MODELLING TRIP GENERATION / TRIP ACCESSIBILITY USING LOGIT MODELS

By

Shucheng Hu

A thesis submitted in partial fulfilment of the requirements of Edinburgh Napier University, for the award of Doctor of Philosophy

School of Engineering and the Built Environment

Edinburgh Napier University

September 2010

ABSTRACT

Trip generation is the first stage of the conventional 'four-stage' transport model. The aim of this stage is to predict total number of trips generated to and from each zone. The two most common techniques for trip generation are linear regression (the dependent variable is a linear-in-parameter function of a number of explanatory variables) and category analysis including multiple classification analysis (based on estimating number of trip generations as a function of household attributes). Both techniques of trip generation rely on the availability of a large socio-economic, mainly revealed preference data set. They also have technical limitations such as the assumption of linearity which might result in unreasonable predictions of trip generation. Any deficiency or inaccuracy in the estimation at this stage will be carried over and will have implications on subsequent stages.

The other stages of the 'four-stage' model employ other techniques including logistic analysis which broadens the scope of the analysis. Logistic regression analysis has been used to model travel choices such as mode, route and departure time but not trip generation. There has not been much research to investigate the appropriateness of using this technique to model trip generation. The main reason for this is that logistic regression predicts probabilities rather than the total number of trips.

In order to be able to model trip generation using logistic regression, the number of trips (trip frequency) can be treated as a set of mutually exclusive categorical variables; therefore the built-in upper and lower limits are incorporated. Therefore, it is not possible to predict a negative number of trips and the estimates of the model will show the underlying probabilities for the actual number of trips. This will also provide a behavioural framework that directly links the number of trips to utility-based consumer and decision-making theory. Logistic regression can be used to model trip generation as binary, multinomial or nested logit frameworks. An added advantage of using this approach is the ability to predict the frequency and number of trips made by each individual.

The aim of this research therefore, is to investigate possible methodologies to improve performance of trip generation modelling. In order to achieve this aim firstly, this research investigates the appropriateness of logistic regression to model trip generation and device a methodology for it. The analysis and comparisons of the results with results from conventional models are examined. Exploring the use of stated preference data to calibrate trip generation models is also studied here. Finally, transport policy measures and enhanced transport accessibility functions have been investigated in trip generation models.

ACKNOWLEDGEMENTS

I gratefully acknowledge the financial assistance provided by School of the Built Environment, Edinburgh Napier University. Thanks are also extended to the TRI for its support during my studies and to the Edinburgh City Centre Management Company for the data used in this study.

Special thanks are given to my supervisor Dr. Wafaa Saleh; without her continuous advice and encouragement, I would have never completed my thesis. Thanks are also given to Professor J. de D. Ortúzar and Professor Howard Kirby for their comments on some parts of this work.

My deepest gratitude goes to my family: to my parents, my wife and my daughters for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 INTRODUCTION

Trip generation is the first phase of the classical 'four-stage' transport model (trip generation, distribution, modal split and assignment). Trip generation is defined as the number of individual trips generated in a given period of time. The purpose of this stage is to predict the total number of trips which are generated from and attracted to each zone. Trip generation analysis provides the means for relating the number of trips in any zone to its land-use and socio-economic characteristics such as land use intensity, characteristics of activities and location within the urban environment. Trip generation models attempt to identify and quantify the trip ends related to various urban activities without taking into account other trip characteristics such as direction, length or duration (FHWA, 1975).

The two most commonly used techniques of trip generation modelling have been linear regression analysis and category analysis. Both approaches have their strengths and weaknesses. In regression analysis, although there are statistical tests for the goodness of fit of the models, the assumption of linearity of each of the independent variables with the dependent variables is restrictive. The lack of built-in upper and lower limits to the number of trips could potentially lead to unreasonable predictions as the model's covariates increase, or could result in negative number of trips when the covariate values are relatively low (Páez *et al.*, 2006). The assumption that the number of trips is approximately continuous can be questioned when typical values for the number of trips are relatively low. The link between number of trips and covariates in a linear regression, while it may be based on hypothetical ideas about the process of trip generation, lacks a behavioural justification such as supported by the theory of random utility (e.g. Ben-Akiva and Lerman, 1985).

Alternatively, in category analysis the large sample size required to calibrate the trip rates as well as the absence of statistical tests for the overall goodness of fit of the models undermines its adequacy (see Stopher and McDonald, 1983; Ortúzar and Willumsen, 2001). Multiple classification analysis (MCA) methods provide improved techniques to overcome some of the shortcomings of category analysis approach. In MCA's, the new cell values are calculated based on the data sample within the given cell, as well as on an overall mean derived from the whole data set. These means could also be weighted average means or least square regressions of the dummy variables. In addition to overcoming the main shortcomings of category analysis approach the MCA methods, allow goodness-of-fit statistical tests that permit hypothesis-testing procedures to be followed, and results to be assessed in terms of the amount of the variability of the dependent variable that is captured in the model.

Logistic regression overcomes many of the restrictive assumptions of ordinary least squares regression (Garson, 2002); in particular, the assumption of linearity between the dependent and independent variables. This technique can be used to model relationships between the response variables which are binary or categorical, with more than two categories and several explanatory variables which may be categorical or continuous. This approach has been widely used to model other travel choices such as choice of mode (Ortúzar, 1983; Bhat, 1995; Bhat, 1998a; Ortúzar and Willumsen, 2001), route choice (Yai *et al.*, 1997), departure time choice (Bhat, 1998b; Saleh and Farrell, 2005) and other travel choices. However, not many applications in trip generation modelling have been reported (see for example Daly, 1997).

Discrete choice models, by treating the number of trips (or the trip frequency) as a set of mutually exclusive and collectively exhaustive categorical variables, incorporates built-in upper and lower limits. They cannot predict a negative number of trips and the estimates of the model show underlying probabilities for actual number of trips, whereas the linear regression model only gives the expectation (and variance) of the number of trips, as implicitly the dependent variable would be a continuous variable. In addition, the model provides a behavioural framework that directly links the number of trips to utility-based consumer and decision making theory.

Logistic regression can be used to model trip generation using binary logit models (whether or not an individual will make a trip), or multinomial logit models (probability of making {0, 1, 2 or more trips}, or probability of making {infrequent, frequent, very frequent trips}, etc. This way, one can investigate the frequency of trips combined with the number of trips made by each individual or household (see Hosmer and Lemeshow, 2000 for further discussions on the applications of logistic analysis). This research investigates modelling trip generation using logistic regression analysis. A number of trip generation models using linear regression, category analysis and logistic regression analysis have been calibrated and compared.

The independent variables that are most commonly considered in trip generation models are mainly socio economic variables (individual or household attributes) as well as attraction opportunities. One of the main criticisms of trip generation models is the absence of any variables that represent the transport policies implemented in zones that affect its accessibility (e.g. public transport, pricing and parking policies). Typically accessibility refers to the "ease" with which desired destinations may be reached and is frequently measured as a function of the available opportunities (such as employment levels and retail or non-retail square footage) moderated by some measure of impedance (such as distance, travel time or cost) (Niemeier, 1997).

Previous researches that have attempted to develop trip generation models that include impacts of transport policies or accessibility are limited. For example, Hanson (1959) calibrated a trip generation (production) model with an accessibility index for each zone in the study area as a measure of the activities in other zones and a measure of travel impedance between each zone pair. Freeman (1976) developed a similar model for trip attractions. In both cases, the accessibility index was a function of opportunities and travel impedance (mainly time or cost). Leake and Huzayyin (1979) proposed a composite measure of accessibility which combined private transport and a public transport accessibility measure. Daly and colleagues (Cohn *et al.*, 1996; Daly, 1997) introduced an accessibility measure in the logit trip generation model, which is the logsum from the mode/destination choice model. Transport policies such as road user charging and parking pricing however, have not previously been explicitly included in a trip generation model.

Congestion charging as well as parking management measures are increasingly being considered as management tools in the UK as well as in most world cities (Litman, 2004; European Commission, 2004). In London, a congestion charging scheme has been implemented since February 2003 to control traffic congestion into the city (Banister, 2003). Recently, the City of Edinburgh had plans to introduce congestion charging in the form of a double cordon as a policy to reduce traffic in the central areas. Although the scheme has been abandoned following a public referendum (CEC, 2005), a number of research studies and investigations have been carried out to investigate public acceptance of the scheme as well as the forecasts of the impacts of the schemes on various types of travel behaviour. In this research, parking costs and congestion charging in Edinburgh have been investigated as accessibility measures in trip generation models using logistic regression.

1.2 JUSTIFICATION OF RESEARCH

Trip generation analysis is the first stage of the conventional four stage model. Any inaccuracies in the estimation of trip generation will be carried over the subsequent stages. Trip generation techniques suffer from a number of deficiencies. ***The aim of this research is to investigate possible methodologies to improve performance of trip generation modelling. In order to achieve this aim a number of objectives have been defined as discussed below.***

In linear regression analysis, the assumption of linearity of each of the independent variables with the dependent variables is a strong restrictive. The lack of built-in upper and lower limits to the number of trips could potentially lead to unreasonable predictions, or could result in negative number of trips when the covariate values are relatively low. The assumption that the number of trips is approximately continuous can also be questioned especially where the number of trips are low. The lack of a behavioural justification in trip generation such as supported by the theory of random utility for example is also a drawback of this stage (e.g. Ben-Akiva and Lerman, 1985). Similarly, in category analysis the large sample size required to calibrate the trip rates as well as the absence of statistical tests for the overall goodness of fit of the models undermines its adequacy (see Stopher and McDonald, 1983; Ortúzar and Willumsen, 2001). Although multiple classification analysis (MCA) methods provide improved techniques to overcome some of the shortcomings of category analysis approach, these methods largely suffer from same limitations of category analysis.

In summary, trip generation analysis, unlike the rest of travel choice analysis, has limitations in terms of the techniques (conventional techniques), data used (only revealed preference data) and type of variables (only socio-economic variables). These limitations have been recognised in the literature and acknowledged to impair the efficiency of trip generation models to produce accurate predictions.

Logistic regression analysis may offer a way forward to overcome some or all of the above mentioned limitations of trip generation techniques. It overcomes many of the restrictive assumptions of ordinary least squares regression (Garson, 2002); in particular, the assumption of linearity between the dependent and independent variables. This technique can be used to model relationships between the response variables which are binary or categorical, with more than two categories and several explanatory variables which may be categorical or continuous. This approach has been widely used to model other travel choices such as mode, route, departure time and other travel choices. However, not many applications in trip generation modelling have been reported. Moreover, this approach would allow the use of other sources of data such as stated preference and stated intention data.

The first objective of this research therefore, is to investigate appropriateness of logistic regression analysis for modelling trip generation.

In order to do that, a number of data sets have been identified and analysed to carry out the investigations. These are presented and discussed in Chapter 5. Secondly, the methodology adopted to model trip generation using logit analysis as well as the calibrated work trip models are presented in Chapter 6.

In order to further assess the performance of the logit models of trip generation, they have been compared with the conventional trip generation models (i.e. linear regression analysis and category analysis). There are a number of multiple classification analysis techniques which have been recently developed but not widely empirically tested.

The second objective of this research is to investigate, analyse and compare trip generation models using logistic regression, linear regression and category analysis including multiple classification analysis.

Calibration of trip generation models using the conventional (linear regression and category analysis including multiple classification) models is presented in Chapter 7. Predictions from all the above models and analysis of the results are presented in Chapter 8.

One of the main criticisms of trip generation models is the absence of any variables that represent the transport policies which are implemented in zones that affect its accessibility (e.g. public transport, pricing and parking policies). As discussed earlier, the independent variables that are most commonly considered in trip generation models are socio economic variables (households/individuals' attributes) as well as attraction opportunities. Congestion charging as well as parking management measures is increasingly being considered as management tools in the UK as well as in most world cities (Litman, 2004; European Commission, 2004). There is empirical evidence that such policies do affect trip generations as well as other travel decisions (e.g. trip distributions, modal choice and route choice). However, most of current trip generation models still ignore this type of variables, and only include mainly socio economic characteristics.

For example, there has been a large number of parking management schemes implemented in the UK over the past few decades to reduce congestion. There are a lot of empirical evidences that these schemes have resulted in a reduction of number of shopping and other trips to the central areas. Therefore, to ignore the impacts of such policies on trip generations and only consider them at later choice decisions would certainly be resulting in inaccurate predictions at this, and all subsequent stages.

The third objective of this research is to investigate the impacts of including factors to represents transport policy in the trip generation models on their performance.

In order to achieve this, a data set from the household and shoppers' survey in Edinburgh, has been used to calibrate linear and logistic regression models of trip generation (shopping trips), taking into account parking costs as transport policy measure. These results are presented in Chapter 9.

Most trip generation models are calibrated from aggregate revealed preference data (Daly and Miller, 2006). This is despite the growing number and extent of applications in other sources of data (e.g. stated preference and stated intentions) and the great number of applications in travel forecasting models using these data. This is mainly because of the nature of trip generation models and modelling techniques used (i.e. linear regression analysis and category analysis). SP techniques offer the opportunity to modellers to test impacts of policy measures on travel behaviour. So in principle there is no reason why these techniques cannot be used in trip generation modelling, especially if logistic regression analysis is used.

For example, in London, a congestion charging scheme has been implemented since February 2003 to control traffic congestion into the city. There are empirical evidences that this scheme has resulted in a reduction of number of shopping and other trips to central London. A similar scheme has been proposed for Edinburgh. And although the scheme has been abandoned, a number of research studies and investigations have been carried out to identify public acceptance of the scheme as well as the forecasts of the impacts of the schemes on various types of travel choices but not including trip generation. It would be interesting therefore to use stated preference techniques to investigate impacts of transport policies on trip generations.

In this research, the fourth objective is to investigate the use of stated preference data for calibrating trip generation models.

In order to achieve this, the SP data from Edinburgh Household Survey is used to calibrate mixed RP/SP logistic regression models for trip generation taking account of introducing road user charging as a policy measure. These results are presented in Chapter 10.

Accessibility refers to the "ease" with which desired destinations may be reached and is frequently measured as a function of the available opportunities (such as employment levels and retail or non-retail square footage) moderated by some measure of impedance (such as distance, travel time or cost) (Niemeier, 1997). Accessibility of the transport system has been recognised and investigated in the literature but also limited to variables representing the characteristics of the transport system but not the perceived level of service of that system.

Finally, in this research therefore, the inclusion of transport accessibility measure in trip generation models is explored and analysed.

A public transport accessibility measure is calibrated as a function of the distance from the city centre and the perceived level of service of the public transport system by the users. These results are presented in Chapter 11.

1.3 RESEARCH OBJECTIVES

The main objectives of this research are to:

1. Examine appropriateness of logistic regression analysis for modelling trip generation in order to overcome any problems related with the conventional methods (i.e. trip generation and regression analysis).
2. Investigate, analyse and compare trip generation models using logistic regression, linear regression and category analysis including more recent multiple classification analysis techniques. This is to further test the statistical significance and hence the appropriateness of logistic regression analysis for trip generation.
3. Investigate and calibrate trip generation models which include transport policy measures to investigate if these models will improve the prediction and statistical significance of trip generation models.
4. Explore the use of stated preference data (SP) to calibrate trip generation models. This is to make use of this data source and to improve the validity and performance of trip generation models similar to other travel demand forecasting models (e.g. modal split models).
5. Investigate trip generation models with enhanced transport accessibility functions to make trip generation models more realistic.

1.4 NOVELTY OF THIS RESEARCH

Limitations in trip generation techniques and analysis have been widely recognised in the literature, yet very limited investigations and innovations of these techniques have been reported to date. Trip generation is the first stage in the analysis and forecasting of demand for travel. Any deficiency or inaccuracy in the estimation at this stage will be carried over and will have implications on subsequent stages. While logistic regression analysis has been extensively used in mode, route, destination and departure time choices, it has not been used in modelling trip generation. Logistic analysis can overcome some the limitations of linear regression analysis and category analysis as discussed above. For example, the assumption of linearity of independent variables, the lack of built-in upper and lower limits to the number of trips and the assumption that the number of trips is approximately continuous can also be questioned.

This research defines a framework for modelling trip generation using logistic analysis.

Moreover, a number of multiple classification analysis techniques which have been recently developed but not widely empirically tested, are used to calibrate and analyse work trip generation models.

The research calibrates trip generation models with independent variables that represent transport policies (such as parking pricing and congestion pricing). This is very important since the absence of the effects of such policies at the trip generation stage would result in inaccurate prediction of travel demand forecasting, even though these impacts are considered at later decision choices such as mode and route choice.

Stated preference data and techniques have been investigated in other travel decision models, but not in trip generation modelling. In this research, trip generation models have been calibrated using mixed SP/RP techniques.

Finally, the research also investigates modelling transport accessibility in trip generation models by including a public transport accessibility measure. This measure reflects the transport users' perceived levels of service of public transport. Transport accessibility can include other measures which reflect the level of accessibility of the transport system.

CHAPTER 2 TRIP GENERATION MODELLING

2.1 INTRODUCTION

In transport modelling, 'trip' or 'journey' (both terms are used interchangeably here) is a one-way movement from a point of origin to a point of destination (Ortúzar and Willumsen, 2001). A *Home-Based (HB) Trip* is one where the home of the trip maker is either the origin or the destination of the trip and a *Non-Home-Based (NHB) Trip* is, conversely, one where neither end of the trip is the home of the traveller. Trip Generation is often defined as the total number of trips generated by households or individuals, be they HB or NHB. A Trip Production is defined as the home end of an HB trip or as the origin of an NHB trip and a Trip Attraction is normally defined as the non-home end of an HB trip or the destination of an NHB trip.

During the 1980s a series of other terms, such as tours and trip chains, appeared in transport modelling; and these correspond better to the idea that the demand for travel is a derived demand (i.e. it depends strongly on the demand for other activities, Ortúzar and Willumsen, 2001) and have been used mainly by discrete choice modellers in practice (Daly *et al.,* 1983). A tour or trip chain can be defined as a sequence of trip segments that start at home and end at home (Shiftan, 1999).

2.1.1 Classification of trips

In practise, trips are often classified by different purposes to obtain better trip generation models. By purpose, personal trips are commonly classified into (Barber, 1985): work trips, shopping trips, social trips, recreational trips, school trips, home trips and business trips. This research focuses on work trips and shopping trips respectively. A *work trip* can be defined as a trip made to a person's place of employment (Barber, 1985); the place of employment may be a manufacturing plant, a public or private institution such as a hospital or university. A *shopping trip* can be defined as a trip made to any social outlet, regardless of the size of the store (or shopping centre) and whether or not a purchase was actually made. Among all trip purposes, work trips used to be most numerous followed by shopping trips (Vickerman and Barmby, 1984). The National Travel Survey data (Department of Transport, 1979) show that shopping trips have increased from 12.7% of all trips in 1965 to 16.6% in 1975/1976 while work trips have fallen from 35.7% in 1965 to 25.7% in 1975/1976. In 1996/1998 shopping trips accounted for 20.3% of total trips and has become more numerous than commuting trips which accounted for 18% of that total (Kershaw *et al.*, 2001).

Work trips and school trips are usually called compulsory (or mandatory) trips and shopping trips, social and recreational trips and some other less routine trips (such as seeing a doctor) are called discretionary (or optional) trips (Ortúzar and Willumsen, 2001). When transport policies are introduced, it would mostly impact on discretionary trips than compulsory trips. Trip generation models for different types of trips can vary either by the factors in the equations or by the value of the coefficients of the same factor.

By time of day, trips are often classified into peak and off-peak period trips and the proportion of journeys by different purposes usually varies greatly with time of day (Ortúzar and Willumsen, 2001). The majority of trips in the AM peak are usually compulsory (i.e. either to work or education) and this is not the case in the off-peak period.

Trips can also be classified by person type, as individual travel behaviour is heavily dependent on socio-economic attributes such as income levels, car ownership and household size and structure (Ortúzar and Willumsen, 2001).

2.1.2 Aggregate and disaggregate approaches

There are two approaches in terms of data aggregation in trip generation models: aggregate trip generation models and disaggregate trip generation models. The aggregation levels are usually defined as area (zonal), household, and person. In aggregate models, a given geographic area, such as neighbourhood or city, are used as the unit of analysis. In disaggregate models, the household or individuals are used (Koppelman and Pas, 1984). Estimating the models at more disaggregate levels improves the transferability of trip generation models (Ortúzar and Willumsen, 2001).

Atherton and Ben-Akiva (1976) emphasized that disaggregate models tend to maintain the variance and behavioural context of the response variable and, therefore, are expected to give better estimates when transferred. Downes and Gynes (1976) pointed out that when the explanatory power of the model is of interest rather than the aggregate forecasts, the disaggregate level should be selected. Wilmot (1995) indicated that disaggregate models are preferred because of their independence from zonal definitions. In Supernak *et al.* (1983) and Supernak (1987), the person level was preferred for trip generation models because of the identity of the response factor (trip) and the generative (the person). One advantage of disaggregate person-level models is the reduced amount of data required for model estimation.

At prediction, however, a degree of aggregation will be required. An empirical test of the forecast performance of household- and person-trip generation was conducted by Badoe and Chen (2004) using data collected in a household-travel-behaviour survey in the Greater Toronto Area of Canada. They conclude that the household is theoretically the preferable analysis unit to use in trip production modelling when the model estimation data are collected in a household travel survey in which the household is the sampling unit. The empirical test indicates that household-trip generation models yield predictions of trips at the household and traffic zone level, respectively, that are marginally more accurate than those yielded by person-trip generation models.

2.2 FACTORS INFLUENCING TRIP GENERATION

According to Levinson (1976) and Bruton (1985), trip making is a function of the following basic factors: the socio-economic characteristics of the trip makers residing in the area, and the land-use pattern and developments in the study area (or the physical characteristics of the area).

The explanatory variables used in trip generation models will differ depending on the type of trip being modelled (Sheppard, 1985). First the potential number of trip makers in a zone should be identified by considering the land use mix or the number of residence. Secondly, the degree to which a potential trip maker's characteristics affect his or her propensity to make a trip should be considered. Lastly, the geographical accessibility of the zone to places where the trip purpose will be satisfied can also affect the number of trip made.

In general, the explanatory variables can include: 1) social-economic characteristics of the trip maker; 2) physical and demographic characteristics of the area; and 3) accessibility and policy-related measures. Some of these variables are important when aggregate data are used and some of them are important in disaggregate (e.g. household and individual) models. These variables are classified into three main groups according to their roles in aggregate models and disaggregate models. The discussions are based on Bruton (1985) and Stopher and McDonald (1982).

2.2.1 Factors affecting aggregate trip generation models

The factors which are important in aggregate (zonal) trip generation models are summarised in Table 2.1 and discussed in the following sections.

Table 2.1 Factors affecting aggregate trip generation models

Factors	References
Location / land use factor	Buchanan and Partners, 1965; Douglas and Lewis, 1970, 1971; Bruton, 1985; Páez et. al, 2006
The social-economic characteristics of the population	Schuldiner, 1962; Taylor, 1968
Density	Stopher and McDonald, 1982; Bruton, 1985
The degree of urbanization	Schuldiner, 1962

2.2.1.1 Land-use factors / area type / location variable

Location reflects the surrounding environment and should ideally measure the spatial separation of households from each of the amenities which they desire, e.g. schools, shops and workplaces (Douglas and Lewis, 1970, 1971). Different uses of land produce different trip generation characteristics. For the purposes of trip generation, the significant land uses include (Bruton, 1985):

- Residential land use, which can be represented in terms of acres of residential land, number of dwelling units, number of dwelling units per acre, number of persons per acre, or total population.
- Commercial and industrial land use, which can be expressed as the numbers employed per unit area of land and the amount of floor space occupied.
- Educational and recreational developments, expressed as the numbers in attendance. The Guildford study, carried out by Buchanan and Partners (1965), included a comprehensive analysis of the effect of the development of the University of Surrey on trip generation and distribution in Guildford.

Where densities are higher, motorized trips are likely to be fewer because opportunities for satisfying activities are closer and both congestion and parking price may be significantly higher, whereas parking availability is lower (Stopher and McDonald, 1982). In addition, various services and home deliveries may be more available, thus reducing the need for some trips. The effect of area type is likely to be greatest on discretionary travel (home-based socio-recreational, home-based other) and least on mandatory travel (home based work or school).

Agyemang-Duah *et al.* (1995) found that suburban living is positively correlated with weekday, home-based shopping trips. Finally, in an elderly trip generation study in the Hamilton CMA by Páez *et al.* (2006), significant spatial variability was detected in the case of work trips, and in the case of non-work trips significant spatial variability within age cohorts was found.

2.2.1.2 The social-economic characteristics of the population

The social-economic characteristics of the population could be expected to produce different movement demands. For example, factory or manual workers could be expected to produce quite different movement characteristics to executive clerical workers. Schuldiner (1962) indicated that a trip generation model based on socio-economic characteristics held some promise. However Taylor (1968) showed that for all modes of travel and a range of journey purposes there appears to be little relationship between the zonal socio-economic characteristics examined by him and trip generation.

2.2.1.3 The degree of urbanization

The degree of urbanization exhibited by an area can be used to represent the level of integration of the household in the local community. Schuldiner (1962) found in his analysis of data relating to Chicago that the index of urbanization, which he derived based on fertility rate, female labour participation rate and the incidence of single family dwellings, appeared to exert a significant effect on trip generation rates. The measure of the degree of urbanization often used is distance from the central area. The argument for the use of this factor is that characteristics of the population and development, and hence the movement demand, change with distance from the central area. For example, within the central area residential development may consist largely of 'temporary' hotel, flat and boarding-house accommodation occupied by young, single or transient persons, while the outer suburbs may consist large of single family dwelling units occupied by married couple with families.

2.2.2 Factors affecting disaggregate household trip generation models

The factors which are important in disaggregate household trip generation models are summarised in Table 2.2 and discussed in the following sections.

Table 2.2 Factors affecting disaggregate household trip generation models

Factors	References
Family income	Stopher and McDonald, 1982; Bruton, 1985; Takyi, 1990
Vehicle ownership	Stopher and McDonald, 1982; Bruton, 1985; Agyemang-Duah *et al.*, 1995; Agyemang-Duah and Hall, 1997; Schmöcker *et al.*, 2005
Household structure	Allaman *et al.*, 1982; McDonald and Stopher, 1983
Household size	Schuldiner, 1962; Stopher and McDonald, 1982; Agyemang-Duah *et al.*, 1995; Takyi, 1979, 1990
Number of children	Agyemang-Duah *et al.*, 1995
Occupied residence	Stopher and McDonald, 1982; Bruton, 1985
Life style and life cycle	Allaman *et al.* 1982; Ortúzar and Willumsen, 2001; Chicoine and Boyle, 1984
Type of dwelling unit	Schuldiner, 1962; Stopher and McDonald, 1982; Bruton, 1985
Value of a property	Bruton, 1985

2.2.2.1 Family income

The ability to pay for a journey affects the number of trips generated by a household (Bruton, 1985; Stopher and McDonald, 1982). Thus families with a high income can generally afford to satisfy more of their movement demands than low-income families. As one would expect, increasing family income leads to greater trip production. Family income tends to be related to levels of motor vehicles ownership. In the analysis of trip generation in a developing country by Takyi (1990), it has been found, when household income was included in the same model with car ownership, its influence on trip making was significantly reduced.

2.2.2.2 Motor vehicle ownership / license ownership

Motor vehicle ownership, or the number of vehicles available for use by each household, has been found to have a significant influence on trip generation (Bruton, 1985). Households with more than one motor vehicle tend to generate more trips per unit than households with only one motor vehicle, although the single-car households tend to utilize their vehicle more intensively. Motor vehicle ownership and family size are to a certain extend related. A large non-motor-vehicle-owning family can be expected to generate fewer trips than the same size family which has access to three motor vehicles. The most common measures of car ownership are the total number of cars per zone, car ownership per person, or car ownership per household.

The acquisition of a vehicle increases substantially the number of trips and motorized trips made by a household (Stopher and McDonald, 1982, also Agyemang-Duah *et al.*, 1995; Agyemang-Duah and Hall, 1997, and Schmöcker *et al.*, 2005), this arises both from substitution of vehicular trips for walk trips and from satisfaction of vehicular trips for walk trips and from satisfaction of previously unsatisfied demand for travel. The trip making rate of increase is nonlinear, with a decrease rate of increase with increasing automobile. Vehicle availability is likely to be the more appropriate measure than ownership because it is a more accurate measure of the potential to satisfy demand for vehicle trips. Also the number of vehicles has nonlinear effects on discretionary trip generation.

The elderly trip generation study in the Hamilton CMA by Páez *et al.* (2006) indicates that license ownership relates positively with trip making frequency and it also shows that license ownership is a stronger predictor of number of trips than car ownership. License ownership and car ownership are found to be the two most important factors affecting elder trip generation, but, it is also found that overall mobility may not necessarily be negatively affected by lack of access to a car, presumably as long as transit remained accessible.

However, the results from the studies by Vikerman and Barmby (1985) and Barmby and Doornik (1989) using the Sussex Household Shopping Survey data show that car ownership has no clear effect on shopping trips.

2.2.2.3 Family size

Household size is defined as the number of persons in the household without regard to age, and it is expected to cause increases in trip making for all trip purposes, although not in a uniform manner (Stopher and McDonald, 1982). The number of trips per person is expected and has been shown to be relatively stable. Schuldiner (1962) in his work on the Modesto area of California has shown that average trip frequency increases with increasing persons per household, at the rate of approximately 0.8 trips per day for each additional person. This increase in the number of trips with family size is, however, related mainly to non-work trips which tend to level off at the four person per dwelling unit family size.

In trip generation analysis in Ghana by Takyi (1990), household size, which reflects the extended family in developing countries, has been found to be the strongest determinant of trip making, together with car ownership and the number of employed persons in the household, although trip rates were not significantly increased for household sizes larger than eight. In this case, household size as a variable performs significantly better than household income for work, school and shopping trips, which makes up more than 60 percent of total household trips.

Agyemang-Duah *et al.* (1995) point out that the household home-based shopping trips increase with increasing household size but at a decreasing rate and household sizes have non linear effects on discretionary trip generation. An earlier study by Takyi (1979) also shows that there is a nonlinear relationship between household size and the average number of trips per household.

2.2.2.4 The number of children

The presence of children in the family may have a dual influence on shopping travel (Agyemang-Duah *et al.*, 1995); on one hand, it may lead to some restrictions on the time available for shopping. Alternatively, it may be regarded as a scale factor leading to increased shopping. When household size is included in the meantime, to some extent one might expect the number of children to have a negative effect; and this is confirmed by their weekday, home-based shopping trip generation study (Agyemang-Duah *et al.*, 1995). An explanation is that children of school age are at school and childcare responsibilities might have some time budget effects on trip making.

2.2.2.5 Occupied residents

It has been found that the proportion of work trips for the gainfully employed groups decreases as the occupational status increases, although the proportion of trips for non-work purposes varies little between various groups with the exception of the unemployed (Bruton, 1985).

The number of workers may be defined as all workers or as full time workers only, where worker is restricted to work outside the home (Stopher and McDonald, 1982). The number of workers will be in direct proportion to and is causative of the number of household work trips. Also, as more members of a household of a given size work, the number of trips for all other purposes is likely to be fewer, except for non-home-based trips, because more activities are likely to be undertaken on the way to or from work.

2.2.2.6 Life style and life (or family) cycle and household-structure

Ortúzar and Willumsen (2001) suggest that life cycle variables could be an important factor for explaining trip generation. This is consistent with the idea that travel is a derived demand and that travel behaviour is part of a larger allocation of time and money to activities in separate locations. For example, the concept of life style may be operationalised as the allocation of varying amounts of time to different (activity) purposes both within and outside the home, where travel is just part of this allocation (See Allaman *et al.* 1982).

It can be tested whether the major break points or stages (such as the appearance of pre-school children; the time when the youngest child reaches school age, the time when all the children of a couple have left home, and the time when all members of a household have reached retirement age) in the life cycle are consistent with major changes of time allocation. Different trip rates can be expected for households and people at various stages of life and, furthermore, age should correlate with employment, having a driver's license, and marital status.

The concepts of life style and stage of family cycle are important from two points of view: first, that of identifying stable groupings (based on age or sex) with different activity schedules and consequently demands for travel; second, that of allowing the tracing of systematic changes which may be based on demographic variations (e.g. changes in age structure, marital or employment status).

Chicoine and Boyle (1984) use the Automatic Interaction Detector program to determine the important components of a life-cycle classification scheme which emphasize the presence of children more than ages of children. They conclude that the advantage of a life-cycle-based trip generation procedure over regression models lies in its simplicity and its ability to handle non-numeric values. It is preferable to a procedure based on family size because it explicitly addresses family structure and thus takes intrahousehold interactions into account. Finally, a life-cycle-based procedure uses readily available data; an income-based procedure is vulnerable to high nonresponse rate if a noncensus data source is used, and such a scheme must be constantly adjusted to account for the effects of inflation.

Allaman *et al.* (1982) use a household structure variable based on the above ideas in trip generation modelling and the household categories are based on the age, gender, marital status, and last name of each household member. It was expected that these categories would have varying effects on trip rates. For example, adults living alone would be less mobility constrained than those adults living with children; but they would have none of the opportunities for trip coordination produced by living with other adult members. More specifically, when trip-generation rates are analysed by purpose groups, differences between the trip-generation rates of these household categories would be expected.

Allaman *et al.* (1982) examined this household-structure concept by using Baltimore survey data with linear regression analysis and suggest that the household-structure variable correlates more strongly with trip rates than almost any other variable, except vehicle ownership. In particular, this should improve the model significantly where it is combined with vehicle ownership and used as a substitute for household size.

McDonald and Stopher (1983) tested this variable using Midwest data by using both analysis of variance and multiple classification analysis (MCA) in contrast to linear regression and conclude that the household–structure variable does not perform significantly better than the other variables tested. The contrary indications may, however, be a result of the different methodologies that were used in the two analyses. They further mention that even had household-structure variable performed satisfactorily in the trip generation analysis, there would be problems implementing it in trip-generation models as, it appears to have problems when forecasting at zonal level, particularly to obtain distribution of households by household-structure category.

2.2.2.7 Type of dwelling unit

The more permanent types of dwelling unit, such as a single family house, reflect a high degree of integration into the local community on the part of the household, and lead to a high rate of trip generation (Bruton, 1985). Conversely the less permanent dwellings, e.g. a hotel room, result in a more limited integration with local affairs, with a lower resultant trip generation rate. Schuldiner (1962) found that this was the case as well but not as marked as expected. However, when family size and car-ownership levels are taken into consideration, the difference in generation rates is not as great as appeared at first sight. Similarly, Stopher and McDonald (1982) suggest that household type has a weak conceptual link, deriving principally from density considerations and some aspects of vehicle availability associated with vehicle storage space.

2.2.2.8 Rateable value of a property

The rateable value of a property is considered indicative of the occupiers' financial status (Bruton, 1985). Thus the greater the annual outgoing in rent, or interest on invested capital, the more likely it is that the occupiers have resources available to spend on travel. Rateable value is related to family income and usually easier to obtain reliable information about it.

2.2.3 Factors affecting disaggregate individual trip generation models

The factors which are important in disaggregate individual trip generation models are summarised in Table 2.3 and discussed in the following sections.

Table 2.3 Factors affecting disaggregate individual trip generation models

Factors	References
License ownership	Páez *et al.*, 2006
Age	Bruton, 1985; Páez *et al.*, 2006
Employment status/job type	Agyemang-Duah *et al.*, 1995; Páez *et al.*, 2006
Telecommuting	Mokhtarian *et al.*, 1995; Henderson and Mokhtarian, 1996; Henderson *et al.*, 1996; Koenig *et al.*, 1996
Teleshopping/ electronic-shopping	Lenz, 2003; Farag *et al.*, 2003

2.2.3.1 License ownership

The elderly trip generation study in the Hamilton CMA by Páez *et al.* (2006) indicates that license ownership relates positively with trip making frequency and it also shows that license ownership is a stronger predictor of number of trips than car ownership. License ownership and car ownership are found to be the two most important factors affecting elder trip generation, but, it is also found that overall mobility may not necessarily be negatively affected by lack of access to a car, presumably as long as transit remained accessible.

2.2.3.2 The age structure of the population

The age structure of the population is often taken into consideration in trip generation analysis on the basis that different age groups produce different movement demands and characteristics (Bruton, 1985). The teenage population 15-20 years, for example, could be expected to produce more journeys of a social and recreational nature than older age groups.

In the elderly trip generation study in the Hamilton CMA by Páez *et al.* (2006), the results also confirm the negative association between increasing age and trip making frequency. However, it is found that this behaviour is not spatially homogenous, in particular with respect to non-work trips. The results also suggest that a sizable segment of the 65+ cohort tends to engage in increased trip making, relative to other cohorts.

2.2.3.3 Employment status and job type

Different employment status may exert different time budget constraint on shopping trips (Agyemang-Duah *et al.*, 1995). Full-time and to some extent, part-time work is expected to have a negative impact on weekday home-based shopping trips. Two opposed effects of unemployment may be hypothesized: one effect is that an unemployed person has more time and therefore can make more shopping trips and the other hypothesis is that because a person is unemployed, he or she does not have enough money for shopping. In their studies of weekday home-based shopping trips in the greater Toronto Area (GTA), it has been found that full-time employment has a negative impact on home-based, weekday shopping trips; however, the effect of unemployment is not statistically significant in the shopping trip generation.

Páez *et al.* (2006) found that full time employment has a positive, but relative small, impact on total trips, but the difference in trip making frequency between blue collar and other workers is negligible. While for work trips the single most important factor is employment status, being employed full time correlates negatively with number of non-work trips made, but the effect is relatively small.

2.2.3.4 Telecommuting

Henderson and Mokhtarian (1996) investigate the impacts of centre-based telecommuting on individual travel behaviour and emissions, using travel diary data from the Puget Sound Telecommuting Demonstration Project. A telecommuting centre, or telecentre, is defined as a facility where employees (from single or multiple organizations) share workplace and equipment for the purpose of reducing the length of the commute from the employee's home to the workplace. An analysis of personal vehicle usage for this small sample of workers showed that the number of vehicle-miles travelled (VMT) was reduced significantly as a result of centre-based telecommuting. The number of personal vehicle trips did not change significantly. In essence, centre-based telecommuters behave as conventional commuters in terms of their number of trips, but more similar to home-based telecommuters in terms of VMT reductions. Home-based telecommuting has been found to reduce the number of daily trips and VMT, leading to substantial savings in personal vehicle emissions (Mokhtarian *et al.*, 1995; Henderson *et al.*, 1996; Koenig *et al.*, 1996). However, home-based telecommuting may not be appropriate for every worker whose job permits it (Bagley *et al.*, 1994), for reasons such as no adequate space and distractions; while telecommuting centres may offer more opportunity of social or professional interaction and provide expensive specialized equipment that can be shared by all telecommuting employees.

2.2.3.5 Electronic shopping

Recently, the impacts of electronic shopping (e-shopping) / electronic commerce on travel behaviour have been studied by some researchers (such as Lenz, 2003; Farag *et al.*, 2003). In the research carried out by Lenz (2003) in the Stuttgart region in Southwest Regional in southeast Germany, it is concluded that there is little hope for larger traffic reduction through e-commerce and that e-commerce will have a stronger impact on traffic and transportation only when it is broadly used for everyday standard shopping. Farag *et al.* (2003) use an Internet survey and Netherlands National Survey data to analyse the possible impact of e-shopping on travel behaviour and their main conclusions include: First, some shopping time will be saved and used for other maintenance or leisure activities instead; Second, e-shopping will affect travel behaviour most in the urbanized western part and in the less urbanized parts of the Netherlands; Finally, a reduction in car travel in the less urbanized areas of the Netherlands and a reduction in walking and cycling in the more urbanized areas of the Netherlands are expected.

2.2.4 Accessibility and policy-related measures

Accessibility and policy-related measures are given in Table 2.4. The detailed discussions are as in the following sections.

Table 2.4 Accessibility and policy-related measures

Factors	References
Traffic demand management (TDM) measures: **1.** Pedestrianisation and traffic calming; **2.** Park and Ride; **3.** Parking restraint policy; **4.** Congestion charging	Still and Simmonds, 2000 **1.** Hass-Klau, 1993; Wiggin, 1993 **2.** Cairns, 1997 **3.** Still and Simmonds, 2000 **4.** Schmöcker *et al.*, 2006
Parking at work	Páez *et al.*, 2006
Public transport cost	Vickerman and Barmby, 1984
Petrol fee	Vickerman and Barmby, 1984
Accessibility	Hansen, 1959; Ben-Akiva and Lerman, 1979; Niemeier, 1997. See Chapter 3 for more references.

2.2.4.1 Toll measures and pedestrianisation and traffic calming

Schmöcker *et al.* (2006) reviewed the impacts of road pricing on retail and analysed the shopping trips into London's central shopping district (Oxford Street area) before and after the introduction of the congestion charging scheme in February 2003. The impact of any traffic demand management (TDM) measure on urban vitality is still in a research stage (Still and Simmonds, 2000) and a reason for this is that these policies mostly do not come as an isolated measure but as a package with other policies, which complicates the impact assessment.

Pedestrianisation and traffic calming has slightly negative impact on the retail sector; however, in the long run, it has proved to be beneficial for turnover (Hass-Klau, 1993; Wiggin, 1993). Park and Ride can lead to small change in land use patterns that encourages the development of out-of-town shopping centres with Scottish case studies; on the other hand, it can attract more car bourn customers from the surroundings to the city centre retailers (Cairns, 1997). Although parking restraint policy is always strongly opposed by retailers, there is no statistical evidence that it is linked to the performance of retailing or of other economic sector (Still and Simmonds, 2000).

Schmöcker *et al.* (2006) indicated that the analysis of the surveys provides some evidence of a negative impact on shopping trips at John Lewis, Oxford Street attributable to congestion charging. The main reasons for the reduction in trip frequency include negative experiences with the congestion charging scheme or a generally bad perception of the scheme. However, it is pointed out that evidence from other travel demand measures on city centre shopping activities suggest that the long-time effects of the congestion charge could be more positive.

2.2.4.2 Parking at work

Páez *et al.* (2006) have found that free parking at work has a positive if modest effect on the number of trips, which could be attributed to the relative ease of making subsequent trips, even if not related to work, once that a secure parking base is available.

2.2.4.3 Travel costs

Vickerman and Barmby (1984) and Barmby and Doornk (1989) have found that travel costs affect trip making consistently and there is a very significant tendency to save travel costs by reducing shopping trips as costs increase.

2.2.4.4 The quality of transportation facilities / services, and the level of accessibility

The quality of transportation facilities / services available to the trip maker in a given area and the resulting level of accessibility affect trip generation. Accessibility has not often been used although most studies have attempted to include it as it offers a way to make trip generation elastic (responsive) to changes in the transport system (Ortúzar and Willumsen, 2001); however, unfortunately this procedure has seldom produced the expected results in the case of aggregate modelling applications, because the estimated parameters of the accessibility variable have either been non-significant or of the wrong sign. Detailed discussions of the applications of accessibility in trip generation models will be given in Chapter Three.

The factors discussed above are mainly used for trip production studies. The factors affecting trip attraction can include (Ortúzar and Willumsen, 2001): roofed space available for industrial, commercial and other services, zonal employment and accessibility. For freight trip productions and attractions, important variables include (Ortúzar and Willumsen, 2001): number of employees; number of sales; roofed area of firm and total area of firm.

2.3 TECHNIQUES OF TRIP GENERATION MODELLING

This section reviews trip generation techniques that have been explored in the literature. These modelling techniques can be classified into four main categories as shown in Table 2.5 by their similarity of methodology:

1. Linear regression analysis;
2. Category analysis and its improvements or modifications;
3. Discrete choice / trip frequency models;
4. Other techniques.

The two most commonly used techniques of trip generation modelling are linear regression analysis and category analysis (FHWA, 1975; Hobbs, 1979; Koppelman and Pas, 1984; Bruton, 1985; Sheppard, 1985). First the methodologies, advantages and disadvantages of these two methods will be reviewed in the first two sections. See Ortúzar and Willumsen (2001) for more detailed discussions about the two techniques.

2.3.1 Linear regression analysis

2.3.1.1 Introduction

In the late 1950's and early 1960's linear regression was the most popular method of predicting what the number of trips generated would be if one of the factors affecting trip generation changed. This approach uses trip data collected at one time to determine a functional relationship between trip generation (which are known as the 'response' or 'dependent' variable of the function) and the characteristics that exhibit a causal effect on it (which are known as the 'explanatory' or 'independent' variables of the function) utilising the principle of least-squares, i.e. the squared sum of the residuals or deviations from the estimated line is minimised. The linear least-squares model is based on the hypothesis that there exists a linear relationship between some dependent variable and one or more independent variables.

2.3.1.2 Linear regression model

A trip generation model based on linear regression analysis predicts the number of trips by residents of zone or household *i*, for travel purpose *p* and for person type *n* as:

$$Y_{ipn} = \theta_0 + \theta_{1i} X_{1np} + \theta_{2i} X_{2np} + \ldots + \theta_{ki} X_{knp} + \varepsilon$$

Where

Y = the number of trips generated by an individual, household or zone;

Table 2.5 Classification of trip generation modelling techniques

Category	Modelling techniques
Linear regression analysis	• Multiple linear regression analysis • Some combinations with other models, see the following categories
Category analysis or cross-classification and its modifications or improvements	• Classic cross-classification model • Multiple classification analysis (MCA) methods • The person-category approach • Generalized linear model • Regression analysis for household strata - a combination of linear regression model with category analysis
Discrete choice models (trip frequency)	• Nested-alternative-logit model • Ordered response model • Ordered logit model • Negative binomial model / count data model • Ordered probit model / mixed ordered probit model • Tobit model - a combination of linear model with discrete choice models
Other techniques	• Growth factor modelling • CHAID (Chi-squared Automatic Interaction Detection) • Hierarchical tree-based regression (HTBR) model • Iteratively specified tree-based regression (ISTBR) model - a combination of linear regression and HTBR • Artificial neural networks • Trip chaining and trip generation model • Activity-based trip generation model • Direct demand modelling • Dynamic trip generation model

X_i = the independent variables (number of households, number of workers, car ownership, etc.);

θ_i = the model coefficients estimated by linear regression. That is, for any given set of observations X_1, X_2, ... , X_k there exists a corresponding observation Y which differs from the regression line ($\theta_0 + \theta_1 X_1 + ... + \theta_k X_k$) by the amount of ε;

ε = the error terms which are commonly referred to as the disturbance terms of the equation. They arise in practice mainly because the model does not take account of all factors which influence the value of Y; thus the ε values account for the net effect of excluded variables and random deviations.

2.3.1.3 The assumptions of the linear regression model

The use of least-squares regression analysis involves a number of important assumptions which mainly include (Douglas and Lewis, 1970):

1. *Distribution of the disturbance terms.* Regarding the disturbance terms it is assumed that their mean and co-variance are zero, their variance is constant and that their distribution is normal. If the variance is not constant then data is said to be heteroscedastic and this may lead to an over-statement of the accuracy of the regression equations.
2. *Collinearity between independent variables.* When two or more variables are inter-correlated (it is known as *multi-collinearity*) it becomes difficult to distinguish their separate effects and sometimes the coefficients of a value or sign may be contrary to intelligent expectation.
3. *Error in variables.* Measurement errors in the independent variables are not allowed for by the model and if present can lead to biased estimates of the equation coefficients.
4. *The shape of the response surface.* It assumes that the dependent variable is a linear function of the independent variables. The independent variables need not be in their original forms and transformations such as the logarithm and reciprocal are sometimes used.

2.3.1.4 The tests of the multiple linear regression model

The statistical validity of trip generation analysis derived through linear regression can be assessed by a series of standard statistical tests:

1. *Multiple correlation coefficient (R).* It indicates the degree of association between the independent variables and the dependent variables. Its square is approximately the decimal fraction of the variation in the dependent variable which is accounted for by the independent variables;
2. *'t' test statistic on regression coefficients.* The significance of the regression coefficient of each independent variable in a regression equation is indicated by the '*t*' test statistic. The value of '*t*' is calculated by dividing the regression coefficient by its standard error, and a value of at least 1.96 is necessary for significance to be established at the 95% level.

In addition, the size of the regression constant should be carefully examined - if it is large then the regression set should be used with caution.

Here is an example (Ortúzar and Willumsen, 2001) of a multiple linear regression analysis model to estimate the number of trips per household using number of workers in the household and number of cars (*t*-ratios are given in parentheses):

$$\underset{(3.6)}{Y = 0.84} + \underset{(8.1)}{1.41X_1} + \underset{(3.2)}{0.75Z_1} + \underset{(3.5)}{3.14Z_2} \qquad R^2 = 0.387$$

where

Y is household peak hour trips;

X_1 is the number of workers in the household; and

Z_1 and Z_2 are two dummies for number of cars with Z_1 taking the value 1 for household with one car and 0 in other cases and Z_2 taking the value 1 for households with two or more cars and 0 in other cases (it should be noted that only n-1 dummy variables are needed to represent n intervals); non-car-owning households correspond to the case where both Z_1 and Z_2 are zero.

This model is a good equation in spite of its low R^2. In the model, the intercept 0.84 is not large (i.e. as compared with 1.41 times the number of workers) and the regression coefficients are significantly different from zero with t-ratios 8.1, 3.2 and 3.5. The positive signs of the coefficients are correct, i.e. more workers in a household, more household trips and so with the cars owned by the households. In this example, it is clear that there is a non-linear relationship between household car ownership and the number of trips made by a household and in this case, a model with dummy variables is preferable to that with a single 'number of cars' linear variable.

2.3.1.5 The fits of the linear regression model

There may be a large number of variables to exert a causal effect on trip generation (Douglas and Lewis, 1970, 1971). Some of them may be interrelated and measure largely the same effect and others may exhibit only minor influence. The objective of trip end modelling is to provide a reliable forecasting tool. In the process of trip end modelling attention should be given to the following:

1. The explanatory variables must lend themselves to future estimation and be incorporated in a meaningful way with particular regard to the sign and magnitude of their coefficients.
2. If two explanatory variables are highly intercorrelated, it is desirable to override any automatic selection procedure in order to include only the preferred variable, i.e., the one that either has more meaning or may be more easily forecasted.
3. Known or anticipated change in trip-making behaviour should be reflected in the model. For example, models for vehicle trips must reflect the rising level of vehicle ownership.
4. Generally it will be necessary to estimate beyond the range of data used to develop the model in order that future situations are still suitable, and
5. Zonal regression models only explain the variation in trip making behaviour which exists between various traffic zones and can only provide reasonable future estimates if the "between zone" variance

sufficiently reflects the true reasons for trip variability. Zones thus should be of homogeneous socio-economic composition and should represent as wide a range of conditions as possible.

2.3.1.6 The effect of zonal, household, and personal regression

A zonal regression can only explain the variation in trip-making behaviour which exists between zones. As the zone size increases, the amount of variation between the zones will decrease (Douglas and Lewis, 1970, 1971).

As the aggregate variables directly reflect the size of zone, their use should imply that the magnitude of the error actually depends on zone size; this heterocedasticity (variability of variance) has been found in practice (Ortúzar and Willumsen, 2001). Using a $1/H_i$ (where H_i is the number of households in zone i) multiplier, allows heterocedasticity to be reduced because the model is made independent of zone size. Similarly, it has also been found that the aggregate variables tend to have higher intercorrelation (i.e. multicollinearity) than the mean variables. However, it is important to note that a model using aggregate variables often yields higher values of R^2, as zone size obviously helps to explain the total number of trips (see Douglas and Lewis, 1970).

As the regression models are to be used to predict future trips generated, reasonable forecasts can only be expected if the models take account of a sufficient high proportion of the total variation in trip behaviour. Ideally, therefore, the zones should be as small as possible to maximise the between zone variance and to reduce the within zone variance which is unaccounted for by the model. However, small zones can result in more expensive models in terms of data collection, calibration and operation; and present greater sampling errors which are assumed to be non-existent by the multiple linear regression models. If sampling errors exist in the independent variables, these can produce biased estimates of the regression coefficients.

If zonal aggregation precedes analysis, then the basic relationship between household characteristics and trip-making behaviour are likely to be obscured. Concentration on the household as the basic unit of analysis provides a more meaningful description of the factors underlying trip-making behaviour.

The household models attempt to explain the total variation between households and can be easily expanded to provide zonal trip end estimates. For base year conditions these estimates can be shown to be as accurate as those obtained from zonal based models. The household models are much more likely to be stable over time and will hence provide more reliable future estimates.

Downes *et al.* (1976) used data from a household survey in the Reading area in 1962 and 1971, to compare two alternative types of trip generation model, one based on household trip rates and the other on person trip rates for each household. Statistical considerations favour models based on person trip data because the error variables in household trip rate data is often found to vary with household size and this can invalidate the analytical procedure used to construct the models. Further examination of the residuals errors of one model of each type confirmed that the person rate model was the better of the two. Therefore, in terms of statistical validity and practical utility, it was concluded that models based on person trip rates were preferable to those based on household trip rates.

2.3.1.7 The advantages and disadvantages of regression analysis

The regression analysis method has the following advantages:

1. Regression models are simple;
2. It is relatively easier to include many variables in linear models; and
3. The linear regression models have statistical measures to evaluate the goodness-of-fit, such as *t*-test, the coefficient of determination (R^2) and *F*-test for the complete model.

On the other hand, the regression analysis method has the following disadvantages:

1. The need to assume a linear relationship between dependent variable and independent variables. It is not easy to detect non-linearity because a linear effect may turn out to be non-linear when the presence of other variables is allowed in the model.
2. There is a class of variables, those of a qualitative nature, which usually shows non-linear behaviour (e.g. type of dwelling, occupation of the head of household, age, and sex). In these models, these variables are usually treated as dummy variables where the independent variables under consideration are divided into several discrete intervals and each of them is treated separately in the model. Or some transformation has to be considered, i.e. to transform the variables in order to linearise their effect (e.g. take logarithms, raise to a power). However, selecting the most adequate transformation is not an easy or arbitrary exercise and it takes time and effort.
3. Problems may be encountered in relation to heteroscedasticity and multicollinearity. For zone-based linear regression, the magnitude of the error depends on zone sizes when aggregate variables are used. By using multipliers, this heteroscedasticity can be reduced because the model is made independent of zone size (Ortúzar and Willumsen, 2001).

2.3.2 Category analysis or cross-classification

This section discusses category analysis and a number of enhanced approaches known as multiple classification analysis. An overview of these approaches is given in Table 2.6.

Table 2.6 An overview of category analysis and its modifications or improvements

Modelling Technique	Brief Description	Selected References
Category analysis or cross-classification	It assumes trip generation rates are relatively stable over time for certain household stratification.	Stopher and McDonald, 1983; Ortúzar and Willumsen, 2001
MCA_1	An improvement to the classic cross-classification; based on analysis of variance (ANOVA); The estimated mean trip rates for cells of the cross-classification table utilize a model fit based on data from all cells.	Stopher and McDonald, 1983; Rickard, 1989; Said and Young, 1990; Said *et al.*, 1991, Guevara and Thomas, 2007; Glass and Stanley, 1986
MCA_2	Weighted averages are used. It corresponds to a numerical correction that tries to consider the fact that the number of observations by category is not equal.	Guevara and Thomas, 2007; Ortúzar and Willumsen, 2001; Clark, 1996
MCA_3	It is based on working estimation of the household trip rates by estimating least squares regressions where the independent variables are all dummy variable; one for each of the categories of the strata variables.	Guevara and Thomas, 2007
MCA_4	The trip rates are calculated as the average number of trips by household for each category. It is equivalent to the estimation of an OLS model with dummy variables representing each category.	Guevara and Thomas, 2007; Guevara and Ben-Akiva, 2006; Goodman, 1973
The person-category approach	A person-level category analysis model.	Supernak, 1979; Supernak *et al.*, 1983
Regression analysis for household strata	This method is a mixture of cross-classification and linear regression model.	Hall *et al.* (1987)

2.3.2.1 The classical model

At the end of the 1960s an alternative method for modelling trip generation appeared and quickly became widely used. The method is known as category analysis in the UK (Wootton and Pick, 1967) and cross-classification in the USA. It originally developed in the Puget Sound Regional Transportation Study (1964) and it is based on reporting trips rates per household for any trip purpose as a function of household attributes. In this method, households are categorised into categories on the basis of a cross classification of their characteristics and applies a constant trip generation rate for each category. The advantages of category analysis include that it is easy to understand and no prior assumptions about the shape of the relationship are required. The difficulty with category analysis is the lack of any effective way to choose the best groupings of household characteristics and hence the best categories. Another drawback of category analysis is the lack of inferential statistics, so there is no way to assess the statistical significance of the explanatory variables in trip generation. Finally, the huge samples required to develop the trip rates also account as a drawback of this method.

The dependent variable Y is measured in trip rates ($t^p(h)$ - the average number of trips with purpose p by members of households of type h or t_{jp} – the number of trips with purpose p by the average person in category j). The main assumption made by category analysis is that mean trip production rates do not change (or at least change very little) over the timescale being considered. One of the appealing properties of category analysis is that household characteristics are often of the discrete or qualitative type and so the categories relate to meaningful household units observed in the real world.

The method proceeds as follows for each zone: first, home interview or census data is collected from households to determine the number of trips generated by each household and the characteristics (income, household size, car-ownership, etc.) of that household; second, the households are then divided into categories according to these characteristics; third, for each category, the mean trip production rate is calculated by adding together the number of trips generated by each household in that category and then dividing by the number of households belonging to that category; and finally, the new number of trips produced by each category zone is then estimated by multiplying the mean trip rate by the new number of households in that category. The new number of trips produced by the zone is then estimated by summing over all categories.

The most commonly used method (Wilson, 1974) to predict the number of households in each category in the future consists in, firstly, defining and fitting to the calibration data, probability distributions for income, car ownership and household structure, etc.; secondly, using these to build a joint probability function of belonging to a household type.

Table 2.7 presents an example (Ortúzar and Willumsen, 2001) of a category analysis model based on four household-size and three car-ownership levels. The table presents the trip rates for each household category. Generally the more people and cars in a household, the more trips would be made by the household.

Table 2.7 Trip rates per household calculated using category analysis

Household size	**Car ownership level**		
	0 car	**1 car**	**2+ cars**
1 person	0.12	0.94	-
2 or 3 persons	0.60	1.38	2.16
4 persons	1.14	1.74	2.60
5+ persons	1.02	1.69	2.60

It should be noted in this example that trip rate values decrease for 0 and 1 car-owning households when household size increase from 4 to 5 or more persons. This is contrary to intuition and may be due to insufficient data for these cells.

2.3.2.2 The advantages and disadvantages of the classical model

The disaggregate cross-classification method has the following advantages (Stopher and McDonald, 1983):

1. Cross-classification groupings are independent of the zone system of the study area;
2. No prior assumption about the shape of the relationships is required (i.e. they do not even have to be monotonic, let alone linear);
3. Relationships can differ in form from class to class of any one variable (e.g. the effect of household size changes for zero car-owning households can be different from that of one car-owning households); and
4. The cross-classification model does not permit extrapolation beyond its calibration classes, although the highest or lowest class of a variable may be open-ended.

But the model has several disadvantages, which are common to all traditional category analysis methods:

1. There is no statistical goodness-of-fit measure for the model, so that only aggregate closeness to the calibration data can be ascertained;
2. Unduly large samples are required; otherwise cell values will vary in reliability because of differences in the numbers of households being available for calibration at each one. It is suggested that at least 50 observations per cell are required to estimate the mean reliably.
3. The least-reliable cells are likely to be those at the extremes of the matrix, which may also be the most critical cells for forecasting;
4. There is no effective way to choose among variables for classification or to choose best grouping of a given variable, except to use an extensive trial-and-error procedure not usually considered feasible in practical studies;
5. The procedure suppresses information on variances within a cell;
6. It is particularly difficult to account for land use and accessibility factors in a cross-classification methodology, both because the number of cells quickly becomes too large and because these variables are particularly difficult to divide into meaningful ranges; and
7. It is very difficult to estimate the future number of households in each category (Ortúzar and Willumsen, 2001).

The fundamental problem of category analysis is the rigid structure that is imposed on the way in which the independent variables operate (Daly, 1997). There is no role for insight into the mechanisms affecting the numbers of trips made: in direct consequence the amount of data that is required is very large. Essentially, the method gives no explanation of trip generation.

Efforts have been made to overcome the shortcomings of the classic cross-classification model and these models are discussed in the following sections. Also see Section 6.3 in this research and Guevara and Thomas (2007) for further information.

2.3.2.3 Multiple classification analysis_1 (MCA_1)

An alternative methodology for calibrating cross-classification models is multiple classification analysis (MCA). The method is based on analysis of variance (ANOVA, Johnson and Leone, 1964), which provides a structured procedure for choosing among alternative independent variables and alternative groupings of the values of each independent variable (See Stopher and McDonald (1983) for details).

Consider a model with a continuous dependent variable (such as the trip rate) and two discrete independent variables, such as household size and car ownership. First, a grand mean can be estimated for the dependent variable over the entire sample of households. Second, group means can be estimated for each row and column of the cross-classification matrix; each of these can be expressed in turn as deviations from the grand mean. Observing the signs of the deviations, a cell value can be estimated by adding to the grand mean the row and column deviations corresponding to the cell. In this way some of the problems arising from too few observations on some cells can be compensated.

If interactions are present, then these deviations need to be adjusted to account for the interactive effects. This is done by taking a weighted mean for each of the group means of one independent variable over the groupings of the other independent variables, rather than a simple mean, which assumes that variation is random over the data in a group.

Because it is based on ANOVA, MCA also has statistical goodness-of-fit measures associated with it. Primarily these consist of an F statistic to assess the entire cross-classification scheme, a correlation ratio statistic for assessing the contribution of each classification variables (Stopher, 1975), and an R^2 for the entire cross-classification model. These measures provide a means to compare among alternative cross-classification schemes and to assess the fit to the calibration data.

In MCA, the cell values are no longer based only on the size of the data sample within a given cell; rather the cell values are based on grand mean derived from the entire data set, and on two or more class means which are derived from all data in each class relevant to the cell in question.

This procedure overcomes a number of the criticisms that have been made of the traditional cross-classification models. Specifically, the method permits a statistically based selection of variables for the cross-classification models, and also allows comparisons to be made between alternative groupings of any given variable.

Second, the method provides a statistically sound procedure for estimating cell means, which reduces the inherent variability of rates computed from different size sample of households and is capable of providing estimates for some cells where data may be lacking in the base data set (the use of this capability does reduce some of the available statistical information).

Third, there are goodness-of-fit statistics from all of these steps in the process that permit more specific comparisons to be made, good hypothesis-testing procedures to be followed, and results to be assessed in terms of the amount of the variability of the dependent variable that is captured in the model.

Finally, and more important, the method takes into account the interactions among the alternative independent variables, which have never been taken into account in standard cross-classification models.

It is mentioned that any phenomenon that has a nonlinear, and possibly discontinuous, functional form, and that is most readily related to variables that are categorical in nature, would be a prime candidate for this method.

Although the problem of not having large number of observations in each cell in the classical category analysis method has been overcome by using this analysis, Guevara and Thomas (2007) point out that it only corresponds to the OLS estimates of a model in which the number of observations by category is exactly the same (Glass and Stanley, 1986), which could hardly be true if surveyed households are, as usual, randomly sampled. The transgression of the assumption may lead to a significant overestimation of the future number of trips and a systematic bias in its socio-economic composition.

2.3.2.4 Multiple classification analysis_2 (MCA_2)

MCA_2 method is presented by Stopher and McDonald as a correction of MCA_1 for cases in which "interaction" among variables (which really means correlation among explanatory variables) is present. In practice, this method corresponds to a numerical correction that tries to consider the fact that the number of observations by category is not equal. This method was described in Ortúzar and Willumsen (1994) and Clark (1996).

MCA_2 differs from MCA_1 in that the average number of trips by household of each stratus is calculated as weighed averages. Guevara and Thomas (2007) indicate that MCA_2 method could improve the estimated coefficients, but hardly turn them into the OLS estimates. The net effect of this method should then be a partial improvement in the estimates.

2.3.2.5 Multiple classification analysis_3 (MCA_3)

The third modified method MCA_3 is the method of linear ordinary least squares (Guevara and Thomas, 2007), which is based on working estimation of the household trip rates by estimating least squares regressions where the independent variables are all dummy variable; one for each of the categories of the strata variables.

As this can be seen as an application of OLS, it is possible to use all the computational and statistical tools available for it in the literature. Particularly, if some distribution of the error is assumed, for example Normal, it would be possible to use statistical tests to identify variables for stratification or the size of each stratum.

To summarize, MCA_3 estimates corresponds to ANOVA (or OLS) estimates correctly calculated when MCA_1 is not applicable because the number of observations by category is not the same. On the other hand MCA_2 method can be seen as a numerical approximation of MCA_3 in cases where MCA_1 is not applicable.

2.3.2.6 Multiple classification analysis_4 (MCA_4)

In MCA_4, the trip rates are calculated as the average number of trips by household for each category. This method, also known as Category Analysis (Ortúzar and Willumsen, 1994), is equivalent to the estimation of an OLS model with dummy variables representing each category (see, for example, Goodman, 1973).

If the underlying model is linear, MCA_3 and MCA_4 are statistically equal. In that case they would both be consistent, but the first would be more efficient because it entails the estimation of fewer coefficients with the same information. Thus, MCA_3 should be chosen. If the underlying model is non-linear, MCA_4 would be consistent but MCA_3 will not, because the omitted attributes would be correlated with the observed linear attributes, causing endogeneity (Guevara and Ben-Akiva, 2006).

Guevara and Thomas (2007) conclude that MCA_1, the MCA method most widely used to estimate trip generations worldwide, should be discarded because it is supported by an assumption with very low probability of occurrence in the real world, the transgression of which may imply a severe bias in transportation systems modelling. The MCA_2 method should be seen as a numerical correction of MCA_1, which improves to some extent its results but is still weak, especially in modelling future scenarios. Thus, MCA_2 should also be discarded. The MCA_3 and MCA_4 methods are considered to be superior to the previous ones, in terms of precision and theoretical basis. The selection of one or another will depend on the case investigated, a decision that can be tested statistically.

In this thesis the MCA_1, MCA_2 and MCA_3 techniques were used to estimate trip generation models (see Chapter 6). Only results obtained from MCA_3 model however, were used in the final comparisons with the other methods.

2.3.2.7 The person-category approach

This approach was originally proposed by Supernak (1979) and it has been argued that, compared to household-based models, it has the following advantages (Supernak *et al.* 1983):

1. A person-level trip generation model is compatible with other components of the classical transport demand modelling system, which is based on trip-makers rather on households.
2. It allows a cross-classification scheme that uses all important variables and yields a manageable number of classes; this in turn allows class representation to be forecasted more easily.
3. The sample size required to develop a person-category model can be several times smaller than that required to estimate a household-category model.

4. Demographic changes can be more easily accounted for in a person-category model as, for example, certain key demographic variable (such as age) are virtually impossible to define at household level, and
5. Person categories are easier to forecast than household categories, as the latter require forecasts about household information and family size.

The major limitation that a person-category model may have is the difficulty of introducing household interaction effects and household money costs and money budgets into a person-based model. However, Supernak *et al.* (1983) argue that it is not clear how vital these considerations are and how they can be effectively incorporated even in a household-based model.

2.3.2.8 Generalized linear model with cross-classification

Said and Young (1990) review the disadvantages of the cross-classification analysis such as the variation in the reliability of trip rate values due to the variation in the number of households available in each cell for calibration and the loss of information when all households within each cell are treated similarly (Kassoff and Deutschman, 1969; Stopher and McDonald, 1983) and they applied generalized linear model framework (GLM) (Dobson, 1983; McCullagh and Nelder, 1983) for estimating work trip rates for households in Kuwait where there are great variations among households for the same nationality (for example, households vary in size between 1 and 50 persons) and between households of different nationality groups (e.g. three different groups) and in the difficulties that exist in the routine use of the cross-classification analysis approach.

The classical linear regression model of the form to be used is (Said and Young, 1990):

$$\mu_{ijk} = \mu + \alpha_i + \beta_{1i} x_{1j} + \beta_{2i} x_{2k}$$

Where

μ_{ijk} is the true mean of household work trip numbers in cell (i,j,k). μ is an overall mean, α_i is the effect of nationality type i and β_{1i} and β_{2i} are regression coefficients allowing for assumed linear effects of x_1 (household size) and x_2 (number of cars owned per household).

Said and Young (1990) point out two criticisms may be made of this classical linear regression model. First the range of its application is limited as sometimes the number of trips is relatively small but all must be positive. Second, the assumption of constant variance within cells is unlikely to be satisfied in practice with cells with the higher mean trip rates being likely to exhibit larger variances.

A possible way to overcome the first problem is to adopt a logarithmic model for the means with the following form (Said and Young, 1990):

$$\log \mu_{ijk} = \mu + \alpha_i + \beta_{1i} x_{1j} + \beta_{2i} x_{2k}$$

If variance heterogeneity among cells exists, the distribution is approximately Poisson. To obtain variance stabilization with Poisson observations, the square root transformation is used.

An alternative approach (i.e. MCA in Section 2.3.2.3) still within the GLM framework for handling grouped data is to use ANOVA models, see for example Dobson (1976). The mean trip rates of cell, (j, k), could be expressed as (Said and Young, 1990):

$$\overline{Y}_{jk} = m + m_j + m_k + m_{jk} + \sum_{l} \varepsilon_{jkl} / n_{jk}$$

Where

m is the grand mean of the true cell means;

m_j are deviations of true row means about the grand mean;

m_k are deviations of true column means about the grand mean; and

m_{jk} represent deviations from additivity of row and column effects about the grand mean.

The use of regression and ANOVA models with grouped data solves the problems related to the difficulty of forecasting household characteristics at the level of detail required for regression models with ungrouped data. However, the use of ANOVA does not take into account the quantitative nature of the two variables used in the example.

An illustrative GLM analysis was described in which trip rates of Kuwaiti households living in villas were utilized. Three regression-type models were fitted which are classical model for untransformed data, classical model with square root transformation of household trip data and a model that assumes a Poisson distribution of individual household trip rates within cross-classification cells with logarithmic link function for their means. The analysis showed that work trip rates of this household group are influenced by car ownership, household size, and the interactive effect of these two variables. It is concluded that the three models produce generally adequate fits; and only cross-classification cells with very low frequencies show significant discrepancies. The differences between the mean trip rate estimates from classical regression models for untransformed data and squared route transformed household trip data and this model are relatively small indicating that there is flexibility in the choice of a particular model for the data and the three models produce generally adequate fits. This analysis is very similar to that of Guevara and Thomas (2007) as discussed in Sections 2.3.2.3 - 2.3.2.6 above. These types of investigations, analysis and proposed approaches show that there are still needs and opportunities in the area of trip generation modelling.

Other applications of generalized linear models in trip generation include Rickard (1989) who describes an application of GLM to railway trips, Said *et al.* (1990), who extend this analysis to include qualitative variables and address the use of GLM with cross-classified household data using regression and ANOVA specifications, and Said *et al.* (1991) who apply this procedure to estimate mean trip rates of households in urban areas with a distinct mix of households group.

2.3.2.9 Regression analysis for household strata – another improvement to the classic model

This is a mixture of cross-classification and regression modelling of trip generation and it may be the most appropriate approach on certain occasions (Ortúzar and Willumsen, 2001). For example, in an area where the distribution of income is unequal it may be important to model impacts of policies on different income groups; therefore it may be necessary to model travel demand for each income group separately throughout the entire modelling process (see Hall *et al.*, 1987) for an example. A general problem of this approach is that some categories have rather few data points.

2.3.3 Discrete choice models

This section discusses discrete choice models that have been considered in trip generation modelling. An overview of these models is given in Table 2.8.

2.3.3.1 Nested-alternative-logit (ordered choice) model and ordered response model /ordered logit model

Sheffi (1979) developed a nested-alternative-logit model in a disaggregate, utility maximization framework for estimating choice probabilities among nested alternatives, i.e., the alternatives available to an individual randomly chosen from the population exhibit some internal choice related ranking: choice of a given alternative implies that all lower-ranked alternatives have been chosen as well. The utility model that corresponds to the choice among ordered integer alternatives is:

Table 2.8 Overview of trip generation – discrete choice models

Modelling Technique	Brief Description	Selected References
Nested-alternative-logit (ordered choice) model	To estimate choice probabilities among nested alternatives, i.e., the alternatives available to an individual exhibit some internal choice related ranking: choice of a given alternative implies that all lower-ranked options have been chosen as well.	Sheffi, 1979
Ordered response model /ordered logit model	A type of discrete choice model which maintains the ordinal nature in the dependent variable in situations where there are more than two responses.	Agyemang-Duah *et al.*, 1995; Agyemang-Duah and Hall, 1997; Schmöcker *et al.*, 2006
Negative binomial model / count data model	The negative binomial distribution is a generalization of the Poisson distribution.	Rickard, 1988; Barmby and Doornik (1989); Washington *et al.*, 2003; Jang, 2005; Guy, 1987;
Ordered probit model / mixed ordered probit model	Simple linear regression analysis would be inappropriate due to the large number of zero trips in the sample, and the difference between making 0 trips and 1 trip might be far more significant than a difference between 5 and 6 trips.	Schmöcker *et al.*, 2005; Long, 1997; Páez *et al.*, 2006; Jones, 1991; Duncan and Jones, 2000.
Frequency choice logit model - 'stop and go' trip generation model / the exponential model	To use a hierarchical structure representing an indefinite number of choices. At each hierarchical level, the choice is whether to make further journeys or stop at the present number (hence the name 'stop-go model').	Daly, 1997; Daly and Miller, 2006; Kouwenhoven *et al.*, 2006
Tobit model	It is a combination of regression and discrete choice models. It differentiates from regression model by the incorporation of truncated or censored dependent variables; it assumes that the dependent variable has a number of its values clustered at limiting value, usually zero.	Cotrus *et al.*, 2005; McDonald and Moffitt, 1980.

$$P_i = (1 - P_{i+1|i}) \cdot \prod_{k=1}^{i} \Pr(U_k \geq U_{k-1})$$
$$= (1 - P_{i+1|i}) \cdot \prod_{k=1}^{i} P_{k|k-1}$$

Where

P_i is the probability that alternative i is chosen;

U_k is the utility of alternative k to an individual randomly chosen from the population;

The model is a product of independent binary choices. Estimating each of the binary probabilities can be carried out through the use of a logit model. The use of a logit model is justified in the case of a binary choice problem since the difficulties arising from the independence of irrelevant alternatives (IIA) property of the multinomial logit (MNL) model do not exist in a binary model (Domencich and McFadden, 1975).

The essence of the model is in capturing the special correlation implied by the definition of nested alternatives and overcoming the difficulty from applying the MNL model to this problem: the IIA property.

This model was applied for estimating probabilities of non-work vehicle trip frequencies by elderly individuals. Sheffi (1979) points out that, in general, a trip generation model might not conform to this model of ordered nested alternatives in two aspects. First, there is a problem with using the entire household as the behavioural unit. Trips might be decided upon simultaneously and carried out by more than one person and the model cannot account for this phenomenon since the "one choice at a time" assumption is basic to its structure. The second difficulty is that multi-destination trip chains (in which a number of trips are combined in a single tour from the residence) cannot be accounted for in the model, and tours have to be counted as trips.

Agyemang-Duah *et al.* (1995) summarized the shortcomings of regression models and category analysis. They point out that the problems with the standard regression model include lack of any built-in upper limit to household trips as the values of explanatory values, such as household size and vehicle ownership, increase and the possibility of the regression models predicting negative trips. The difficulty with category analysis is the lack of any effective way to choose the best groupings of household characteristics and hence the best categories and also lack of inferential statistics and thus no way to assess the statistical significance of the explanatory variables in trip generation. Also both models treat the number of trips per household as a continuous dependent variable, but to develop a behavioural basis for trip generation, the dependent variable must be discrete rather than continuous. The possible solutions to this problem include to use Poisson regression models, which have been shown to be appropriate in applications to count data, especially when the count for some observations is small or 0 (Guy, 1987), and to use one of the family of discrete choice models, which are based on a probabilistic theory of choice among a finite set of options. Also there is a definite order to the trip-making decision. If a person makes two trips, that person also necessarily makes one trip. The ordered response model, which maintains the ordinal nature in the dependent variable in situations in which there are more than two responses, is adopted in their study of home-based shopping trips in the greater Toronto area.

The ordered response model has the following advantages over the standard regression models (Agyemang-Duah *et al.*, 1995): first, the property that choice probabilities are necessarily between 0 and 1 means that in prediction mode, the model cannot forecast negative or infinite trip. The second advantage is that the model predicts the whole distribution of the response levels unlike the standard regression approach, which will at best predict the mean of the dependent variables. And thirdly, the model offers a way to exploit the ordering of information.

Schmöcker *et al.* (2006) developed an ordered logit model to estimate the reduction of shopping trips a person makes in response to a congestion charge in

London and the levels of frequency reduction include slight decrease, decrease, significant decrease and very significant in shopping trip frequency.

2.3.3.2 Count data model / negative binomial model

Gourieroux *et al.* (1984) point out that the classical linear regression model (CLRM) is not appropriate for analyzing trip frequency, a discrete variable which can only take non-negative values for three reasons: firstly, the observation set is not that of the CLRM; secondly, the assumption of normality for the error term cannot be made; and thirdly, the predictions from CLRM could allow for impossible values.

Barmby and Doornik (1989) propose to model the number of trips, T_i, as a Poisson variable. This would have two distinct advantages. Firstly, the model could not predict a negative number of trips for certain values of the regressed variables. Secondly, the estimates of the model show underlying probabilities for actual number of trips, whereas the linear regression model only gives the expectation and variance of the number of trips, as implicitly the dependent variable would be a continuous variable. A Poisson model could be described as (Barmby and Doornik, 1989):

$$T_i \sim f(t_i) = \frac{e^{-\lambda_i} \lambda_i^{t_i}}{t_i!}; \quad t_i = 0, 1, 2, \ldots$$

$$E(T_i) = \lambda_i = \exp(X_i' \beta); \quad i = 1, \ldots, n$$

Where X_i is a vector of characteristics of the household which defines the mean of the distribution.

Barmby and Doornik (1989) indicate that a generalization of the Poisson distribution, the negative binomial distribution, could be a better choice to constructing a statistical model for trip frequency. The simple Poisson distribution assumes that the variance is constrained to be equal to the means, and this would be too restrictive for the data that are characterised by over-dispersion or under-dispersion, according to whether the variance is less than or greater than the mean. Also to generate the Poisson form for the probability function, the events must have occurred independently through time. The over (under) dispersion is circumvented, by modelling λ, the Poisson parameter, as a Gamma distribution, $h(\lambda)$. The new distribution of the observed number of trips can be obtained by mixing the distribution as:

$$g(t) = \int_0^\beta f(t;\lambda)h(\lambda)d\lambda$$

The resulting form of a negative binomial distribution is (Barmby and Doornik, 1989):

$$T_i \sim g(t_i) = \frac{\Gamma(t_i + \gamma_i)}{\Gamma(t_i + 1)\Gamma(\gamma_i)}\left(\frac{\gamma_i}{\gamma_i + \mu_i}\right)^{\gamma_i}\left(\frac{\mu_i}{\gamma_i + \mu_i}\right)^{t_i} \quad t_i = 0, 1, \ldots$$

The above model can be parameterized as (Cameron and Trivedi, 1986):

$$\mu_i = \exp(X_i'\beta)$$
$$\gamma = \frac{1}{\alpha}[\exp(X_i'\beta)]^k \quad \alpha > 0$$
$$E(T_i) = \exp(X_i'\beta)$$
$$VAR(T_i) = E(T_i) + \alpha[E(T_i)]^{2-k}$$

It can be seen now that the variance and mean are no longer constrained to be equal, and the parameters α and k will determine the form of the relationship between $E(T)$ and $VAR(T)$.

As there is a maximum number of trips in the record, an upper truncation is taken into account in estimating the Negative Binomial model. In general, if $T \sim f(t)$, truncation at T* will result in a truncated density of the following form (Barmby and Doornik, 1989; see also Cohen, 1961):

$$g(t) = \frac{f(t)}{\sum_{j=0}^{T^*} f(j)}; \quad t = 0, \ldots, T^*$$

Figure 2.1 shows a comparison of the predictions of the Negative Binomial model and the regression model in fitting observed data. For the first model, the implied relative frequencies are computed as the mean of the implied individual probabilities. Though both the normal frequency curve implied by the regression results, and the Negative Binomial lack the flexibility to pick up the bimodality in the observed data at trip level one, the latter tracks the relative frequencies of the observed data better than does the normal curve.

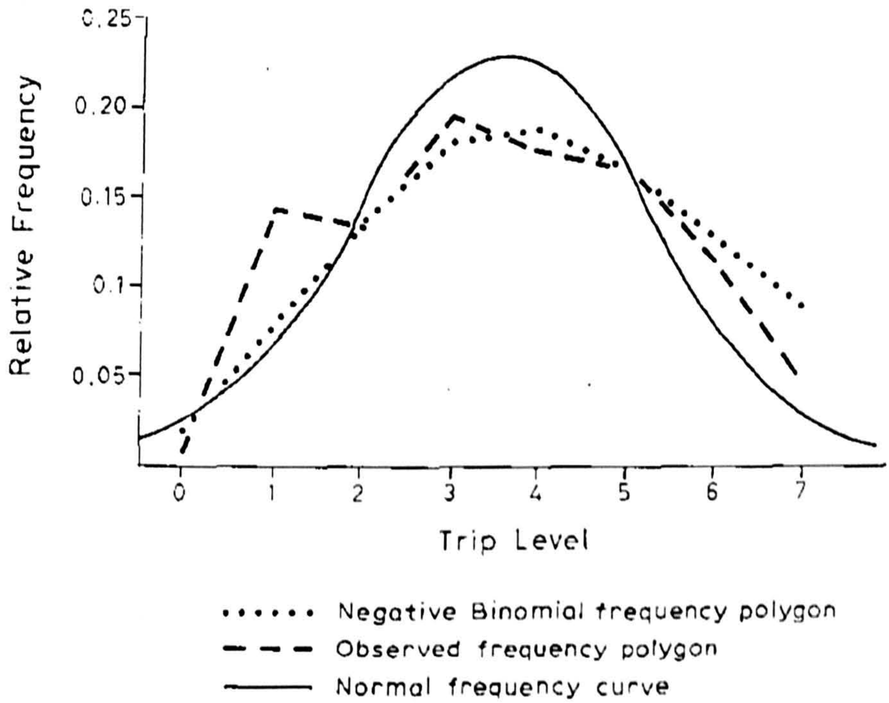


Figure 2.1 Comparison of predictions using the negative binomial model and the linear regression model

Source: Barmby and Doornik (1989)

Rickard (1988) compares the use of the Poisson distribution and the Negative Binomial to model long-distance rail trips as a generalized linear model (GLM) (McCullagh and Nelder, 1983). She finds the Negative Binomial distribution to be the more appropriate, and postulates that this is because the overall distribution is the sum of those of a number of sub-groups, each following its own Poisson distribution.

Jang (2005) also developed a Negative Binomial model and a modified count data model for trip generation to overcome over-dispersion of the Poisson model due to the assumption that the conditional variance of the dependent variable equals the conditional mean. Zero inflated models, which use a logistic mixing distribution to add to the zero mass of the probability density function (Cameron and Trivedi, 1990), are developed including the zero inflated Poisson (ZIP) model and the Zero inflated Negative Binomial (ZINB) model. These models allow for two sources of over-dispersion and extra zero resulting in individual heterogeneity in the positive set and are at work in determining the number of zero counts. The zero inflated model is a natural extension of the Poisson (or Negative Binomial) specification and is given by (Jang, 2005):

$$\Pr[y_i = 0] = \varphi_i + (1 - \varphi_i)e^{-\lambda_i}$$

$$\Pr[Y = y_i \mid Y > 0] = (1 - \varphi_i)\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \qquad y_i = 1,2,...$$

where $\ln \lambda_i = \beta' X_i$.

This distribution can also be interpreted as a finite mixture with a degenerate distribution whose mass is concentrated a zero. The proportion of zeros, φ_i, is added to the Poisson (or Negative Binomial) distribution, and other frequencies are reduced by a corresponding amount.

The zero inflated Negative Binomial model (ZINB) is selected as the optimal model through Vuong test and is used to calibrate non home based trips at household level and has shown improved variable estimation and decreased errors.

2.3.3.3 Ordered probit model /mixed ordered probit model

Schmöcker *et al.* (2005) used an ordered Probit model to estimate trip generation of elderly and disabled people in London taking the daily trip frequency as a latent variable. In the study, a model for total trips as well as models for specific trip purposes (namely work trips, shopping trips, personal business trips and recreational trips) were estimated. It is pointed out that simple linear regression analysis would be inappropriate due to the large number of zero trips in the sample, and the difference between making 0 trips and 1 trip might be far more significant than a difference between 5 and 6 trips. So an ordered Probit model was used as it provides a technique to estimate regression models for this sort of data. Alternatively, an ordered logit model would also be suitable. As the difference between a logit and probit model is in the assumption of the distribution of the error terms: a probit model assumes a normal distribution, whereas logit assumes a Gumbel distribution, Long (1997) concludes that the choice between logit and probit is mainly a matter of convenience as both models normally come to the same result. It is also mentioned that another method would be to use a Poisson or Negative Binomial model for count data (Washington *et al.*, 2003).

Páez *et al.* (2006) point out that the ordered probit model, by treating the number of trips (or the trip frequency) as a set of mutually exclusive and collectively exhaustive ordinal categorical variables, incorporates built-in upper and lower limits. In addition, the model provides a behavioural framework that directly links the number of trips to utility-based consumer and decision making theory.

In their elderly trip generation study, Páez *et al.* (2006) used a mixed ordered probit model, which is part of a family of models alternatively know as random coefficients, variance components, multilevel, or hierarchical models (see Jones, 1991; Duncan and Jones, 2000). The models of this family are characterized by their ability to accommodate random variation of the coefficients, which makes them suitable for exploring spatial variation in individual trip rates, and in particular the relationships between these rates and the factors that influence them (e.g. location and age). The use of the mixed ordered probit models allows for a mixture of variables at different levels of geography.

Despite the intuitive appeal of these models, an erstwhile constraint to their application was the complexity of the estimation procedures (Páez *et al.*, 2006). Hedeker and Gibbons (1994) have developed methods for estimating the mixed ordered probit model that use numerical quadrature techniques. As an alternative to this approach, Train (2003) provides a discussion of simulation techniques, whereby random numbers are generated to obtain a simulated log-likelihood function that can be maximized to obtain estimates.

2.3.3.4 Frequency choice logit model – 'stop and go' trip generation model / the exponential model

Daly (1997) indicated that a model with a logit form is suitable for predicting the total number of trips by first calculating the probability that each individual will choose to make a trip. The total travel volume is then obtained by multiplying the number of individuals of each type by their probabilities of making a trip. The logit model represents the choice of each individual whether or not to make a trip, and therefore it is particularly suited to dealing with disaggregate data.

To model higher trip frequencies, Daly (1997) proposes the use of a hierarchical structure representing an indefinite number of choices. At each hierarchical level, the choice is whether to make further journeys or stop at the present number (hence the name 'stop-go model'). A separate model is found preferable to model the first choice, as possibly strong difference exists between the 0 and 1+ choice where the remaining choices could then be modelled. Also because normally there is little data on travellers making multiple journeys, it is necessary to model the remaining choices with a single 'stop-go' model (i.e. which predicts the same probability of stopping at every level of the hierarchy). If the probability of an individual n making at least one journey is p_n, and then the probability of making a further journey at each stage is q_n, then the expected number of journeys is (Daly, 1997):

$$D_n = \frac{p_n}{1-q_n}$$

When the 0/1+ model and the 1/2+, 2/3+ models are the same, the stop-go reduces to a geometric model with parameter $1-p_n$.

Suppose the 'stop' alternative for an individual has utility V_n^0, which may incorporate all non-accessibility (e.g. socio-economic) effects and the 'go' alternative has a utility of $\lambda \cdot V_n^*$, a multiple of the logsum (Ortúzar and Willumsen, 2001). Then the probability of travel is (Daly and Miller, 2006):

$$p_n = q_n = \frac{\exp(\lambda \cdot V_n^*)}{\exp(\lambda \cdot V_n^*) + \exp(V_n^0)}$$

And then

$$D_n = \exp(\lambda \cdot V_n^* - V_n^0) = \alpha \cdot \exp(\lambda \cdot V_n^*)$$

Where α does not depend on accessibility. The exponential model has the same expectation of the forecast number of trips as a stop-go model in which the two model components are identical. The model can be considered to be an implementation for forecasting of the simplified stop-go or geometric model and it has a secure basis in utility theory.

Daly (1997) also investigated an accessibility measure by calculating the logsum of destination choice in an integrated trip generation, mode and destination choice model using the hierarchical structure. A number of applications of this approach have been developed including Cambridge Systematics Europe (1981), HCG and TOI (1990) and Cohn et al (1996).

Daly and Miller (2006) compare this derivation of exponential trip generation with a Poisson model and note that while both models give rise to a mean trip rate that is an exponential of the logsum, the probability distributions for the actual number of trips made by an individual are very different. The mode of the Poisson distribution occurs around the mean, whereas the mode of the geometric distribution is always zero. The geometric distributions also have a larger variance. The difference of the two models is less for lower trip rates. For the Poisson model, the link to utility theory has yet to be established. Larson (2003) found a corresponding problem with the Poisson model in some of his tests on Norwegian data, where he found it necessary to introduce an initial binary choice model for the 0/1+ choice.

Daly and Miller (2006) point out that the geometric model cannot be recommended for trip generation in urban and regional contexts. As behaviourally, the decision whether to travel at all (0, 1+ trips) is usually found to be quite different from the decision whether to make a further trip (1/2+, 2/3+ etc.). For long-distance travel, however, a single model is acceptable. The exponential model is an exact implementation of the geometric model, where each step is modelled by a binary choice, and it can represent the actual behaviour accurately.

It is also noted that the exponential model is different from a constant-elasticity model. While the difficulty in the elasticity model is to define a zero for generalized cost, i.e. defining exactly which components should and should not be included; this difficulty does not arise in the exponential model.

2.3.3.5 Tobit model

Cotrus *et al.* (2005) explored the use of regression and Tobit models in trip generation in two metropolitan areas and two time periods in Israel, and investigated their spatial and temporal transferability. Hald (1949) first presented the model that, in its final form, is called the Tobit model (Tobin, 1958). Tobit models differentiate from regression models by the incorporation of truncated or censored dependent variables. Tobit analysis assumes that the dependent variable has a number of its values clustered at a limiting value, usually zero. Tobit models can be presented as discrete/continuous models that first make a discrete choice of passing the threshold and second, if passed, a continuous choice regarding the value above the threshold. As shown by McDonald and Moffitt (1980), Tobit analysis can be used to determine the changes in the value of the dependent variable if it is above the limit, as well as changes in the probability of being above the limit.

Cotrus *et al.* (2005) indicate that Tobit models tend to present the mechanism of trip generation more realistically, capturing and estimating (partially) non-travellers. As a combination of regression and discrete choice models, the Tobit model may be more suitable for implementation in trip generation modelling than discrete choice and regression models, particularly because Tobit is better formulated to differentiate non-travellers from travellers. However, non-travellers are underestimated which may be partly due to the fact that the best Tobit model has not been obtained.

2.3.4 Other trip generation techniques

Other trip generation approaches and modls include This section looks at other techniques that have not been included in the above three categories and an overview of these models is given in Table 2.9.

2.3.4.1 Growth factor modelling

Growth factor modelling is one of the techniques that have been proposed to model trip generation which may be applied to predict the future number of journeys. Its basic equation is $T_i = F_i t_i$, where T_i and t_i are future and current trips in zone *i* respectively, and F_i is a growth factor which is related to variables such as population, income and car ownership. The method is very crude. It is therefore only used in practice to predict the future number of *external* trips to an area; this is because there are not too many in the first place (so errors cannot be too large) and also because there are no simple ways to predict them.

Table 2.9 Overview of other trip generation modelling techniques

Modelling Technique	Brief Description	References
Growth factor modelling	To use a growth factor rate to predict the future number of journeys.	Ortúzar and Willumsen, 2001
Criterion-based segmentation modelling tool – CHAID (Chi-squared Automatic Interaction Detection)	Presented in the form of a tree, each final node represents a group of homogenous households concerning daily trip making; Allows to identify significant interaction effects between categories of explanatory variables.	Kass, 1980; Strambi and Bilt, 1998; Bilt, 1997
Hierarchical tree-based regression (HTBR) model and iteratively specified tree-based regression (ISTBR) model	HTBR is a tree-based method more adept at treating multicollinearity among variables; interactions between independent variables are also less troublesome. Iteratively specified tree-based regression (ISTBR) combines desirable properties of OLS with HTBR.	Washington and Wolf, 1997; Washington, 2000
Artificial neural networks	Computing system made up of number of simple highly interconnected processing elements that process information by dynamic state response to external inputs.	Caudill, 1987; Fahgri and Hua, 1992; Tillema *et al.*, 2004
Approaches to model trip chaining and trip generation	Trip generation and trip chaining integrating concepts from activity-based analysis. Structure of the model system is recursive, depicting a sequential decision-making mechanism assuming that the number of discretionary trips is dependent on the number of mandatory trips.	Goulias *et al.*, 1990
Direct demand modelling	The model subsumes trip generation, distribution and mode choice.	Kraft, 1968
Dynamic trip generation model	Model examines dynamic characteristics of a household trip generation, i.e., the correlation of trip making over time. The generalized method of moments procedure is used.	Meurs, 1990; Anderson and Malave, 2005

Activity-based trip generation model	Model developed to estimate trip productions from the analysis of complete travel/activity patterns; classifies travel patterns with respect to activity, spatial, and temporal characteristics.	Wang (1997)

2.3.4.2 Criterion-based segmentation modelling tool - CHAID

Strambi and Bilt (1998) identify the difficulties with the applications of conventional trip generation models which are typical of segmentation problems: identification and categorization of explanatory variables and of the interactions between them and explore the use of CHAID (Chi-squared Automatic Interaction Detection), to analyze household trip generation rates. CHAID is a criterion-based segmentation modelling tool originally developed by Kass (1980) and CHAID models are presented in the form of a tree, each final node representing a group of homogenous households concerning daily trip making. CHAID can automatically identify significant interaction effects between categories of predictor/explanatory variables which provide the opportunity to avoid flaws in model specification, in particular, biases resulting from omitting relevant interactions.

An application to data from an origin-destination survey for Sao Paulo produced interesting results (Bilt, 1997), in agreement with theoretical expectations and amenable to interpretation based on the likely activity-travel patterns of each group of households generated by the technique. CHAID can be used as an exploratory technique for aiding model development or as a model itself. The application of CHAID as a modelling tool requiring a highly disaggregate projection of the population may become possible considering the advances in methods for the generation of synthetic populations.

2.3.4.3 Hierarchical tree-based regression (HTBR) model and iteratively specified tree-based regression (ISTBR) model

Washington and Wolf (1997) explored the use of a hierarchical tree-based regression (HTBR) model in trip generation and compared it to ordinary least squares regression. HTBR is one of the two types of tree-based methods: classification trees, which are designed to partition data, based on the discrete nature of categorical or class data and regression trees, to partition (regress) data on the basis of continuous response data. It is sometimes referred to as classification and regression trees, or CARTs (Breiman *et al.*, 1984).

HBTR is more adept at treating multicollinearity among variables because it handles them automatically within the tree construction process (Washington and Wolf, 1997). Interactions between independent variables are also less trouble-some in HBTR. In the estimation of an ordinary least squares (OLS) regression model, which derives its name from the criterion used to draw the best fit regression line: a line such that the sum of the squared deviations of the distances of all the points to the line is minimized (Garson, 2006), the modeller must specify the correct functional interaction between variables to account for their synergistic effect, where in HTBR interactions are handled automatically. HBTR methods treat non-additive and non-linear behaviour better than do OLS methods. HTBR is superior to OLS regression as discrete variables take on significantly more than two levels. Washington and Wolf (1997) pointed out that OLS regression, whose estimated coefficients are easily interpretable, is generally a more intuitive tool than HBTR for explaining phenomenon. However, theory is better developed for OLS regression than for HBTR, and therefore, HTBR's shortcomings include a lack of formal methods for: analysis of residuals and outliers, dealing with omitted influential independence variables, efficiency, bias, consistency of estimated model parameters, finding statistically significant tree depth, testing of working hypotheses, and model selection and refinement criteria.

Washington (2000) presents an iterative modelling method that combines some desirable properties of OLS with hierarchical tree-based regression (HBTR).

This combined approach, named iteratively specified tree-based regression (ISTBR), is shown to provide insights into data structure provided by hierarchical tree-based regression, while retaining the desirable parametric properties of OLS. ISTBR helps the analyst to identify potentially important interactions, nonlinearities, and non-additive behaviour between the response variable and the predictor variables. Specifying linear regression models using the ISTBR modelling approach differs from traditional linear modelling in that the modelling results are driven by data – exposing second- and higher-order interactions, nonlinearities, and non-additive behaviour between variables. Best subsets and stepwise regression procedures, in contrast, rely on a priori identification of important interactions and specifications of a functional form of the independent variables. ISTBR equips the modeller with improved tools for exploring and identifying alternative model specifications and affords the analyst insight into systematic patterns in data that might otherwise go undetected.

2.3.4.4 Artificial neural networks

Artificial neural networks (ANNs) is a computing system made up of a number of simple, highly interconnected processing elements that process information by dynamic state response to external inputs (Caudill, 1987). Fahgri and Hua (1992) presented a demonstration of the applicability of ANNs in zonal trip generation forecasting, using the ADALINE (i.e., Adaptive Linear Element) and the back-propagation ANN models. ADALINE is a combinatorial logical circuit that accepts several inputs and produces one output, operating with a least mean square error-correcting learning rule. Back propagation has at least one hidden layer and during the learning process, the error information is propagated back from the output layer through the network to the first hidden layer. Back propagation is a powerful technique for constructing nonlinear transfer functions between a number of continuously valued inputs and one or more continuously valued outputs. One of the obvious differences between ADALINE and the regression method is the handling of the optimization of the weights and the coefficients. The regression method pursues the coefficients that will produce the minimum error on the surveyed data, which can be considered the training data sets for the ADALINE model. The training of ADALINE pursues the best value of the weights that will allow the model to obtain good results on the testing data sets, but not on the training data sets. Even if a set of weights will allow the model to perform well on the training data sets, unless those values of the weights will allow the model to reach the approximate error minimum on the testing data sets, those weights are not considered good. The results obtained by ANNs techniques outperformed those obtained by conventional regression models.

Tillema *et al.* (2004) investigate modelling trip generation using neural networks to see whether neural networks can out-perform traditional regression methods or not with the smallest data sets. The neural networks are tested in two situations with regards to the data availability; (i) data is scarce; and (ii) data is sufficiently at hand. The question of whether neural networks can be used in trip generation modelling is answered positively. However, neural networks do not overall out-perform classical regression models in situations where data is scarce. The advantages over regression models are negligible.

2.3.4.5 Approaches to model trip chaining and trip generation

Goulias *et al.* (1990) developed a model system of trip generation and trip chaining by integrating concepts from activity-based analysis. The structure of the model system is recursive, depicting a sequential decision-making mechanism assuming that the number of discretionary trips is dependent on the number of mandatory trips.

First, the number of trips for mandatory activities can be expressed as a linear function of exogenous variables alone (i.e. income and structure of the household). Second, the number of trips for discretionary activities may be represented by a linear function of the number of mandatory trips as well as exogenous variables. The statistical significance of each variable can be used to identify possible causal links between the exogenous and endogenous variables. Finally, the number of trip chains is formulated as a linear function of the number of trip by purpose. And then, the number of trip chains can be converted into home-based and non-home-based trip rates based on simple identity.

One advantage of this method is that it reflects a possible multistage decision-making process that may be followed by households when making trips. Another important property of the model system is that it explicitly considers the interface among trips made for different purposes, thus integrating home-based and non-home-based trip generation in a coherent manner. However, the model system needs further development to be a component of a comprehensive procedure of travel demand forecasting. For example, the model system cannot be used to predict the sequence in which trips for different purposes are linked. Consequently, it is unable to estimate home-based and non-home-based trip generation by purpose.

2.3.4.6 Direct demand modelling

The conventional sequential 4-step model classic methodology requires the estimation of relatively well-defined sub-models (Ortúzar and Willumsen, 2001). An alternative approach is to develop directly a model subsuming trip generation, distribution and mode choice. This is very attractive as it avoids some of the pitfalls of the sequential approach. There are two types of direct demand models: purely direct, which use a single estimated equation to relate travel demand directly to mode, journey and personal attributes; and a quasi-direct approach which employs a form of separability between mode split and total (O-D) travel demand.

The earliest forms of direct demand models were of the multiplicative kind. The SARC (Kraft, 1968) model, for example, estimates demand as a multiplicative function of activity and socioeconomic variables for each zone pair and level-of-service attributes of the model serving them. The model is very attractive in principle, as it handles generation, distribution and modal split simultaneously, including attributes of competing modes and a wide range of level of service and activity variables. Its main problem is the large number of parameters needed to cash in on these advantages.

The approach can further be enhanced to combine generation (i.e. choice of frequency), distribution (i.e. choice of destination) and mode choice in one combined model. It is possible to use the nested logit model structure for this modelling. The direct demand model, as it is calibrated simultaneously for these sub-models, would not suffer from the problems of having to cope with the errors in trip-end totals and those generated by poorly estimated intra-zonal trips.

Recently, the logit frequency model is re-introduced in the direct demand models, which combines generation (i.e. choice of frequency), distribution (i.e. choice of destination) and mode choice in one combined (i.e. nested) logit model; examples include Daly and others in Europe (Daly, 1997) and Iglesias et. al (2008) in Chile. In the latter correct accessibility measures were derived for intercity trip generation.

2.3.4.7 Dynamic trip generation models

Meurs (1990) reviewed the problems with conventional models such as the omission of variables in the models when they are correlated with the included explanatory variables and the models are static when based upon cross-section data, and examined the dynamic characteristics of a household trip generation, i.e., the correlation of trip making over time. The basic models considered in the research are the serial correlation and the state-dependence models. As part of the correlation of the error-terms over time is due to time-invariance of unobserved heterogeneity, unobserved heterogeneity is taken into account using random effects. The generalized method of moments procedure is used for estimation of the models: it is asymptotically efficient and does not require assumptions about the initial conditions. It is concluded that trip making in total and by transit was best described using state-dependence models; and trip making by car by a model with lagged exogenous variables.

Anderson and Malave (2005) developed a zonal time-dependent dynamic trip generation model for a medium-sized urban community, which is necessary to supply data to support the dynamic traffic-assignment models. The results show that a 15-min model performs better, with model predictions closer to the average number of trips being made from the zone, than a 5-min model, because of the aggregation involved. However, both models can predict time-dependent trip making with the community.

2.3.4.8 An activity-based trip generation model

Wang (1997) developed an activity-based trip generation model to address shortcomings of the conventional trip-based approach such as problems with conventional generation models resulted from a fundamental incapability to address temporal and spatial characteristics of activities and the trips which they generated, and the sequencing and scheduling of trips and activities, and interactions between household members, are ignored in the standard model. The model was developed to estimate trip productions from the analysis of complete travel/activity patterns and it classifies travel patterns with respect to activity, spatial, and temporal characteristics. The results obtained show that there is temporal stability of activity patterns in similar life cycle groups in the 1985 and 1994 Portland test data and it is concluded that patterns are a viable structure on which to base future forecasts.

2.3.5 Temporal and spatial transferability of the models

Transferability is an issue in two dimensions, space and time (Agyemang-Duah and Hall, 1997). Temporal transfer occurs when a model estimated in one time period in a specific geographic context is used in future forecasting in the same area and spatial transfer involves applying a model estimated on data from one particular spatial entity to another geographic context. Transferability can help to reduce substantially the need for costly full scale transportation surveys in different metropolitan areas or different areas in the same metropolitan area, and thus to allow for cost-effective analyses of transportation plans and policies. The following summary is based on a discussion by Ortúzar and Willumsen (2001):

Transport models, in general, are developed to assist in the formulation and evaluation of transport plans and projects. While on some occasions use has been made of descriptive statistics for examining travel trends, most developments have used cross-sectional data to express the amount of travel in terms of explanatory factors. A key assumption of this approach is that the model parameters will remain constant (or stable) between base and design years (Ortúzar and Willumsen, 2001). A number of researchers have examined the assumption and found the transferability of models in time (i.e., their temporal stability) satisfactory (see Downes and Gyenes 1976; Karasmaa and Pursula 1997) when trips by all modes are considered together. Unsatisfactory results, however, were obtained in other studies (see Doubleday 1977; Copley and Lowe 1981).

Geographic transferability should be seen as an important attribute of any travel demand model for the following reasons (Ortúzar and Willumsen, 2001):

1. It would suggest the existence of certain repeatable regularities in travel behaviour which can be picked up and reflected by the model;
2. It would indicate a higher probability that temporal stability also exists; this is essential for any forecasting model; and
3. It may allow reducing substantially the need for costly full-scale transportation survey on different metropolitan areas.

Not all travel characteristics can be transferable between different areas or cities such as the average work trip duration should be a function of area size, shape and the distributions of workplaces and residential zones over space. However, trips reflect the need for individuals' participation in various activities outside home and if trip rates are related to homogeneous groups of people, they can be expected to remain stable and geographically transferable (Ortúzar and Willumsen, 2001).

A number of studies found spatial transferability of models satisfactory (Wilmot 1995; Supernak, 1979, 1981). Supernak (1979, 1981) reported the successful transferability of the personal-category trip generation model, both for Polish and American conditions. Rose and Koppelman (1984) examined the transferability of a discrete choice trip generation model, allowing for adjustment of modal constants using local data, and concluded that context similarity appeared to be important determinant of model transferability; also, because their results showed considerable variability, they caution that great care must be taken to ensure that the transferred model is usable in the new context.

Agyemang-Duah and Hall (1997) investigate the performance of a directly transferred ordered response model (without updating the transferred coefficients) and assess the effectiveness of a technique for revising the constant terms and scalars in the model by using small-sample data from the region to which the model is to be applied. The analysis focuses on shopping trip generation in Metropolitan Toronto. The results of this spatial transferability analysis show that a directly transferred ordered response model performs reasonably well in predicting the aggregate shares in the application (new) context. Revising the constant terms and the scalars in the model substantially improves the predictive ability of the transferred model.

On the other hand, Smith and Cleveland (1976) and Daor (1981) found spatial transferability unsatisfactory. Cotrus *et al.* (2005) indicate that in order for trip generation models to be transferable they need to account for variables not included in the current models: income, land use and spatial structure, the economy, the transportation system and accessibility, more detailed socio-economic and life cycle variables. If we could estimate a perfect disaggregate model accounting for all factors that affect trip generation and with appropriate segmentation, it would likely be transferable. With this data lacking, models are not transferable, because unobserved variables affect coefficients of observed variables with which they are correlated. They point out that household survey conducted on a regular basis will be more useful if the design stays constant. Differences in the structure, variables, range, investigation period, definition of the variables, and database structure affect the transferability of the estimated models.

2.4 THE GAPS IN CURRENT TRIP GENERATION TECHNIQUES

As discussed above, although in regression analysis there are statistical tests for the goodness of fit of the models, the assumption of linearity of each of independent variables with the dependent variable is restrictive. Furthermore, the lack of built-in upper limits for trip rates as the values of the explanatory variables increase, and the possibility of predicting negative trips, both mean that regression models are not wholly suitable for trip generation analyses (Agyemang-Duah and Hall, 1997; Páez *et al.*, 2006). The assumption that the number of trips is approximately continuous can be questioned when typical values of the number of trips are relatively low (Páez *et al.*, 2006). The link between number of trips and covariates in a linear regression, while it may be based on hypothetical ideas about the process of trip generation, lacks a behavioural justification such as supported by the theory of random utility (e.g. Ben-Akiva and Lerman, 1985). A number of research investigations have been carried out which demonstrate the importance of including behaviour data and modelling approaches for the prediction of trip generation. For example Vickerman and Barmby (1985) investigated the use of behaviour approach and a choice model to investigate trip generation. Bhat (1999) investigated the use of repeated choice observations models in analysing evening commuting trips. Golob (2000) developed a simultaneous model of household activity participation and trip chaining. Wallace et al. (2000) investigated the effects of travellers and trip characteristics on trip chaining, with implications for transportation demand management strategies and Misra et al. (2003) used a continuous time representation and modelling framework for the analysis of nonworker activity-travel pattern.

Other forms of the model include the Poisson distribution which assumes that the variance is constrained to be equal to the means; this would be too restrictive for the data that are characterised by over-dispersion or under-dispersion. Also to generate the Poisson form for the probability function, the events must have occurred independently through time. In Tobit models, non-travellers can be underestimated. Ordered probit model is also suitable for modelling trip generation, however, the complexity of calculations of the model makes it not very attractive. Alternatively, classical category analysis, is undermined by the large sample sizes required to calibrate reliable trip rates as well as the absence of statistical tests for the overall goodness of fit of the models. MCA methods provide further developments of the principles of category analysis despite the heavy reliance on large amount of data. Logistic regression techniques have been investigated in this study for simplicity and ease of estimation. Moreover, insufficient empirical evidence exists to confirm that any one model form is superior to another in trip generation modelling.

Logistic models have been widely used to model travel behaviour choices such as mode, departure time, destination, route and residential location choice and commute behaviour. For examples, Bhat (1998b) studied mode and departure time for urban shopping trips and Wen and Koppelman (2001) investigate inter-city travel mode choice. Small (1982) modelled the arrival time of car commuters and Abkowitz (1981) modelled departure time choice for the commute to work. Freedman and Kern (1997, investigated workplace and residential location decisions and Sermons and Koppelman (2001) also investigated residential location choice and commute behaviour. Finally, Hensher and Greene (2003) analysed urban commute travel route choice and Rizzi and Ortúzar (2006) examined interurban route choice. For more discussions of the logistic models, see Chapter 4.

However, very limited applications of logistic regression in trip generation modelling have been reported (see for example Daly, 1997). Logistic regression can be used to model trip generation using binary logit models (whether or not an individual will make a trip), or multinomial logit models (probability of making {0, 1, 2 or more trips}, or probability of making {infrequent, frequent, very frequent trips}, etc.). This way, one can investigate the frequency of trips combined with the number of trips made by each individual or household. Logistic regression overcomes the restrictive assumption of ordinary least squares regression (Garson, 2002) that is the assumptions of linearity between the dependent and independent variables. This technique can be used to model relationships between the response variables which are binary or categorical, with more than two categories and several explanatory variables which may be categorical or continuous.

Utilising discrete choice framework to model trip generation, the number of trips (or the trip frequency) are treated as a set of mutually exclusive and collectively exhaustive categorical variables, incorporating built-in upper and lower limits. The estimates of the model show underlying probabilities for actual number of trips, which cannot be a negative number, whereas the linear regression model only gives the expectation and variance of the number of trips, as the dependent variable would be a continuous variable. In addition, the model provides a behavioural framework that directly links the number of trips to utility-based consumer and decision making theory. This research considers investigates the development of trip generation models using logistic regression analysis and also incorporating policy sensitive measures such as road user charging and parking fees.

Accessibility of the transport system has been investigated. A number of researchers have calibrated functions to represent transport accessibility (for example see Leek and Huzayyin 1979). However, most of the investigated functions included mainly factors which are representing the level of service of the transport system such as frequency of buses, travel time distances. Transport policies and their impacts on the accessibility have not been much investigated at the trip generation stage. Impacts of transport policies however have been investigated at other travel choice decisions such as mode, route and destination choices. A major disadvantage of this is that the changes to the network are basically assumed to not have any effects on trip production and trip attractions. This assumption may hold for compulsory trips, but it may not be so in case of discretionary trips.

2.5 SUMMARY

This chapter discusses some basic definitions in the trip generation modelling. The main factors which affect trip generation have been reviewed. These include various socio-economic characteristics of the trip makers residing in the area, the physical characteristics of the area, and transport infrastructure and transport services / accessibility (this is discussed in Section 3.4). Also a discussion of the approaches of data aggregation in trip generation modelling is presented.

Section two reviews the two most commonly used techniques of trip generation modelling (i.e. linear regression analysis and category analysis). For regression analysis, it covers the assumptions, statistics and models development, as well as the comparison of the effects at different types of aggregation (zonal, household and personal) and its advantages and disadvantages. For category analysis, the classic model and its advantages and disadvantages, the improvements and personal-category model are discussed. The new class of MCA methods which overcome a number of limitations of the classical MCA model have also been overviewed. Also, the temporal and geographic transferability of the trip generation models is discussed and other trip generation techniques that have appeared in the literature are briefly described.

Finally, the gaps in current trip generation techniques and the main aim of this study are presented. We briefly introduce logistic analysis and its applications in travel choice models and their potential use in trip generation modelling.

CHAPTER 3 MODELLING ACCESSIBILITY IN TRIP GENERATION MODELS: LITERATURE REVIEW

3.1 INTRODUCTION

This chapter reviews the main approaches for modelling transport accessibility and its application in trip generation models. Section 3.2 discusses the concept of accessibility and the factors that influence it. Section 3.3 reviews the different approaches to accessibility measures in the literature. In Section 3.4, a discussion of how transport accessibility has been included in the trip distribution, modal split and trip assignment stages of the classic four stage transport model is presented, while Section 3.5 reviews how different accessibility measures have been incorporated into trip generation models. Section 3.6 gives a general discussion of accessibility and its appropriateness for inclusion in trip generation modelling. Finally, Section 3.7 discusses the gaps in earlier research and approaches for treating transport accessibility measures.

3.2 CONCEPT OF ACCESSIBILITY

Accessibility is a concept used in a number of fields such as transport planning, urban planning, geography and marketing. Typically, accessibility refers to the "ease" with which desired destinations may be reached and is frequently measured as a function of the available opportunities moderated by some measure of impedance (Niemeier, 1997). Opportunities may be expressed as employment levels and retail or non-retail square footage depending on the application; impedance is usually denoted by travel time or possibly distance.

The types of opportunities depend upon whether origins or destinations are being considered (Halden *et al.* 2000). Origin accessibility considers the opportunities available to an individual or a business, thus the opportunity term is based upon the land use at alternative destinations. Destination accessibility considers the catchments for a destination, thus the opportunity term is based upon the land uses (i.e. employment, education, health, shopping, etc.) and the type of person or traveller at alternative origins.

Halden *et al.* (2000) point out that all accessibility measures relate to specific locations, origin or destination, and include representation of defined opportunities and a separation element between these opportunities and the locations. Generally, accessibility measures consist of four different components: land-use component, temporal and individual components, and transport component (Geurs and Ritsema van Eck, 2001).

3.2.1 The land-use component

The distribution of opportunities in space influences the level of accessibility (Geurs and Ritsema van Eck, 2001). For example, if all jobs and dwelling are equally distributed over a certain area or clustered in the (city) centre of a given area, there will be different impacts on people's level of job accessibility. The land-use component of accessibility can be split into two elements: the spatial distribution of supplied destinations and their characteristics (such as location of offices, capacity) and the spatial distribution of the demand for activities and their characteristics (such as locations of dwellings). Both the distributions of supplied opportunities and the demand for opportunities can influence accessibility.

The types of opportunities include (Halden *et al.* 2000): (1) employment, education and training, e.g. employment locations, jobs centres and colleges, etc.; (2) health and social, e.g. hospitals and social security offices, etc.; and (3) shopping and leisure, e.g. shopping centres and cinemas, etc.

In handling the land-use component of accessibility the demarcation of the research area must be decided. Halden *et al.* (2000) indicate that the extent of the zoning system and the level of detail will depend upon the policy issues being examined and how much effort can be afforded on the analysis. For example, strategic transport improvements require a wide geographical coverage and a fairly coarse zoning system may be adequate.

3.2.2 The temporal component

The temporal component of accessibility involves the availability of activities at different times of day or weeks, seasons, years, etc. and the times in which individuals participate in specific activities (Geurs and Ritsema van Eck, 2001). It originates in the space-time studies of the urban activity system from Gägerstrand (1970) and Chapin (1974). The time component and land-use component of accessibility are interdependent because individuals can only be at one location at a given time and travel consumes time. In potential accessibility measures, the temporal component is usually implicitly dealt with by varying the transport component throughout the day.

3.2.3 The individual component

The characteristics of individuals play an important role in the level of access to social and economic opportunities. Three groupings of determinants are often identified: needs, abilities and opportunities (Vlek and Steg, 1996). Geurs and Ritsema van Eck (2001) summarize that: (1) needs for travel and access to opportunities depend on their characteristics, such as age, income, and household situation; (2) abilities of people are related to level of physical capacity (e.g. cognitive, intellectual or physical disabilities) and to specific skills needed to access a transport mode (e.g. qualifications to drive a car); and (3) opportunities of people are related to income and travel budgets. In general, the individual component of accessibility is incorporated into accessibility measures by stratifying the population according to a selection of relevant characteristics (e.g. age, gender).

3.2.4 The transport component

In general, the transport component of accessibility consists of three elements (Geurs and Ritsema van Eck, 2001): (1) the supply of infrastructure, its location and characteristics (e.g. maximum travel speed, public transport timetables, travel costs); (2) the demand for passenger and freight travel; and (3) the characteristics of resulting infrastructure use, i.e. the outcome of the confrontation between infrastructure supply and travel demand, resulting in the spatial distribution of road traffic, and the travel time, costs and effort to reach a destination.

Deterrence functions, including barriers to accessibility, can be measured as time, travel cost, distance, or generalised cost/time (Halden *et al.* 2000). They aim to represent each factor or barrier perceived by each population group. This must include the relative deterrent effect of different types of travel and the costs associated with each, including issues such as the greater deterrent effect of time waiting for a vehicle when compared with the same time spent travelling in a vehicle. It is usually helpful to look separately at the deterrence functions for car available and non-car available trips. This is because many trips involve a combination of several modes and for non-car available trips the car options need to be excluded from the calculation.

The deterrence factors affecting travel (or access without travel) for people to activities include (Halden *et al.*, 2000): (1) transport availability, physical accessibility, affordability and acceptability, etc.; (2) other extraneous factors such as topography, severance, crime and fear of crime; and (3) information and personal knowledge, skills, willing to travel, etc. These generic categories can be used as a guide to identify factors for the deterrence function. For example, deterrence factors affecting public transport use can be categorized into: (1) time factors, e.g. travel time, scheduling of activities and transport services, and time budgets; (2) cost factors, e.g. public transport fares, (3) reliability; (4) security; (5) quality; and (6) information and booking.

In terms of the mathematical formulation of the deterrence functions, Geurs and Ritsema van Eck (2001) summarize the forms of distance decay functions that have been used in most of accessibility studies:

1. A negative power or reciprocal function (i.e. $F(d_{ij}) = d^{-\alpha}$), where d is the distance and α is a constant, which has, for example, been used by Hansen (1959), Patton and Clark (1970), Davidson (1977) and Fotheringham (1982).
2. A negative exponential function (i.e. $F(d_{ij}) = e^{-\beta d}$), where β is a constant, which has, for example, been used by Wilson (1971), Dalvi and Martin (1976), Martin and Dalvi (1976) and Song (1996).
3. A modified version of the normal function (i.e. $F(d_{ij}) = 100 * e^{-d^2/u}$), where u is a constant. This function has, for example, been used by Ingram (1971) and Guy (1983).
4. A modified logistic function (i.e. $F(d_{ij}) = 1 + e^{a+b*\ln d}$), where a and b are constants (Bewley and Fiebig, 1988). This function has been used by Hilbers and Verroen (1993).

The choice of which specific distance decay function to use depends on (a) the specific characteristics of the function and (b) the study area and the nature of the empirical data (Geurs and Ritsema van Eck, 2001). For example, Hilbers and Verroen (1993) indicate that the following aspects were relevant in their studies:

1. The steepness of the function. A negative and a negative exponential function decay very rapidly, suggesting a strong sensitivity to short distances. From a behavioural point of view, a very strong decay at short travel distances or times does not seem realistic, i.e. the perception of distance will probably not be very different between a 3-minute and a 6-minute trip. Fotheringham (1982) states that a power function gives a more accurate description of the perception of distance at an interurban level than an exponential function, which may be more accurate on an intra-urban level. Hilbers and Verroen (1993) state that, in general, a conventional logistic function will give a better behavioural explanation of distance decay because of its S-shaped form.

2. The functions' point of inflection. Some functions (such as the conventional logistic function) have a fixed point of infection halfway the maximum trip likelihood and this implies that the perception of distance is assumed to be the same for short and long travelling distances.
3. The value of the trip likelihood at zero distance. For the estimation of the trip likelihood it is necessary that the function reaches the maximum trip likelihood when the distance is zero.

In summary, the accessibility of a location is influenced by and interacting with four components (Geurs and Ritsema van Eck, 2001): land-use, transport, the individual and the temporal components. Accessibility is a location factor for inhabitants and firms (i.e. land-use component) which influences travel demand (transport component), people's economic and social opportunities (individual component) and the time needed for activities (temporal component).

Each trip has other characteristics which make the generalisation for the purpose of analysis difficult (Halden *et al.* 2000). For example, the reason for not making a walking or public transport trip may be the need to carry goods, the weather, the perceived quality of the route, including personal security and safety considerations, or simply a lack of knowledge of available options. These factors can be affected by transport policy decisions, so it is desirable if appraisal can take account of them in a meaningful way. To ensure a robust approach, calibration against observed behaviour should provide a firm foundation on which to build. Also, as travel patterns are not static, observations of travel behaviour should ideally take account of trends in trip making rather than simply observed demand.

3.3 REVIEW OF ACCESSIBILITY MEASURES

This section gives an overall review of different approaches to accessibility measures which can be classified to a number of classes. For example, for practical application purposes, Halden *et al.* (2000) classify accessibility into three generic but overlapping types of indicators: simple indicators, opportunity indicators and value or utility indicators (also see Handy and Niemeier, 1997).

3.3.1 Simple indicators

With simple indicators, the representation of transport and/or opportunities within the accessibility equation is simplified by defining threshold measures of the travel cost, time, etc., required to reach a given number of opportunities. Simple measures are fairly easy to understand and are most useful for local walking and cycling trips including assessing access to public transport services. The disadvantage however, is mainly the limited scope of these measures. The commonly used indices categorised under simple indicators include:

3.3.1.1 Catchment/contour indices

Catchment/contour indices count the number of people, jobs, shops etc., within a threshold travel cost (distance, time etc.) from a defined location. They are used for a wide variety of planning purposes for both land use and travel infrastructure and are often used by developers to consider the potential commercial viability of a potential development location.

3.3.1.2 Access to public transport

Rather than looking at transport network accessibility to destinations, they indices measure walking access time to the public transport services. Walking time or distance thresholds to public transport services are set and summed across all the available services. The quality of public transport being accessed is categorised on a scale which takes account of service frequency, type of service (i.e. rail/bus/light rail etc.) and service reliability. Although of limited scope, the simplicity of this approach has proved attractive and the calculation and mapping procedures have been automated and marketed by various organisations (LPAC, 1994).

3.3.1.3 Peripherality indices/rural accessibility

These identify thresholds in terms of cost, distance, time etc., from defined types of opportunities. These are usually calculated from major centres of population such as towns or cities or public services such as hospitals, but have also been used to study accessibility to transport networks including the European Community Trans European Networks.

3.3.1.4 Time space geographic measures

These measures simplify travel behaviour and choice in terms of the opportunities available within a limited time budget. The threshold is therefore the travel time available for a particular individual or group. These are widely used in logistics planning for freight but are equally applicable to people accessibility issues.

Developed by Hagerstrand (1970) within the *space-time* framework, the constraint-oriented approach is based on the fact that individual accessibility has both *spatial* and *temporal* dimensions. This approach considers the temporal dimension of activities which leads to indicators that account for the individuals' time constraints and the recognition of multipurpose activity behaviour by a space-time prism. However, Wang (1996) points out that this approach is not realistic as it assumes a constant speed in all directions and variable speed makes the model exceedingly burdensome to handle, and also the activity schedules are usually incomplete and do not cover the whole spectrum of activities

An example of the simple measures, given by Halden *et al.* (2000), is discussed here. The accessibility measure for a location (*i*) is calculated as the sum of the opportunities available at alternative locations (*j*) within defined threshold:

$$A_i = \sum_j O_j \delta_{ij}$$

Where A_i is the accessibility measure for a location i; O_j are the opportunities available at locations j; $\delta=1$ if the opportunities are within the threshold, and 0 otherwise.

3.3.2 Opportunity indicators

Opportunity indicators sum all the available opportunities and weight them by a measure of deterrence based upon how easily the opportunities can be reached. Opportunities also have the benefit of being easy to understand since, like the simple measures, they are expressed in terms of number of jobs or number of people for example. They have many potential uses including: the comparison of accessibility changes for different population groups, the identification of the catchments for destination, and the comparison of accessibility for car available and non-car available trips. The following sections briefly review a number of examples of opportunity indicators.

3.3.2.1 Hansen indices - the potential to opportunities or the gravity approach

The simple measures above are all special forms of Hansen indices incorporating thresholds to simplify data or analysis requirements. Hansen indices have had wide application within research and are used within transport models to estimate trip distribution (Halden *et al.*, 2000).

Indicators based on spatial opportunities available to travellers are among the first attempts to address the behavioural aspects of travel. The *potential to opportunities* or the *gravity* approach is the most utilised technique among accessibility indicators (see, Dalvi and Martin, 1976; Linneker and Spence, 1991; Geertman and Ritsema Van Eck, 1995; Bruinsma and Rietveld, 1998; Brunton and Richardson, 1998; Kwan, 1998; and Levinson, 1998)). An early attempt was made by Hansen (1959), who claimed that accessibility is the "potential of opportunities for interaction" or literally "a generalization of population-over-distance relationship". The concept of potential to opportunities is closely associated with the gravity models based on the interaction of masses and has been extensively discussed by Rich (1978). Advantages of Gravity or Opportunities measures include ease of comprehension and ease of calculations and the ability to differentiate between locations. Also, they are less demanding on input data than other indicators that reflect behavioural aspects. Some disadvantages of this class of indicators are their sensitivity to the choice of demarcation area and their deficient treatment of travellers with dispersed preferences.

3.3.2.2 Shimbel measures

This is a specific case of the Hansen indices in which all specified opportunities are assumed to have the same weighting.

Graph Theory measures (Garrison, 1960; Muraco, 1972; Vickerman, 1974) consider the degree of node (i.e. the number of links incident to each node) or the associated number (i.e. the number of links in the shortest path from a particular node to its most remote mode which is taken as a reference point, König, 1936). Shimbel (1953) suggested a measure to overcome the problem of taking the most remote node as a reference point, and this measure takes account of all possible destinations for each node. The Shimbel measure is simply the sum of the cost (e.g. time, etc.) to each of the opportunities and it indicates the accessibility of each node with respect to its linkage to all other nodes in the network.

3.3.2.3 "Economic potential" measures

Where the opportunities being considered in the Hansen index are regional incomes and the deterrence function is measured in distance, the accessibility index is sometimes described as the economic potential of a location (Keeble *et al.*, 1982).

Here is an example of the opportunity measures given by Halden *et al.* (2000). The opportunity measure for a location (*i*) is calculated as the sum of the opportunities available at locations (j) multiplied by a deterrence function based upon the travel time between i and j:

$$A_i = \sum_j O_j \exp(-\lambda t_{ij})$$

Where

A_i is the accessibility measure for a location i;

O_j are the opportunities available at locations j;

$\exp(-\lambda t_{ij})$ is the deterrence function;

t_{ij} is the travel time between i and j; and

λ is the factor for correction sensitivity to travel time, where a higher value means that travel time is more of a deterrent. The calibration of λ is usually undertaken as part of the trip distribution stage. However, even without location calibration the accessibility analysis can still be useful, since default values of λ can be used to give meaningful results (Halden *et al.*, 2000).

3.3.3 Value or utility based indicators

Value measures seek to define the attractiveness of the available opportunities to represent their value as a transport choice. They are expressed in generalised time or cost so findings can be more difficult to interpret. However, by providing a direct measure of the value of transport systems they could be powerful appraisal tools.

These indicators measure the value to a group of the choices available to them. The main difference with the opportunity measures is that additional opportunities only provide an increase in accessibility if they provide some additional value. If there is already a surfeit of opportunities available, adding more opportunities will result in little change in the index.

Utility-based indicators have their roots in travel demand modelling. Ben-Akiva and Lerman (1979) states: "accessibility logically depends on the group of alternatives being evaluated and the individual traveller for whom accessibility is being measured." In that sense, the shortcoming of gravity-based indicators becomes obvious, as all individuals within the same zone will experience the same amount of accessibility, regardless of the differences between their perceived utility of alternatives. Ben-Akiva and Lerman (1979, 656) continue: "for any single decision, the individual will select the alternative which maximises his/her utility." The measure of accessibility defined in this way is in monetary units, which enables the comparison of different scenarios. Williams (1977) noted that utility-based accessibility is linked to consumer welfare. By definition, a person's consumer surplus is the utility, in money terms that a person receives in the choice situation (Jong *et al.*, 2005). The consumer surplus associated with a set of alternatives is, under the logit assumptions, relatively easy to calculate. If the unobserved component of utility is independently and identically distributed extreme value and utility is linear in income, the expected utility becomes the log of the denominator of a logit choice probability, divided by the marginal utility of income, plus arbitrary constants, this is called the 'logsum'. McFadden (1975) and Small and Rosen (1981) showed how this measure can be derived in the discrete choice situation for the multinomial logit (MNL) model when income effect is not present. The advantage of this approach is that it is supported by relevant travel behaviour theories. Some disadvantages include the demand of extensive data on locations and individuals' travel behaviour and their choice sets.

Another utility-based accessibility measure is the activity-based accessibility measure (ABA, Dong *et al.*, 2006), which measures accessibility to all activities in which an individual engages, incorporating constraints such as scheduling, and travel characteristics such as trip chaining. The ABA is an extension of the logsum accessibility measures frequently derived from joint destination and mode choice models. Compared with more traditional measures of accessibility it is successful in (a) capturing taste heterogeneity across individuals; (b) combining different types of trips into a unified measure of accessibility; and (c) reflecting the impact of scheduling and trip chaining on accessibility.

3.3.4 An alternative accessibility measure – stated preference (SP) accessibility measure

Ortúzar *et al.* (2000) review the access measure with a microeconomic base and propose an alternative measure, in the perspective of approaching what the individuals perceive as access. Stated preference tools with their ability to manage the set of available alternatives, not only in terms of definition, but also in relation to the variation of the relevant attributes considered, are used to collect the data specifically focused on the problem of access perception. An access perception model was developed using multinomial logit modelling techniques and the study considered explicitly the full set of household members as decision makers. The variables used included travel time to work and to study by an individual, walking distance to the nearest underground station or bus stop, value of the house rental, number of workers and students in the household, and frequency of trips to work and to study. It was concluded that this measurement instrument was capable of discriminating between location effects in terms of the included variables. The parameters from this method can be taken as referential for evaluation purposes or as a comparison with those parameters calibrated from actual location-choice data including other location characteristics.

3.3.5 Some issues in the specification of accessibility measures

Handy and Niemeier (1997) discuss a number of interrelated issues that need to be resolved in the specification of the accessibility measure, regardless of the class of measure: the degree and type of disaggregation, the definition of origins and destinations, and the measurement of attractiveness and travel impedance.

The question of disaggregation is particularly important and has multiple dimensions. The most fundamental dimension is spatial disaggregation. Typically, accessibility is measured by zone, thus grouping individuals and household by proximity. The smaller the zone, the greater the disaggregation. All else being equal, smaller zones should result in more accurate estimates of accessibility for the individuals and households in the zone, as accessibility can vary greatly across small distances. Accessibility can also be measured separately for each household or individual, an approach which emphasizes the individual or household as the decision-making unit (Hanson and Schwab, 1987; Guy, 1983).

Accessibility measures can also be disaggregated according to socio-economic characteristics; this is important given that different segments of the population care about different sets of opportunities and may evaluate them differently (Wachs and Kumagai, 1973; Niemeier, 1997). In general, some differentiation of individuals and households by selected characteristics should result in more accurate accessibility measures.

The purpose of the trip or the type of opportunity represents another dimension of disaggregation. At the most aggregate level, accessibility to employment regardless of type is measured as employment serves as an indicator of overall activity. Finer levels of disaggregation distinguish between work and non-work opportunities (Guy, 1983; Hanson and Schwab, 1987).

The second issue that arises in developing accessibility measures is the origin and destination of the accessibility measure, i.e., the question of from where and to where accessibility will be measured. Usually home-based indicators are used. Thus, accessibility is measured for a resident who begins or ends his or her trip at home. Given the increasing importance of non-home-based trips, the appropriateness of a home-based measure must be revaluated (Lerman, 1979).

The set of potential destinations to include must also be determined. The desired level of disaggregation with respect to types of opportunities is the first criterion by which destinations are screened; for example, if the intent is to measure accessibility to shopping, then only shopping destinations should be included. But the set of destination opportunities to include also depends on assumptions as to the perceived choice set, in other words, the set of potential destinations that residents perceive to be available to them (Morris *et al.*, 1979). Researchers must ensure that the destination opportunities used in any accessibility measure reflect the needs of residents (Voges and Naude, 1983). Research on activity-based modelling points to the need for careful definition of choice sets and suggests that spatial and temporal constraints must be considered so that the focus is on 'constrained-choice sets' (Ben-Akiva *et al.*, 1987; Hanson and Schwab, 1986; Jones *et al.*, 1983).

The measurement of travel impedance presents yet another specification issue to resolve. Distance or time, common measures of impendence, can be estimated by straight-line distance (Baxter and Lenzi, 1975), network models (Sherman *et al.*, 1974) and field surveys (Wickstrom, 1971). If travel time is used, a choice must be made as to whether uncongested (or, off-peak) or congested (or peak) times will be used. The use of a generalized transport cost function, incorporating both time and monetary costs, is often an improvement over the use of time alone. Difference in travel time and cost by mode must also be addressed. One approach is to calculate accessibility separately for different modes – car accessibility and public transport accessibility. A more challenging approach is to incorporate car and public transport travel times as well as the opportunity to travel by other modes into one measure of accessibility.

The final specification issue surrounds the measurement of the attractiveness of an opportunity. This may simply be the existence of a particular opportunity, as measured by the number of establishments, or it may be either its physical or its economic size, as measured by area or employment, for example. Research on shopping behaviour shows that many characteristics of a potential destination (such as the quality and price of products or the quality of service), are important for destination choice (Bucklin, 1967; Guy and Wrigley, 1987).

3.4 ACCESSIBILITY IN THE FOUR-STAGE MODEL

This section discusses how accessibility has been included in the classic four stage transport model except the first stage - trip generation, which will be discussed in a later section.

3.4.1 Trip distribution and accessibility

Changes in network costs involve important changes in relative transport prices. The cost element may be considered in terms of distance, time or money units. In trip distribution, usually, the generalized cost of travel is used to combine all main attributes related to the disutility of a journey and it is typically a linear function of the attributes of the journey weighted by coefficients which attempt to represent their relative importance as perceived by the traveller. If the generalized cost is measured in money units then the time coefficients are sometimes interpreted as *values of time* as their units are money/time. The generalized cost of travel represents an interesting compromise between subjective and objective disutility of movement. It is meant to represent the disutility of travel as perceived by the trip maker; in that sense the value of time should be a perceived value rather than an objective, resource-based, value (Ortúzar and Willumsen, 2001).

3.4.2 Modal split and accessibility

Different accessibility measures have been used in modal split models (Bruton, 1985). In the trip-end modal split models developed in the early 1960s, such as the Puget Sound and the South-eastern Wisconsin Regional Land Use Transportation Study, accessibility indices were used as a measure of the quality of service provided by the alternative modes of transport. These indices measure the ease with which activity in one area can be reached from a particular zone on a specific transportation system. For example, the accessibility from zone i to zone j is defined as the product of trip attractions in zone j multiplied by the friction factor for the zonal interchange. These products are then summed from zone i to all other zones in the area to obtain the accessibility index for zone i. The friction factor is derived from door-to-door travel time, which, for motor vehicles, includes walking at origin and destination, 'unparking' and parking time, and driving time, while, for public transport, includes walking and waiting time at origin; time spent travelling on the vehicle; changing time between vehicles where applicable, walking time at destination.

3.4.3 Traffic assignment and accessibility

The basic premise in assignment is the assumption of a rational traveller, i.e. one choosing the route which offers the least perceived (and anticipated) individual cost (Ortúzar and Willumsen, 2001). In route choice, two factors are commonly considered: time and monetary cost. Monetary cost is often deemed proportional to travel distance. The majority of traffic assignment programs allow the user to allocate weights to travel time and distance in order to represent drivers' perceptions of these two factors. The weighted sum of these two values then becomes a generalised cost used to estimate route choice. In the case of public-transport assignment the generalized cost of travelling may include the in-vehicle travel time, the walking time to and from stops (stations), the waiting time at stops, the interchange time, an intrinsic 'penalty' or resistance to interchange which is measured in time units, fare charged to travel.

From the above sections we see that accessibility measures have been incorporated in trip distribution, modal choice and trip assignment models. Any change in the transport network (such as transport infrastructure, level of service of public transport) could be reflected in the change at these stages.

3.5 ACCESSIBILITY IN TRIP GENERATION MODELS: LITERATURE REVIEW

This section discusses some previous work, which attempts to model impacts of different accessibility measures on trip generation models. An overview of the accessibility measures for private transport and public transport is presented in Table 3.1 and Table 3.2. In most of these attempts the modellers consider characteristics of public transport services and transport infrastructures/ networks. Impacts of transport policies on accessibility measures have not been considered however. More detailed discussions of these measures are given in the following sections.

Mansfield (1969) incorporated journey time and money cost of travel variables in his linear regression model of recreational trip generation to a single destination (the Lake District). The purpose of the study was to investigate how the demand for pleasure journeys was affected by changes such as a reduction in journey times consequent on the opening of a new motorway. It showed that the demand for recreation trips appears highly elastic with respect to changes in total travel costs (money costs and the value of journey time).

Two accessibility measures proved to have a significant effect on trip generation; the first was named 'accessibility index' and used the reciprocal of the total minimum travel time from one zone to other zones to express the efficiency of the highway service. The second was the 'transit accessibility measure' which used the sum of the transit service frequency available at the zone. Both measures were used in the trip generation stage of the Baltimore Metropolitan Area Transportation Study (Wilbur Smith and Associates, 1964). However, Leake and Huzayyin (1979) argue that the measure of transit accessibility does not reflect the distribution (length of routes) operating in each zone, and also the measure does not make any reference to zone size and hence cannot distinguish between zones of different shape and area.

A public transport accessibility measure was developed in the London Traffic Survey (1966) to the off-peak frequency of buses (its square root) in a zone and the square root of the area (to compensate for the unequal size of the zones). This measure was tried in the trip generation phase of the study, but did not significantly improve the trip generation relationships that were established. Leake and Huzayyin (1979) point out that although this measure takes into account zone size (area), it does not reflect route length in each zone.

Singer (1973) adopted a doubly constrained gravity model which uses a combined generation and distribution function. Daly (1997) indicates that although the doubly constrained gravity form uses a theoretically correct functional formulation, it effectively links the elasticity of the trip generation model rigidly to that of the distribution model, a constraint which cannot be accepted on behavioural grounds.

Table 3.1 Overview of private transport accessibility measures in trip generation modelling

Private Transport Accessibility Measure	Applications and Conclusions	References
Total travel costs – money costs and value of journey time	Highly elastic for recreational trips to a single destination	Mansfield, 1969
The reciprocal of total travel time from one zone to the others	Significant	Wilbur Smith and Associates , 1964
The total travel distance or time between zones	Little improvement	Leake and Huzayyin, 1979
Relative accessibility and stratification of zones according to location	Adds little to the statistical strength of zonal regression trip production and attractions equations	Nakkash and Grecco, 1972
Attraction-accessibility index (number of establishments, squared ($1/d_{ij}^2$) deterrence function)	It is the most satisfactory; however, accessibility did not play a clear role in explaining trip rates in OLS model	Vickerman, 1974
Gravity-type index - combining destination attractiveness and travel time (at zone level)	Not significant in ordered response model of household shopping trip generation	Agyemang-Duah and Hall, 1997
A function of the size of the attraction *i* and the separation of zone *i* from all other zones *j*	Person trip attractions	Freeman, 1976
Doubly constrained gravity model using a combined generation-distribution function	Links the elasticity of trip generation to trip distribution	Singer, 1973
A 'logsum' from a choice model over the possible modes and destinations	Significant correlation obtained for only one of the two areas studied by zonal regression models	LGORU, 1975
The 'logsum' of destination choice	Using the hierarchical structure and statistically significant coefficients obtained	Cambridge Systematics Europe, 1981; HCG and TOI, 1990 ; Cohn *et al.*, 1996

In his shopping trips study with data from Oxford, Vickerman (1974) used one Shimbel accessibility measure (in terms of distances and bus travel time), two accessibility indices for levels of bus service (one is the average off-peak bus frequency to the City Centre and the other is bus-miles per hour available in each zone, standardised by zonal population to allow for different zone sizes and also to reflect the demand on available services, thus indicating the standard of comfort), and a combined attraction-accessibility index; this uses the number of establishments and the squared ($1/d_{ij}^2$) deterrence function. The index is summed for each origin zone over all zones, including the origin zone, so that the strong influence of the home zone is included; the distance for the home zone was taken as the average internal distance to the zone centroid.

Table 3.2 Overview of public transport accessibility measures in trip generation modelling

Public Transport Accessibility Measure	Applications and Conclusions	References
Sum of the transit service frequency	Significant	Baltimore Metropolitan Area Transportation Study, 1964
Off-peak frequency of buses (its square root) in a zone and the square root of the area	Not significant	London Traffic Survey, 1966
Shimbel measure in terms of distances and bus travel time; Average off-peak bus frequency to the City Centre; Bus-miles per hour in each zone standardised by zonal population	Accessibility did not play a clear role in explaining trip rates in OLS model.	Vickerman, 1974
Public transport: service frequency and zonal coverage by bus routes; A composite measure for both private and public transport	Significant improvements when modelling public transport and 'all modes' trips; greatest impact for home-based 'other' purposes (non work) trips	Leake and Huzayyin, 1979

Vickerman (1974) concluded that in many respects the attraction-accessibility associated with the spatial interaction model is the most satisfactory, particularly if it can be calibrated in a form constrained only at the production end and using exogenous attraction weights related to consumer expenditure and choice range at the destination. He rejected the doubly constrained gravity model, preferring to model trip generation as an explicit step separate from spatial interaction. Based on linear regressions on data from Oxford, the results showed that accessibility did not play a clear role in explaining trip rates, although some significant results were found. Wilson (1971) suggested that different forms of spatial interaction models might be appropriate for different trips purposes. For example, a model of journey to work would consider the number of workers resident at the origin zones and the number of jobs at the destination zones, and a model of journey to shop would consider the purchasing power of the residents at origin zones and a measure of the attractiveness of shops at the destination zones.

Agyemang-Duah and Hall (1997) used an accessibility index in an ordered response model of household shopping trip generation. The accessibility index was a single factor combining destination attractiveness, measured as the number of retail shopping employees in each zone, and travel time. This factor was calculated at the level of the traffic zone and the exponential function was used as deterrence function. They obtained a negative sign (i.e. counterintuitive) of the estimated coefficient of the accessibility index (Ortúzar and Willumsen, 2001 report negative signs of similar accessibility measures used in regression models). It is pointed out that a possible cause is that the number of vehicles owned by a household and the accessibility index are not truly independent.

In LGORU (1975), accessibility was incorporated into a zonal linear regression model of trip generation for two small rural areas with an accessibility measure calculated as a 'logsum' from a choice model over the possible modes and destinations available for travel from the origin zones. However, a significant correlation was obtained for only one of the two areas studied.

Nakkash and Grecco (1972) examined the effect of accessibility on both trip production and attractions using Hanson accessibility. Regression models were developed based on zonal variables and two considerations were introduced: (1) the concept of relative accessibility, and (2) stratification of zones according to location. The equations developed indicated that for home-based productions the inclusion of accessibility variables and stratification by location made virtually no improvement over those not incorporating such aspects. However, when relative accessibility was excluded, and stratification by location included, there was a general improvement in the model. Similar results were obtained for the trip attraction equations, but the effects were much stronger. It was concluded that the use of this index added little to the statistical strength of the regressions.

Kitamura (1991) expressed concern that the above aggregate, zone-level accessibility measures would be problematic due to too little variation between zones (and no variation within zones) and that they are two insensitive to detect the effect of accessibility on trip frequency.

Freeman (1976) indicated that the Hanson accessibility index can be seen to be associated with the production end of trips and is suitable for the analysis of person trip productions and not suitable for the analysis of person trip attractions. He advised that the index required for person trip attractions should be a measure of the accessibility to activities in zone i from all other zones j and should be defined as a function of the size of the attraction i and the separation of zone i from all other zones j. In situations of large zone sizes the accessibility of a zone to itself i.e. intra-zonal accessibility can be taken into account. The relative attraction accessibility of a zone can be calculated using the attraction accessibility in the zone divided by the sum of attraction accessibilities in all zones. By allocating personal trip attractions to zones on the basis of zonal relative attraction accessibility, the number of person trips attracted to any zone may be obtained.

Leake and Huzayyin (1979) point out the weaknesses of the Graph Theory measure: (1) 'distance' between nodes should be measured in terms of real travel distance, or generalized cost, not links in the path between them which has no sound logical basis (Muraco, 1972; Vickerman, 1974); (2) these measures cannot reflect public transport levels of service in terms of service frequency; (3) it is very difficult to produce a combined measure based on the Graph Theory approach to reflect all modes of transport (Vickerman, 1974); and (4) they principally measure nodal accessibility of the network and are difficult to modify for measuring household accessibility.

Leake and Huzayyin (1979) summarize the weaknesses of the activity-accessibility measures: First, problems associated with the determination of the power of the travel resistance term incorporated in these measures: a) there is a prior need to calibrate a 'gravity-type' trip distribution model in order to determine the power of the travel resistance term; b) alternatively, an arbitrary travel resistance function may be used; c) the assumption of a stable travel resistance has to be made to enable future accessibilities (Nakkash, 1969). Secondly, problems associated with the activity measure: a) as the suggested measure of activity (employment, labour force, shopping floor area, etc.) may be one of the socio-economic variables of the trip generation model, the potential for high inter-correlation between the accessibility measure and one or more of the socio-economic variables is likely to exist (Vickerman, 1974); b) as different types of activity measure are recommended for different trip purposes, this may result in the difficulty to establish an accurate accessibility measure for use in trip generation equations modelling combined trip purposes.

Practically, it is impossible to establish the activity-accessibility measure at the household level, since the determination of appropriate travel resistance functions would necessitate calibrating a gravity trip distribution model at this disaggregate level, as against the normal practice of calibrating at the zonal level. Furthermore, measures determined at a zonal level should not be used in a disaggregate trip generation model (Doubleday, 1976; Huzayyin, 1978).

Leake and Huzayyin (1979) proposed transport accessibility measures for private transport and public transport respectively and a composite accessibility for both of them. They pointed out that the efficiency of the private transport system depends primarily on the layout of the road network (network structure) and the ease/difficulty of travel on its various links. Private transport accessibility is then based on either the travel distance (total shortest route travel distance between zones) or travel time (minimum total travel time between zones). When revised, they can reflect the structure of the road network. Public transport accessibility should reflect the level of service provided by the public transport system in terms of frequency (bus/hr) and coverage by bus routes. So it considers the number of public transport routes, the number of modes, the length of each route, and the frequency of each mode. Also it can consider the area of the zone. By combining a selected private transport accessibility measure with one of the public transport accessibility measures, a composite measure can be formed.

The results from the above research indicate that the greatest impact of accessibility always occur in the case of home-based 'other purposes (non work)' trips. This shows the sensitivity of this category of trip productions to the characteristics of the urban transport system. This research has shown that for certain trip types the introduction of an accessibility measure can result in significant improvements in the explanatory power of a trip production model. This was particularly noticeable when modelling public transport and 'all modes' trips. However, little improvement was obtained when modelling private trips.

Leake and Huzayyin (1979) conclude that many failures to improve significantly the explanatory power of trip generation models by introducing an accessibility measure may have been due to unsatisfactory accessibility measure formulations, inadequate data, or a combination of both.

Daly (1997) also investigated an accessibility measure by calculating the logsum of destination choice in an integrated trip generation, mode and destination choice model using the hierarchical logit modelling structure. A number of applications of this approach have been developed including Cambridge Systematics Europe, 1981; HCG and TOI, 1990 and Cohn *et al.*, 1996. A coefficient of accessibility was also introduced in the Norwegian National long-distance tour generation Model (HCG and TOI, 1990), with a coefficient value ranging from 0.07 to 0.33 for the modelled five trip purposes. In the 'ProMise' model developed for Netherlands Railways (Cohn *et al.*, 1996), statistically significant coefficients for accessibility were calibrated in the tour generation models for the 'optional' travel purposes, i.e. non work, business or educational travel. Logsum coefficients ranging from 0.03 to 0.11 were obtained for 0, 1+ and stop-go models.

3.6 GENERAL DISCUSSIONS OF ACCESSIBILITY MEASURES AND ITS APPROPRIATENESS FOR INCLUSION IN TRIP GENERATION MODELLING

3.6.1 Introduction

Leake and Huzayyin (1979) outlined the basic requirements of an accessibility measure when used in a trip generation model as: (1) it should be easy to understand and logically expressed; (2) it should reflect the efficiency of private transport and the service levels provided by public transport; and (3) two different sets of accessibility measures are required for private transport and public transport which should be possible to combine into one measure representing accessibility by all modes of transport for use in trip generation models.

They further claim that any accessibility measure to be introduced into a trip generation model should be in harmony with the used trip generation modelling technique:

1. The measure should be capable of accurate calculation, i.e. no errors to satisfy one of the assumptions of the least squares method;
2. The accessibility measure should not be highly correlated with any of the socio-economic variables;
3. The accessibility measure should be capable of being established at both the zonal and household levels so that it can be included in a trip generation model calibrated at either level of aggregation;
4. The measure should be capable of reflecting accessibility for each of the traditional trip purposes as well as any combination of trip purposes.

It should be noted however, that these requirements mainly consider factors and attributes which represent existing characteristics of the transport system. New policies implemented have not been considered by almost all the researchers who investigated accessibility in trip generation models. Moreover, the perceived levels of service by the users have not been considered. It should be noted the other travel choice models, the perceived levels of service of the transport system are often used as well as or instead of the actual level of service because of their importance (Ortúzar and Willumsen, 2001). The advantage of using actual characteristics of the transport system is that data is easier to collect and it is more convenient. The disadvantage, however, is that the actual characteristics of the transport system could be differently perceived than the actual characteristics and also differently perceived by different types of users. In addition, in all previously investigated accessibility indicators, there was no inclusion of policy variables (for example road pricing, parking pricing, etc.) The following section discusses the gaps in previous approaches.

3.6.2 Gaps in previous approaches

As discussed above, although the impacts of various transport policies such as pricing, public transport and management measures have been investigated at the trip distribution, modal choice and route choice stages, these have not been applied at the trip generation stage.

The Hanson accessibility measure and the Freeman attraction accessibility measure consider the opportunities in zones and the travel impedance between zones. The Leake and Huzayyin accessibility measures consider the layout of the road network, the ease/difficulty of travel, and the level of service by the public transport system in terms of service frequency.

Transport system characteristics only are included in terms of the "observed" characteristics of the public transport services as well as transport infrastructures/network in most models. How people really think and their perceptions and experiences that underlie attitudes, beliefs and consequent behaviour are not considered. Although accessibility is determined by patterns of land use and by the nature of the transportation system, two people in the same place may evaluate their accessibility differently, as wants and tastes vary (Handy and Niemeier, 1997).

Also, transport policies such as pricing measures and their impacts have not been considered in any previous research. Some policies have been considered as opportunities (i.e. policies aim at increasing trip generation to/from specific zones, such as public transport measures, pedestrianisation etc.), while others can be seen as impedance as they may reduce some types of trip generation (e.g. pricing measures). Thus when transport policies are introduced they would impact on accessibility as well as trip generation.

Therefore the general requirements for a transport accessibility measure could be summarised as:

1. The measure should be capable of accurate calculation;
2. The measure should not be highly correlated with any of the socio-economic variables;
3. The measure should be capable of reflecting accessibility for each of the trip purposes;
4. Variables which reflect perceived level of service of transport systems should be included in the measure;
5. Policy variables which reflect further characteristics of the transport systems should be included in the measure.

In this research, journey times and public transport cost are included for work trip generation models in Chapter 6. Policy measures such as parking cost and congestion charge have also been investigated in trip generation modelling (see Chapter 7 and Chapter 8 respectively). Finally, a perceived public transport accessibility measure taking account of people's opinions and perceptions of public transport services and its impacts on trip generation modelling has been investigated (see Chapter 9).

3.7 SUMMARY

This chapter first introduces the concept of accessibility which is related to spatial distribution of land use, the transport infrastructure and public transport services, temporal and individual factors. Different approaches of travel impedance can be suggested to reflect the sensitivity to the distances.

With simple indicators, the representation of transport and/or opportunities within the accessibility equation is simplified by defining thresholds (e.g. number of relevant opportunities within a given travel cost or time). Opportunity indicators sum all the available opportunities and weight them by a measure of deterrence based upon how easily the opportunities can be reached. Value measures seek to define the attractiveness of the available opportunities to represent their value as a transport choice.

Accessibility has been included in trip distribution, modal choice and route choice models of the classic four stage transport models, where usually a generalized cost function including a measure of accessibility, is used. This function can easily reflect the changes to the transport network which are caused by the introduction of transport policies. When transport policies are introduced they would impact on trip generation as well as the other stages of the transport model. These types of impacts have not been widely explored at the trip generation stage in previous research.

In this chapter, a review of how accessibility measures have been incorporated in trip generation models is presented. The results from the studies that incorporated different measures and the strengths and weakness of these measures were discussed. In Chapters 7 and 8 respectively, policy measures such as parking cost and congestion charge have been investigated in trip generation modelling.

CHAPTER 4 MODELLING TECHNIQUES OF TRIP GENERATION

In this chapter, generalised linear models which unify diverse statistic techniques (e.g. linear regression and logistic regression) and their suitability for different response variables and explanatory variables are discussed. Also, the logistic regression technique for trip generation is reviewed including choice theories, different types of discrete choice models, joint estimation of revealed preference (RP) and Stated Preference (SP) data, and methods to evaluate the performance of models. Finally, the suitability of using logistic regression in modelling trip generation is discussed.

4.1 GENERALIZED LINEAR MODELS

The term 'Generalized Linear Model' (GLM) is due to Nelder and Wedderburn (1972), who showed how linearity could be exploited to unify apparently diverse statistical techniques. Generalized linear models are specified by three components (Agresti, 1990): a random component, which identifies the probability distribution of the response variable; a systematic component, which specifies a linear function of explanatory variables that is used as a predictor; and a link describing the functional relationship between the systematic component and the expected value of the random component.

The random component of a GLM considers independent observations $Y = (Y_1, \ldots, Y_N)'$ from a distribution in the natural exponential family. That is, each observation Y_i has a probability density function, or mass function, of the form

$$f(y_i;\theta_i) = a(y_i)b(y_i)\exp[y_i Q(\theta_i)]$$

This family includes several important distributions as special cases, including the Poisson and binomial. The value of the parameter θ_i varies for $i = 1, \ldots, N$, depending on values of the explanatory variables. The term $Q(\theta_i)$ is called the "natural parameter" of the distribution. The systematic component of a GLM relates a vector $\eta = (\eta_1, \ldots, \eta_N)'$ to a set of explanatory variables through a linear model:

$$\eta = X\beta$$

Here X is a matrix of values of the explanatory variables for the N observations, and β is a vector of model parameters. The vector η is called the linear predictor.

The third component of a GLM is a link between the random and systematic components. Let $\mu_i = E(Y_i)$, $i = 1, \ldots, N$. Then μ_i is linked to η_i by $\eta_i = g(\mu_i)$, where g is any monotonic differentiable function. Thus the model links expected values of observations to explanatory variables through the formula

$$g(\mu_i) = \sum_j \beta_i x_{ij}, \quad i = 1, \ldots, N$$

The function $g(\mu) = \mu$ gives the identity a link $\eta_i = \mu_i$, specifying a linear model for the mean response. The link function that transforms the mean to the natural parameter is called the canonical link. For it, $g(\mu_i) = Q(\theta_i)$, and $Q(\theta_i) = \Sigma_j \beta_i x_{ij}$.

In summary, a GLM is a linear model for a transformed mean of a variable having distribution in the natural exponential family.

Both linear regression and logistic regression are special cases of Generalised Linear Models (Dobson 2001). Linear regression is the standard method for relating a continuous response variable to several continuous explanatory (or predictor) variables. Linear models have the form

$$E(Y_i) = \mu_i = x_i^T \beta; \qquad Y_i \sim N(\mu_i, \delta^2)$$

where $Y_1, \ldots, Y_N$ are independent random variables. The link function is the identity function, i.e., $g(\mu_i) = \mu_i$. The model is usually written as

$$y = X\beta + e$$

where $e = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$ and the e_i's are independently, identically distributed random variables with $e_i \sim N(0, \delta^2)$ for $i = 1, \ldots, N$.

In this form, the linear component $\mu = X\beta$ represents the 'signal' and e represents the 'noise'. Multiple linear regression, analysis of variance (AVOVA) and analysis of covariance (ANCOVA) are all of this form, and together are called general linear models. Multiple linear regression is used to analyse one continuous response variable and multiple explanatory variables. ANOVA is used for a continuous response variable and categorical or qualitative explanatory variables (factors). And ANCOVA is used when at least one of the explanatory variables is continuous. For details about linear regression, see Neter *et al.* (1996).

Logistic regression is used to model relationships between a response variable which is binary or categorical, with more than two categories, and several explanatory variables which may be categorical or continuous. The link function for logistic regression is

$$g(\pi) = \log\{ \pi/(1-\pi)\}$$

where π is the response probability.

Binary logistic regression is used for binary response variables. Multinomial logistic regression is used for responses with more than two nominal categories; ordinal logistic regression, for ordinal categories, is also included in logistic regression. For details about logistic regression, see Hosmer and Lemeshow (2000) and Agresti (1990).

A summary of the main methods of statistical analysis for various combinations of response and explanatory variables is shown in Table 4.1.

Table 4.1 Major methods of statistical analysis

Response	Explanatory Variable	Methods
Continuous	Nominal, >2 categories	Analysis of variance
	Nominal & some continuous	Analysis of covariance
	Categorical &continuous	Multiple regression
Binary	Categorical & continuous	Logistic regression
Nominal with >2 categories	Nominal	Contingency tables
	Categorical & continuous	Nominal logistic regression
Ordinal	Categorical & continuous	Ordinal logistic regression

Source: Edited from Dobson (2001)

4.2 CHOICE THEORIES AND UTILITY MAXIMISATION

Most of the current discrete choice models are based on utility maximisation concepts. Discrete choice models essentially deal with the decision making process of a decision maker who is faced with a number of mutually exclusive alternatives (Ben-Akiva and Lerman, 1985). Therefore, four elements are defined in the choice process: 1) decision maker; 2) alternatives; 3) attributes of alternatives; and 4) decision rule. The decision maker can be an individual, a household, a company or any other decision-making unit. The alternatives, which are referred to as the 'choice set', are the set of alternatives available to the decision making from which to choose in the context of mode choice, car purchase choice, etc. Usually there are two general types of choice sets: for one type the choice set is continuous such as in the case of "commodity bundles" (e.g. the set of the amounts of milk, bread and butter) and for the other it is discontinuous where the choice set is three television sets denoted A, B and C.

To fit within a discrete choice framework, the set of alternatives needs to exhibit three characteristics (Train, 2003): first, they must be mutually exclusive from the decision maker's perspective, i.e. choosing one alternative necessarily implies not choosing any of the other alternatives. Second, the choice set must be exhaustive; in that all possible alternatives are included (i.e. the decision maker necessarily chooses one of the alternatives). Third, the number of alternatives must be finite (i.e. the researcher can count the alternatives and eventually be finished counting).

The first and second criteria are not restrictive. Usually an appropriate definition of alternatives can assure that they are mutually exclusive and that the choice set is exhaustive. For example, a set of alternatives might not be exhaustive because the decision maker has the option of not choosing any of them. But if an extra alternative, defined as 'none of the other alternatives' is added, then this expanded choice set is exhaustive. The above two conditions can be often satisfied in several different ways. The appropriate specification of the choice set in these situations is governed largely by the goals of the research and the data that are available to the researcher.

The third condition, that the number of alternatives is finite, is restrictive and this is the defining characteristic of discrete choice models and distinguishes their realm of application from that for regression models. With regression models, the dependent variable is continuous, which means that there are an infinite number of possible outcomes. When there are an infinite number of alternatives, discrete choice models cannot be applied.

4.3 RANDOM UTILITY THEORY

Random utility theory (Domencich and McFadden, 1975; Williams, 1977; Manski, 1977) is the most commonly used theoretical basis of the decision rule theories. In the random utility approach, it is assumed that an individual's preference among available alternatives can be represented with a utility function. The individual (n) has a choice amongst several possible alternatives (J). Random utility theory assumes that each individual obtains some utility from each alternative U_{in}, $i = 1, \ldots, J$. Moreover, the individual is assumed to choose the alternative, which maximises his/her utility. Thus, the behavioural model is, that an individual n will choose alternative i if and only if the utility of alternative i is greater than the utility of each other alternative in the choice set:

$$U_{in} > U_{jn} \forall j \neq i$$

However, the modeller does not possess complete information about all the elements considered by the individual making a choice, therefore, U_i is assumed to be represented by two components:

$$U_i = V_i + \varepsilon_i$$

where V_i is the deterministic (observable) element of the utility which is a function of the measured attributes; and ε_i is the random term (unobservable element) of the utility which accounts for the unobserved attributes of alternatives, unobserved taste variations, measurement errors and imperfect information.

The probability of an individual choosing alternative i is simply the probability that the utility of that alternative is greater than the utility for any other alternative.

$$P_i = \Pr ob(U_i > U_k) \forall k \neq i$$

That is

$$P_i = \Pr ob(\varepsilon_k < \varepsilon_i + V_i - V_k) \forall k \neq i$$

The residues ε are random variables with a certain distribution which can be denoted by $f(\varepsilon) = f(\varepsilon_1,...,\varepsilon_N)$.

Different discrete choice models are obtained from different specifications of this density $f(\varepsilon)$, that is from different assumptions about the distribution of the unobserved portion of utility (Train, 2003). Logit, by far the most widely used discrete choice models, is derived under the assumption that ε_i is independent and identically distributed (IID) extreme value for all i and the critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives.

4.4 THE LOGISTIC REGRESSION MODELS

4.4.1 Introduction

The logistic regression model originated from the odds concept in gambling contexts. Widely used by professional gamblers, the odds is the expected number of times an event will occur to the expected number of times it will not occur. Odds of 4 means 4 times as many occurrences as non-occurrences. Odds of 1/5 means that we expect only one-fifth as many occurrences as non-occurrences. There is a simple relationship between probabilities and odds. If p is the probability of an event and O are the odds of the event, then

$$O = \frac{p}{1-p} = \frac{\text{Probability of event}}{\text{Probability of no event}}$$

Logistic regression is popular in part because it enables the researcher to overcome many of the restrictive assumptions of ordinary least squares (OLS) regression (Garson, 2002): it does not assume a linear relationship between the dependent and the independent variables; the dependent variable need not be normally distributed (but does assume its distribution is within the range of the exponential family of distributions, such as normal, Poisson, binomial, gamma); and normally distributed error terms are not assumed.

Logistic regression analysis has been widely used in mode choice, route choice and destination choice of the traditional four-stage transport models and other transport models such as car ownership model and departure time choice. However it has not been much investigated in trip generation modelling.

The main types of logistic regression are discussed in the following sections. See Hosmer and Lemeshow (2000) and Train (2003) for the further details.

4.4.2 Binary logistic regression

Binary choice models deal with a special case where the choice set contains exactly two alternatives. Of the binary choice models, the binary logit model arises from the assumption that $\varepsilon_n = \varepsilon_j - \varepsilon_i$ is logistically distributed. Under this assumption, the choice probability for alternative i is given by:

$$P(i) = \frac{1}{1+e^{-(V_i - V_j)}} = \frac{e^{V_i}}{e^{V_i} + e^{V_j}}$$

V_i and V_j can be linear in their parameters where

$$V_i = \beta_0 + \Sigma \beta X$$

where X are the independent variables representing the attributes, and βs are unknown parameters that need to be estimated.

4.4.3 Multinomial logit (MNL) model

The multinomial logit (MNL) model is the simplest and most popular practical discrete model and it is used for cases where the choice set has more than two alternatives. he MNL model assumes that the error terms are independently, identically Gumbel distributed across cases (also known as type I extreme value) which results in a simple and elegant closed-form model (Domencich and McFadden, 1975).

The MNL model is derived through the application of utility maximisation concepts to a set of alternatives from which one, the alternative with maximum utility, is chosen. A general expression for the multinomial logistic regression is:

$$P(Y = i|x) = \frac{e^{V_i}}{\sum_{j=1}^{k} e^{V_j}}$$

Where

$P(Y = i|x)$ is the probability that an individual will choose alternatives i;

V_i is the deterministic component of the utility of alternative i for the individual; and k is the number of alternatives.

Bhat (2000) summarizes the three basic assumptions that underlie the MNL formulation. The first is that the random components of the utilities of the different alternatives are independent and identically distributed (IID) with a type I extreme-value (or Gumbel) distribution. The assumption of independence implies that there are no common unobserved factors affecting the utilities of the various alternatives. This assumption is violated when some common underlying unobserved factors impact on the alternative utilities and this has implications for competitive structure. The second assumption of the MNL model is that it maintains homogeneity in responsiveness to attributes of alternatives across individuals (i.e. an assumption of response homogeneity). More specifically, the MNL model does not allow sensitivity (or taste) variations to an attribute (e.g. travel cost or travel time in a mode choice model) due to unobserved individual characteristics which, however, can and generally affect responsiveness. Ignoring the effect of unobserved individual attributes can lead to biased and inconsistent parameter and choice probability estimates (Chamberlain, 1980). The third assumption of the MNL model is that the error variance-covariance structure of the alternatives is identical across individuals (i.e. an assumption of error variance-covariance homogeneity). This assumption may not be appropriate if the extent of substitutability among alternatives differs across individuals. Error variance-covariance homogeneity implies the same competitive structure among alternatives for all individuals, an assumption that is generally difficult to justify.

The MNL models satisfy the axiom of independence of irrelevant alternatives (IIA) which can be stated as: where any two alternatives have a no-zero probability of being chosen, the ratio of one probability over the other is unaffected by the presence or absence of any additional alternative in the choice set (Luce and Suppes, 1965). This property holds that for a specific individual the ratio of choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives which can be shown as:

$$\frac{P(j)}{P(i)} = \frac{e^{V_j}}{e^{V_i}} = e^{V_j - V_i}$$

When IIA reflects reality (or an adequate approximation to reality), considerable advantages are gained by its employment (Train, 2003). First, because of the IIA, it is possible to estimate model parameters consistently on a subset of alternatives for each sampled decision maker. Since relative probabilities within a subset of alternatives are unaffected by the attributes or existence of alternatives not in the subset, exclusion of alternatives in estimation does not affect the consistency of the estimator. Another practical use of the IIA property arises when the researcher is only interested in examining choices among a subset of alternatives and not among all alternatives, and this would save the researcher considerable time and expense developing data on other alternatives.

The MNL model has the property of uniform cross elasticities – that is, the cross elasticities of all alternatives with respect to a change in an attribute affecting only the utility of alternative j are equal for all alternatives $i \neq j$. For the linear-in-parameters multinomial logit model, the convenient form which is known as the incremental logit can be used to predict changes in behaviour on the basis of the existing choice probabilities of the alternatives and changes in variables.

The specification of a multinomial logit model consists of a number of distinct steps. First, universal choice set C need to be defined for problem under study which may require some judgements about which alternatives can be ignored. The next step is to define the choice set for each individual and this is generally done by applying reasonable judgements about what constitutes the feasibility of an alternative in any particular situation. And finally, the particular variables entering into the utility functions must be defined.

Another issue of specification is about the functional form. Although the linear function is probably adequate in many contexts, there are others such as destination choice where non-linear functions are deemed more appropriate (Foerster, 1981; Daly, 1982). In the literature three approaches have been proposed: 1) the use of conjoint analysis in real or laboratory experiments to determine the most appropriate form of the utility form (Lerman and Louviere, 1978); 2) the use of statistical transformation, letting the data 'decide' to a certain extent (Gaudry and Wills, 1978); and 3) the constructive use of econometric theory to derive functional form (Train and McFadden, 1978; Jara-Díaz and Farah, 1987) and the final form can be tied up to evaluation measures of user benefit. In general, non-linear forms imply different trade-off to those normally associated with concepts such as the value of time (Bruzelius, 1979) and model elasticities and explanatory power may vary dramatically with function (Ortúzar and Willumsen, 2001)

The closed form of the MNL models makes it straightforward to estimate, interpret, and use. As a result, the MNL models has been used in a wide variety of travel and travel-related choice context, including mode, destination, car ownership, and residential location as well as choices in non-travel contexts. The MNL mode is one of the main techniques used in this study.

4.4.4 The nested logit model

The nested logit model is closed-form model, which relaxes the assumption of independent and identically distributed random-error terms in the MNL models to provide a more realistic representation of choice probabilities. It was the first closed-form alternative to the MNL and have been the most widely used alternative (Williams, 1977; Daly and Zachary, 1978).

Ortúzar (2001) and Carrasco and Ortúzar (2003) review the development of the nested logit (NL) model. Ortúzar (2001) mentions several authors whose work predates the model's actual theoretical formulation Wilson (1969, 1974), Manheim (1973) and Ben-Akiva (1974) all used intuitive versions that – although based on concepts such as marginal probabilities and utility maximization – did not have a rigorous construction of the functional forms and a clear interpretation of all the model parameters. Domencich and McFadden (1975) generated structured models of nested logit form which had an incorrect definition of 'composite utilities'. It was Williams (1977) who first made an exhaustive analysis of the NL properties, especially composite utilities (or inclusive values), showing that all previous versions had important inconsistencies with microeconomic concepts. He also reformulated the NL, and introduced structural conditions associated with its inclusive value parameters, which are necessary for the NL's compatibility with utility maximizing theory. With these, he formally derived the NL model as a descriptive behavioural model completely coherent with basic micro-economic concepts. Other authors, whose seminal work completed the fundamental theoretical development of the NL, are Daly and Zachary (1978), who worked simultaneously and totally independent from Williams, and McFadden (1978, 1981) who later generalized the work of both Williams and Daly and Zachary.

The nested logit (NL) model, which was further developed and applied by (Ortúzar, 1983; Hensher 1986; Daly 1987; Bierlaire *et al.* 1997; Koppelman and Wen 1998; Hensher and Greene 2002), is an extension of the multinomial logit model and it allows dependence or correlation between the utilities of alternatives in common groups (Williams, 1977; Daly and Zachary, 1978; McFadden, 1978). Derivation of the nested logit model is based on the same assumptions as the MNL model (Koppelman and Sethi, 2000), except that correlation of error terms is assumed to exist among predefined groups of alternatives. Such error correlations arise if an unobserved factor influences the utility of all members of the group. The nested logit model can be written as the product of a series of MNL choice models defining each level in a tree structure.

To be consistent with utility maximisation, the structural parameters at the highest level and the ratios of the structural parameters at each lower nest are bounded by zero and one. The estimated parameters at each node represent the ratio between the structural parameter at that node and at the next higher node in the tree. A value of one for any ratio of structural parameters implies that the alternatives in that nest are uncorrelated and can be directly connected to the next higher node. If all structural parameter ratios equal one, all the alternatives can be directly linked to the root of the tree; i.e., the structure collapses to the MNL.

The nested logit model, by allowing correlation among subsets of utility functions, alleviates the IIA problem of MNL in part. The model is suitable to use with correlated alternatives in a number of situations. Examples include, model choice models, where there are similarities between public transport alternatives (see for example, Forinash and Koppelman, 1993), car ownership models, where there may be similarities between types of vehicles for purchase (see for example, Mohammadian and Miller, 2003).

Other forms of models include the ordered logit models where the potential responses are ordered. For example, the rating of books from 1 to 7, where 1 is the worst you have ever read and 7 is the best and 6 is higher than 5, which is higher than 4. A standard logit model could be specified with each potential response as an alternative. However, the logit model's assumption of independent errors for each alternative is inconsistent with the fact that the alternatives are ordered: with ordered alternatives, one alternative is similar to those close to it and less similar to those further away (Train, 2003).

The ordered nature could be handled by specifying a nested logit, mixed, or probit model that accounts for the pattern of similarity and dissimilarity among the alternatives. However, such a specification, while it might provide fine results, does not actually fit the structure of the data, as the traditional derivation for these models starts with a specification of the utility associated with each alternative. For more discussions of these types of models see Páez *et al.* (2006).

4.5 MIXED RP/SP MODELS FOR MODEL ESTIMATION

4.5.1 Introduction

Before any realistic modelling process can be implemented, data must be collected or obtained on the characteristics of the transportation system to be modelled as well as the characteristics of the users. The data requirements and the choice of data types depend upon the objectives of the study, the time and resources available, and the characteristics of the study area. There are a number of data types/ sources which could be used to estimate choice models. In this research a number of data types have been utilised including national and household surveys, stated preference (SP) and revealed preference (RP) data as discussed in 4.10.

Revealed preference data (RP) and stated preference (SP) data have been widely utilised and used to calibrate travel choice models. Generally, stated preference data are analysed in the same way as revealed preference data, that is, using discrete choice analysis. However, SP data is different from RP data because usually respondents evaluate more than one choice scenario and thus contribute more than one observation. Therefore, because a number of observations are taken from each respondent in an SP choice study assuming independence between observations will be a weak approximation. In the case of RP data, only one observation is taken from each respondent, hence, it is fair to assume that there is independence between observations. Revealed preference and stated preference data are subject to different types of errors and hence it is unlikely that both sources of data will have the same distribution for the error term. Stated preference data may not be valid for prediction but could be useful for identifying and estimating underlying preferences that determine actual behaviour (Morikawa, 1989). Hence, there are strengths and weaknesses associated with both sources of data, and it may be desirable to combine the stronger features of RP and SP data. This may lead to improvements in the modelling exercise and provide a deeper understanding of choice behaviour.

4.5.2 Mixed RP/SP models

Mixed revealed preference (RP) and stated preference (SP) models which use RP and SP data have been used in many transport demand analyses (Cherchi *et al.*, 2005; Espino *et al.*, 2006). RP data are based on individual choices and allow the analyst to characterise actual travel behaviour. SP data are based on individuals' stated behaviour under hypothetical scenarios and are useful when the problem is to examine the demand for new alternatives or measure the effect of latent variables.

There are advantages and limitations to each type of data (see for example Ortúzar and Willumsen, 2001). Revealed preference data have the advantage that they reflect actual choices. However, such data are limited to the choice situations and attributes of alternatives that currently exist or have existed historically and they are not available for new situations. The advantage of stated preference data is that experiments can be designed to contain as much variation in each attribute as the researcher thinks is appropriate. The limitations of stated preference data include that there is no guarantee that people would do what they say they would actually do if they are faced with the choice situations presented to them.

The combined use of both types of data allows to exploit their respective advantages and to overcome their specific limitations (Ben-Akiva and Morikawa, 1990; Bradley and Daly, 1997; Louviere *et al.*, 2000). Stated preference data provide the needed variation in attributes, while revealed preference data ground the predicted shares in reality.

There have been many examples of application of mixed RP/SP models (Brownstone *et al.*, 2000; Bhat and Castelar, 2002; Cherchi and Ortúzar, 2002, 2006a, 2006b; Espino *et al.*, 2006). Brownstone *et al.* (2000) used mixed logit models of stated and revealed preferences for alternative-fuel vehicles. Bhat and Castelar (2002) used a unified mixed logit framework to analyse congestion pricing in the San Francisco Bay area. Cherchi and Ortúzar (2002) investigate incorporating interaction effects in mixed RP/SP models, and they further investigate how to fit mode specific constants in the presence of new options in RP/SP models (Cherchi and Ortúzar, 2006a). Cherchi and Ortúzar (2006b) estimate income, time effects and direct preferences in a multimodal choice context using mixed RP/SP models. Espino *et al.* (2006) analyse demand for suburban trips using a mixed RP/SP model with latent variables and interaction effects.

The mixed use of RP/SP data to estimate choice models requires that the variances of the error terms in RP and SP are equal; the quotient between those variances is known as "scale parameter" and denoted by λ (Ben-Akiva and Morikawa, 1990). Bradley and Daly (1997) proposed an estimation method based on the construction of an artificial nested logit (NL) structure (also see Louviere *et al*, 2000) where RP alternatives are placed just below the root and each SP alternative is placed in a single-alternative nest with a common scale parameter λ. The following sections summarise this method.

4.5.3 Comparisons of preference data

4.5.3.1 Conceptual framework

Louviere *et al.* (2000) show that the scale factor, which is inversely related to the error variance, is a measure of the statistical information contained in preference data. Therefore, they develop a conceptual framework based upon RUT to compare differences in choice or preference data sources. In this approach, it is assumed that the sample of respondents in a survey make choices from experimentally designed pairs of alternatives, each of which describe a product or a choice. The associated design matrix in this case is assumed to be X_1. Now assume that a second source of preference or choice data is also available. An example of this could be a reporting of a different independent sample of respondents on their last purchases from the choice options and the attributes associated with each option. The associated design matrix in this case is assumed to be X_2. Further assume that X_1 and X_2 have some common attributes (X_{c1}, X_{c2}) while other attributes are alternative specific (Z_1, Z_2). Figure 4.1

below shows a representation of this framework (also see Louviere *et al.* (2000) for more discussions of the approach).

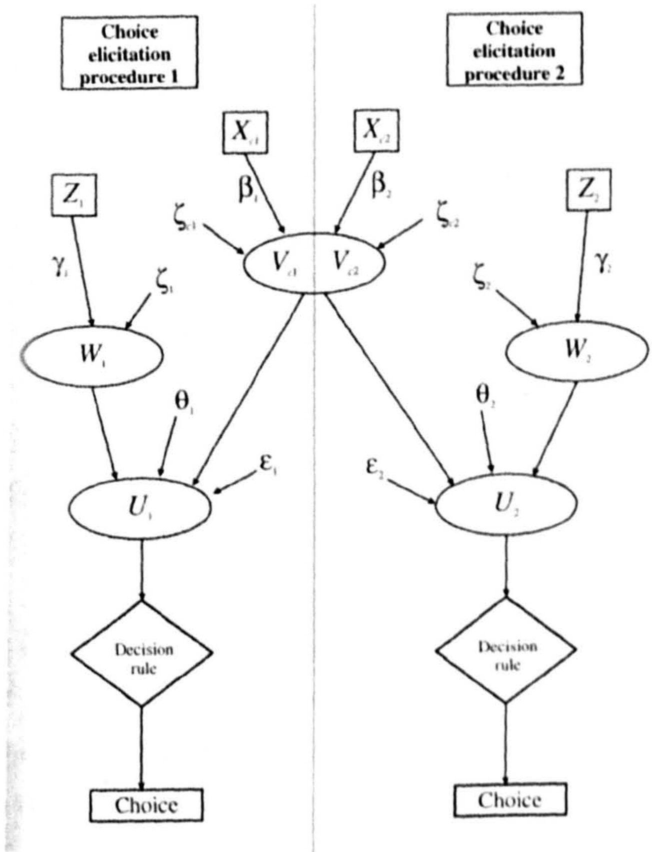


Figure 4.1 Conceptual framework for preference data comparison (Louviere *et al.*, 2000)

4.5.3.2 Preference regularities

Louviere *et al.* (2000) represented the consumer behavior and the preference measurement using the concept of preference regularities (PR). They claim that the existence of PR should be evaluated on the basis that the marginal common utility partworths measured in each source be equal to a multiplier for all common attributes. They developed a formal definition of this PR and illustrated how it could be applied to different data sources. They also proposed a basic test for the existence of preference regularities which is a generalization of the likelihood ratio test. A simple graph of marginal utilities (or parameter values (where utility functions are linear in the parameters) is plotted which could be used as a simple exploratory analysis tool to investigate the appropriateness of combining both data types.

Figure 4.2 graphically illustrates a proportionality condition that underlies the definition of *PR* in the two data sources. That is, if preference regularity holds between the two data sources, the marginal common utilities should be linearly related with a positive slope. Then, the graphic of the estimated parameters should plot as a straight line intersecting the origin (the slope is equal to λ_2/λ_1, i.e. the ratio of error variance of set 2 to that of set 1). The 'cloud' of points should occupy quadrants I and III, but not II and IV of the graph. If the cloud of points is too dispersed or too many parameters have opposite signs in the data sources (implying points in quadrants II and IV), therefore this provides evidence that parameter equality between data sets are less likely.

A key issue in the proposed approach is the recognition of the fact that it is not the absolute magnitudes of common utilities *per se* that matters in comparing multiple measures, but rather the comparability of the implied sensitivity of the measures to changes in attribute levels. If the two preference data sets contain the same underlying preference structure, but differ significantly in the magnitudes of random error, the two sets of estimated parameters will appear to differ significantly in absolute magnitude (Louviere et al., 2000).

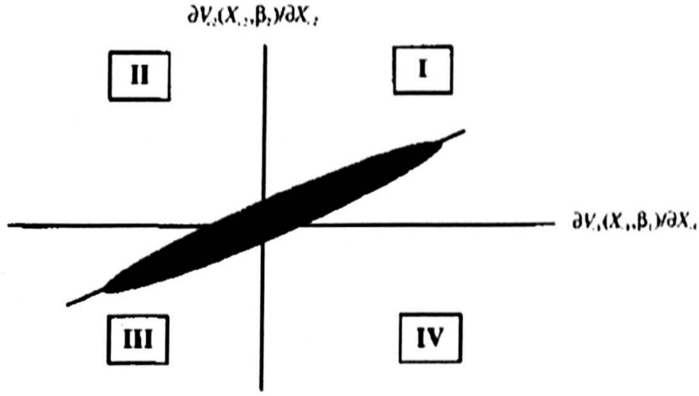


Figure 4.2 Preference regularity hypothesis generation by definition PR (Louviere *et al.*, 2000)

Further statistical tests that take into account the errors in the estimates could also be used to make references about preference regularities (see Louviere *et al.*, 2000 for further discussions).

4.5.4 Mixed RP/SP model estimation

There are a number of procedures or approaches for mixed RP/SP model estimation including a manual method using existing MNL software and the NL trick method. Firstly, the manual method, originally proposed by Swait and Louviere (1993), estimates the desired model parameters and the relative SP scale factor by manual search. This process first defines a range of values of λ^{SP} within which one expects the log likelihood function to be maximised, and then implements a one dimensional search to obtain an estimate of the relative scale factor of the SP data, and the estimates of λ^{SP} are obtained from the model solution that maximises the value of the log likelihood function. This method trades-off statistical efficiency for ease of implementation.

Secondly, a full information maximum likelihood (FIML) method which estimates model parameters and relative scale factor(s) simultaneously and optimise with respect to all parameters. Bradley and Daly (1992) and Hensher and Bradley (1993) proposed an artificial tree structure (i.e., the NL trick) to obtain an estimate of the scale factor of one data set relative to that of the other. The artificial trees can be extended to multiple data sources.

In the NL trick approach, the joint estimation of a choice situation using two types of data involves a choice outcome associated with the RP data and a number of choice outcomes associated with the SP data. The hierarchical structure (Hensher and Bradley, 1993), given in Figure 4.3, ensures that each of the parameter estimates associated with the SP data are scaled by the ratio of the variances. The different thetas on each dummy node are constrained to take the same value, a requirement for the scaling conditions. Different theta's can be allowed for each additional type of SP data sets.

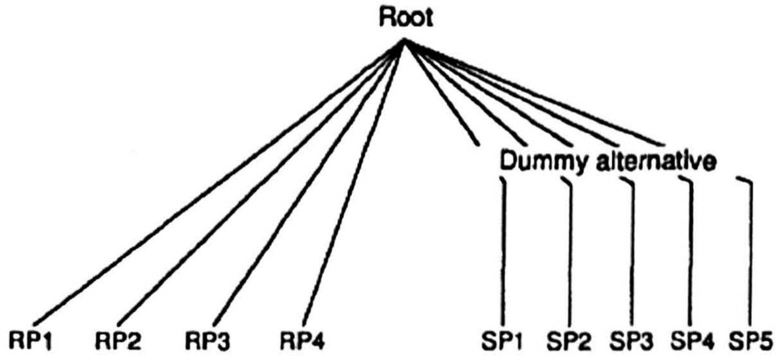


Figure 4.3 The estimation structure (Hensher and Bradley, 1993)

In a recent review by Hensher *et al.* (2008), they investigate the mixed RP/SP modelling using the nested logit 'trick'. In the approach, the modelling strategy assumes that the observations are independent, a condition of all GEV models. However, this condition is not strictly valid within a stated preference experiment with repeated choice sets and between each SP observation and the single RP data point. Hensher *et al.* (2008) suggest the replacement of the NL 'trick' method with an error components model that can accommodate correlated observations as well as reveal the relevant scale parameter for subsets of alternatives. Such a model can also incorporate ''state'' or reference dependence between data types and preference heterogeneity on observed attributes.

In some choice situations however, where there is no problem of repeated observations from the same respondents, one can possibly still use the NL trick model as discussed above. For example, in cases where the SP data is simply one observation to indicate potential future behaviour as the case of the SP data set used in this research (see Section 8.3 for further discussion).

A second potential source of error in the NL trick model is the state or reference dependence that is mainly resulting from preference heterogeneity between data types, which is possible to be positively or negatively affecting the preferences and hence the responses. A positive effect maybe a result of habit persistence while a negative effect could be the result of frustration with the inconvenience associated with the introduction of new policy measures (see Hensher *et al.*, 2008 for further discussion). In the case where the impact of the reference dependence might be negative, the implications of the reference dependence is less severe in the models.

In Chapter 8 of this thesis a mixed RP/SP model is calibrated using the NL trick approach. That was because the SP data consists of one response from each individual and therefore there was no problem of repeated observations. In addition, the impacts of the reference dependence is expected to be negative, if any, which would then results in less errors in the model.

4.6 THE METHOD OF MAXIMUM LIKELIHOOD

The most commonly used method of estimating the parameters of a logistic regression model is the method of maximum likelihood (Ryan, 1997). Maximum likelihood (ML) is based on the idea that although sample could originate from several populations, a particular sample has a higher probability of having been drawn from a certain population than from others (Ortúzar and Willumsen, 2001). Therefore the ML estimates are the set of parameters which will generate the observed sample most often.

To illustrate this idea a sample of n observations of a given variable $\boldsymbol{Z} = \{Z_1, \ldots, Z_n\}$ drawn from a population characterised by a parameter θ (mean, variance, etc.). As $\boldsymbol{Z}$ is a random variable it has associated a density function $f(\boldsymbol{Z}/\theta)$ which dependent on the values of θ. If all of the values of $\boldsymbol{Z}$ in the sample being independent, the joint density function can be written as

$$f(Z_1, Z_2, \ldots, Z_n / \theta) = f(Z_1 / \theta) f(Z_2 / \theta) .. f(Z_n / \theta)$$

The usual statistical interpretation of this function is with $\boldsymbol{Z}$ as variables and θ fixed. Inverting this process, the precious equation can be interpreted as a likelihood function $L(\theta)$; maximising it with respect to θ, the result is called maximum likelihood estimate because it corresponds to the parameter value which has the greatest probability of having generated the observed sample. Maximum likelihood can easily be extended to situations where the population is characterised by several parameters.

Suppose a sample of Q individuals is randomly obtained, for which their choice (0 or 1) and the value of x_{jkq} for each available alternative is observed, so that individual q is observed to choose alternative i.

As the observations are independent the likelihood function is given by the product of the model probabilities that each individual chooses the option they actually selected:

$$L(\theta) = P_{21}P_{32}P_{23}P_{14}\cdots$$

Defining the following dummy variable: $g_{jq} = 1$ if A_j was chosen by q; 0 otherwise. The above expression may be written more generally as

$$L(\theta) = \prod_{q=1}^{Q} \prod_{A_j \in A(q)} \left(P_{jq}\right)^{g_{jq}}$$

To maximise this function we differentiate partially with respect to θ and equate it to 0. We normally maximise $l(\theta)$, the natural logarithm of $L(\theta)$, which is more manageable and yields the same optima. Therefore, the function we seek to maximise is (Ortúzar 1982):

$$l(\theta) = \log L(\theta) = \sum_{q=1}^{Q} \sum_{A_j \in A(q)} g_{jq} \log P_{jq}$$

When $l(\theta)$ is maximised, a set of estimated parameters is obtained which is asymptotically distributed.

There are two reasons for this popularity of maximum likelihood (Allison, 1999): First, ML estimators are known to have good properties in large samples. Under fairly general conditions, ML estimators are consistent, asymptotically efficient, and asymptotically normal. Consistency means that, as the sample size gets larger, the probability that the estimate is within some small distance of the true value also get larger. No matter how small the distance is or how high the specified probability is, there is always a sample size that yields an even higher probability that the estimator is within that distance of the true values. One implication of consistency is that the ML estimator is approximately unbiased in large samples. Asymptotic efficiency is that, in large samples, the estimates will have standard errors that are, approximately, at least as small as those for any other estimation method. And, finally, the sampling distribution of the estimates will be approximately normal in large samples, which means that you can use the normal and chi-square distributions to compute confidence intervals and *p*-values.

The other reason for ML's popularity is that it is often straightforward to derive ML estimators when there are no other obvious possibilities. One case that ML handles very nicely is data with categorical dependent variables.

The method of maximum likelihood will generally perform well for large sample sizes. But for small data sets or data sets in which the average value of Y is close to zero or one, it can produce poor results, or even fail to converge (Ryan, 1997).

4.7 EVALUATING THE PERFORMANCE OF THE LOGISTIC MODELS

The criteria used to evaluate the performance of each model are as follows: 1) the sign of the coefficient (is it as anticipated); 2) the *t*-ratio for the coefficient (is it significant at the 95% confidence level?); 3) calculation of a likelihood ratio test; and 4) inspection of ρ^2 values for model goodness of fit.

4.7.1 Statistical significance of the coefficients

For discrete choice models the t-statistic is generally used to test significance for a single coefficient in a model. Sufficiently large values of t (typically bigger than 1.96 for 95% confidence levels) lead to the rejection of the null hypothesis and hence to accepting that the attribute has a significant effect. In discrete choice models t-statistics are asymptotic results (not exactly t-test), which imply that the test are only valid for very large samples.

4.7.2 Sign of the coefficient value

An informal test is to examine the sign of the coefficient estimates to judge whether it conforms with a *priori* notions or theory. Current practice recommends to include a relevant policy variable with a correct sign even if it fails any significance test and the reason is that estimated coefficient is the best approximation available for its real value and the lack of significance may just be caused by lack of enough data.

4.7.3 The likelihood-ratio (ρ^2) index

The asymptotic *rho*-squared (ρ^2) index, which varies between 0 and 1, similar to R^2 in linear regression, can be used to measure the goodness of fit for the model. It is noted that value of ρ^2 of between 0.2 and 0.4 are considered extremely good fits. The adjusted likelihood ratio index $\bar{\rho}^2$ (rho-squared bar) can be used to overcome the shortcoming that ρ^2 will always increase or at least stay the same whenever new variables are added to the utility functions.

4.7.4 Likelihood ratio test

The likelihood-ratio test is used in the same way that F test is used in regression models for joint tests of several parameters. It uses the ratio of the maximised value of the likelihood function for the full model (L_1) over the maximised value of the likelihood function for the simpler model (L_0). The likelihood-ratio test statistic equals:

$$-2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1)$$

This log transformation of the likelihood functions yields a chi-squared statistic with K degrees of freedom.

4.8 THE CHARACTERISTICS OF TRIP GENERATION AND THE SUITABILITY OF LOGISTIC REGRESSION FOR MODELLING TRIP GENERATION

In trip generation models, the response variable is the number of trips that people make which can range from zero to n. If n is large, the response variable can be seen as continuous and multiple linear regression can be applied with the prior assumption that there is a linear relationship between the response variable and the explanatory variables (Ortúzar and Willumsen, 2001).

However, n often is not very large. When n equals to one that is people choose making a trip or not, binary logistic regression may be preferable (Daly 1997). When n is larger than one, but limited (usually it is), that is people have several trip frequency choices, multinomial logistic regression can be applied as these choices are mutually exclusive from the traveller's perspective, i.e. the traveller chooses only one alternative; they are exhaustive i.e. all possible alternatives are included; and the number of alternatives is finite. When some transport policy is introduced, it would impact on trip frequency and it is important to investigate the change of trip frequency as well as the number of trips.

In trip generation modelling, the explanatory variables can be categorical (e.g. employment status, sex, type of dwelling) and continuous (e.g. income, age) and it is convenient to include both categorical and continuous explanatory variables in logistic regression.

While in linear regression models, the response variable is the number of trips, in logistic regression models the probability of an individual / household making a trip(s) is investigated and the total number of trips an individual / household makes can be obtained by the summation of the trip frequencies multiplied by their corresponding probabilities.

Discrete choice models, by treating the number of trips (or the trip frequency) as a set of mutually exclusive and collectively exhaustive categorical variables, incorporate built-in upper and lower limits. The models also provide a behavioural framework that directly links the number of trips to utility-based consumer and decision-making theory.

Some earlier attempts have been made to model trip generation / frequency using discrete choice models where the concept of trip frequency choice is introduced and the dependent variable is the probability of making the actual number of trips. As discussed in Chapter 2, Sheffi (1979) developed a nested-alternative-logit model in a disaggregate utility maximization framework for estimating probabilities of trip frequencies by elderly individuals. Barmby and Doornik (1989) and Jang (2005) used a count data / negative binomial model to estimate trip frequency. Daly (1997) proposed the use of a binary logistic model to estimate the probability that an individual will choose to make a trip and the use of a hierarchical structure, representing an indefinite number of choices, to model choice of frequency with what he called a 'stop-go' model. He and colleagues have made several applications of this approach in Europe (Bradley and Daly, 1997). The logistic regression models considered in this research include binary, multinomial (MNL) and nested logit (NL) models.

4.9 SOFTWARES FOR LOGISTIC REGRESSION

There are a number of software packages available for logistic regression modelling such as Alogit (Daly, 1992), SPSS, STATA, SAS and LIMDEP. For a discussion of some of these software packages see McDermott (1995). Alogit and SPSS are mainly used in this research.

4.10 SUMMARY OF SECTION, GAPS IN RESEARCH AND KNOWLEDGE IN TRIP GENERATION MODELLING AND STRUCTURE OF THESIS

In this section, a summary of the research knowledge and the identified gaps in trip generation analysis and modelling are discussed.

From the discussions presented in the last three chapters, it is clear that trip generation analysis and modelling are currently carried out using revealed preference socio economic data and using two main approaches; linear regression and category analysis. In linear regression analysis, the assumption of linearity of the independent variables with the dependent variables, the lack of built-in upper and lower limits to the number of trips, and the assumption that the number of trips is approximately continuous can all be questioned and could potentially lead to unreasonable predictions of trip generation (Páez *et al.*, 2006).

Similarly, most of category analysis trip generation models employ the basic category analysis techniques (CA and MCA_1) despite their apparent weaknesses. Although there have been further more recent advances in Multiple Classification Analysis techniques (MCA_2, MCA_3 and MCA_4, Guevara and Thomas, 2007), these have not been widely tested empirically. Using these techniques including the improved multiple classification analysis (MCA) methods, the large sample size required to calibrate the trip rates as well as the absence of statistical tests for the goodness of fit of these models undermines their adequacy. Logistic regression overcomes many of the restrictive assumptions of ordinary least squares regression and category analysis. This approach has been widely used to model other travel choices such as choice of mode, route choice, departure time choice and other travel choices. However, not many applications in trip generation modelling have been reported (Cohn *et al.*, 1996; Daly, 1997).

The lack of a behavioural justification in trip generation such as supported by the theory of random utility has been investigated and a large number of investigation attempts have been reported to date to include behavioural dimensions in modelling trip generations. For example, Vickerman and Barmby (1985) investigated the use of behaviour approach and a choice model to investigate trip generation. Bhat (1999) investigated the use of repeated choice observations models in analysing evening commuting trips. Golob (2000) developed a simultaneous model of household activity participation and trip chaining. Wallace et al. (2000) investigated the effects of travellers and trip characteristics on trip chaining, with implications for transportation demand management strategies and Misra et al. (2003) used a continuous time representation and modelling framework for the analysis of non worker activity-travel pattern.

Moreover, one of the main criticisms of trip generation models is the absence of any variables that represent transport policies that no doubt affect the trips generated (e.g. public transport, pricing and parking policies). Schmocker et al. (2005) studied the changes in the frequency of shopping trips in response to a congestion charge in London and the and found that within the sample surveyed the congestion charging scheme had caused a significant number to shop less often in central London and only a few to shop more often in the Oxford Street area. Kelly and Clinch (2006) investigated the potential impact of parking-pricing on trip generation by purpose and the results show there is no differential effect of a price change on business relative to non-business trips in the short run at the lower levels of increase in non-street parking price. However as the prices increases, significant results emerge; the users making trips for business purposes are less likely to cease parking in the area as a result of a price change relative to those making non-business trips. These policies are increasingly being considered as management tools in most world cities, and their impacts are always considered in mode, route, destination and departure time choices. Not many investigations of their impacts on trip generations have been reported though.

Most trip generation models are calibrated from aggregate revealed preference data despite the growing applications of other sources of data such as stated preference especially in travel demand forecasting, mainly because of the nature of trip generation models (Ortúzar and Willumsen, 2001; Daly and Miller, 2006; and Kouwenhoven, et al., 2006). SP techniques offer the opportunity to modellers to test impacts of policy measures on travel behaviour. So in principle there is no reason why these techniques cannot be used in trip generation modelling, especially if logistic regression analysis is used. It would be very useful to use stated preference techniques to investigate impacts of transport policies on trip generations as well as other choice models.

Finally, although accessibility of the transport system has been recognised and investigated in previous trip generation models as a function of the available opportunities or impedances (such as distance, travel time or cost), these were all variables representing the characteristics of the transport system but not the perceived level of service of the system (Ortúzar and Willumsen, 2001; Daly, 1997).

In summary, trip generation analysis, unlike the rest of travel choice analysis, has limitations in terms of the techniques (conventional techniques), data used (only revealed preference data) and type of variables (only socio-economic variables). These limitations have been recognised in the literature and acknowledged to impair the efficiency of trip generation models to produce accurate predictions.

The main aim of this research has been to investigate possible methodologies to improve performance of trip generation modelling (see further discussions in Chapter 1). In order to achieve this aim a number of objectives have been defined as discussed below:

1. Examine appropriateness of logistic regression analysis for modelling trip generation
2. Investigate, analyse and compare trip generation models using logistic regression, linear regression and category analysis including more recent multiple classification analysis techniques
3. Investigate and calibrate trip generation models which include transport policy measures
4. Explore the use of stated preference data (SP) to calibrate trip generation models
5. Investigate trip generation models with enhanced transport accessibility functions

The rest of the thesis is structured as follows. In Chapter 5, a number of data sets have been identified and analysed to carry out the investigations. The methodology adopted to model trip generation using logit analysis as well as the calibrated work trip models are presented in Chapter 6. Calibrations of trip generation models using the conventional (linear regression and category analysis including multiple classification) models are presented in Chapter 7. Predictions from all the models and analysis and comparisons of the results are presented in Chapter 8. A data set from Edinburgh Household Survey has been used to calibrate linear and logistic regression models of trip generation (shopping trips), taking into account parking costs as transport policy measure. These results are presented in Chapter 9. An SP data from Edinburgh Household Survey is used to calibrate mixed RP/SP logistic regression models for trip generation taking account of introducing road user charging as a policy measure, and presented in Chapter 10. A public transport accessibility measure is calibrated as a function of the distance from the city centre and the perceived level of service of the public transport system by the users which is discussed in Chapter 11. A discussion of the results of the research is summarised in Chapter 12 and the research is concluded in Chapter 13.

CHAPTER 5 DESCRIPTION AND PRELIMINARY ANALYSIS OF DATA USED IN THE STUDY

5.1 INTRODUCTION

The main objectives of this research are to calibrate and compare trip generation models including logistic regression analysis and to investigate impacts of including transport policies and transport accessibility in trip generation models (see Chapter 1).

Therefore the data needed had to include the following information:

1. Trip generation patterns
2. Socio-economic characteristics
3. Transport policies and their impacts on trip generation
4. Transport accessibility and its impact on trip generation

It was initially planned to collect the data for this research using a specifically designed questionnaire. A detailed questionnaire was designed to be carried out to collect data from a small sample in Edinburgh to investigate potential impacts of transport policies on shopping trip generation activities in Edinburgh.

The aim of the travel survey was to investigate travel to shopping and to test the impacts of various transport policies that include parking management, parking pricing, congestion charging and improvement of public transport on the number of shopping trips.

A questionnaire (See Appendix 1) was designed which consists of the following four sections:

(1) Travel survey for shopping trips in Edinburgh;

(2) People's attitudes on transport and transport policies;

(3) The potential impacts of such policies on shopping trips to the city centre; and

(4) Socio economic information of household and individual.

The questionnaire was sent out to be piloted. Unfortunately however, in addition to the low response rate, the majority of the returned questionnaires were incomplete. It became obvious then that the collection of enough data to carry out the analysis of this research would be very difficult. The alternative was to use data from existing surveys such as the National Travel Survey (NTS), the Edinburgh Household Survey (HS), the Scottish Household Survey (SHS) or the Edinburgh Shoppers' Survey. It was not possible to use only one set of these data since each of them has its limitations as well as its advantages as discussed below.

The National Travel Survey is a household survey of travel covering residents of Great Britain (GB) and include information on the purpose of each trip made, the modes of transport used, the timing of the trip, and the origin and destination, demographic data, such as age, sex, and other information relevant to travel such as income, employment status, ownership of cars and other vehicles, details of driving licences and the availability of local public transport. More discussions of this survey are included in the following sections. The information is collected on a national level and therefore does not reflect regional characteristics. However, this very large data set allowed the calibration and analysis of the trip generation models using the three techniques (logistic regression, linear regression and category analysis) as discussed in Chapters 6, 7 and 8.

The Edinburgh Household Survey (HS) included information on the socio economic data and the impacts of congestion charging on shopping behaviour in the city centre. More discussions of this survey are included in the following sections. The availability of this data allowed the calibration of trip generation models which include transport policies (in this case parking charges and congestion charging). See Chapters 9 and 10 for discussions of these models.

The Scottish Household Survey (SHS) data is a continuous survey based on a sample of the general population in private residences in Scotland (Hope, 2002). The aim of the survey is to provide representative information about the composition, characteristics and behaviours of Scottish households, both nationally and at a more local level. The sample is being drawn from the small user file of the Postcode Address File (PAF). As part of the main questionnaire, a travel diary collects information about personal travel on the day prior to the interview. One randomly chosen adult per household in the sample is selected to complete the travel diary. There were 686 individuals available in Edinburgh for their travel information from Monday to Friday. This data was not used however in the analysis since it did not provide any information on the impacts of transport policies on the frequency of shopping trips while the Edinburgh Household Survey did.

Finally, the Edinburgh Shoppers' Survey was principally designed to provide a snapshot of spending patterns in the City Centre. More discussions of this survey are included in the following sections. As it was a survey of all visitors to the City Centre, it included tourists, day visitors and those who go there for work, as well as shoppers. This survey provided information on the perceived accessibility to travel to and from the central area of Edinburgh. This information was used to calibrate a trip generation/ accessibility model as presented in Chapter 11.

5.1 NATIONAL TRAVEL SURVEY DATA

Part of the data used in this research was taken from the National Travel Survey (NTS, Kershaw *et al.*, 2001). This is a household survey of travel covering residents of Great Britain (GB) where every household member in the sample is asked to keep a seven-day diary of all personal travel within GB. Parents are asked to keep the diary for young children. Diary details include the purpose of each trip made, the modes of transport used, the timing of the trip, and the origin and destination. The household member are also interviewed to provide background demographic data, such as age, sex, and other information relevant to travel such as income, employment status, ownership of cars and other vehicles, details of driving licences and the availability of local public transport.

The NTS is based on a random sample of private households. First, postal sectors are chosen and these are 'stratified' so that the sample is representative at the regional level by car ownership and social-economic group. Then households are chosen at random in each of these sectors. This results in a 'clustered' sample, which is necessary to reduce the costs of interviewers' travelling time. Survey takes place throughout the year starting on a random, but pre-determined, day of the week. The data used for this analysis were from the 2002/2004 surveys where there were 23,817 households covering the whole UK. In total there are 55,552 individuals, of which 61.6% belong to the 16-64 age group. In a week, these individuals make 903,826 trips with different purposes.

This study investigates work trips (i.e. commuting and business) per household in a day (Wednesday), therefore only those households with at least one worker were chosen (i.e. 1,4091 households). Furthermore, as the dataset represented the whole of the UK and there are large variations in household characteristics, only urban areas of 50,000–250,000 residents (i.e. 2,706 households) were used to obtain more homogeneous data with fewer variations. Some general statistics of the whole dataset and the selected dataset are presented here which include journey purposes, household size, car ownership and household income.

5.1.1 The distribution of trips by journey purpose

Table 5.1 presents the distribution of trips from the NTS (1996-1998 and 2002-2004) by journey purpose.

Table 5.1 National Travel Survey trips by journey purpose

Journey Purpose	1996/98 NTS data (%)	2002/04 NTS data (%)
Commuting	18.0	16.7
Business	4.1	3.7
Education	5.5	5.8
Shopping	3.5	8.8
Non-food shopping	16.8	10.4
Personal business	6.9	9.0
Visit friends at private places	14.0	12.2
Entertain-public places	4.0	5.1
Escort education	3.8	4.1
Escort shopping	3.3	3.8
Other	20.1	20.4

In the 2002-2004 survey, commuting and business trips accounted for about 20.4% of trips, which shows a reduction of 1.7% from the 1996-1998 survey (22.1%). Education trips represented 5.5% and 5.8% of all trips in the two datasets respectively while travelling to shopping trips accounted for 20.3% and 19.2% of all trips (that is shopping plus non food shopping trips). Although there is no big change for the total number of shopping trips, it shows a shift of the different types of shopping, i.e. a 5.5% increase in food shopping and a 6.4% decrease in non-food shopping. Other significant changes are trips for personal business which had an increase of 2.1% and visit friends at private places which had a decrease of 1.8%. In this research trip generation models for commuting and business trips have been investigated.

5.1.2 Number of workers in household

Table 5.2 and Table 5.3 present the number of people employed in the household in the 2002-2004 NTS dataset (n=20,214) and in the sample used for model calibration in this study (n=1,979) respectively. In the complete dataset (Table 5.2), about 30.3% of the households have no workers, 29.6% of households have one worker, either in full time or part time employment, and 40.2% have two or more workers, either in full time or part time employment.

Table 5.2 Number of workers in household in the complete NTS survey (n=20,214)

No. of Workers in the Household	Full Time (%)	Part Time (%)	All (%)
None	38.2	73.4	30.3
1	38.3	23.7	29.6
2	20.1	2.7	32.7
3	3.4	0.2	7.5

Table 5.3 Number of workers / household in the data used for model calibration (n=1,979)

No. of Workers in the Household	Full Time (%)	Part Time (%)	All (%)
None	11.2	62.5	0
1	54.8	33.2	43.3
2	29.2	3.8	45.5
3	4.8	0.5	11.2

In the sample used for model calibration, about 11% of the households have no full time workers, 54.8% have one full time worker and 34% have 2+ workers. 37.5% have at least one part-time worker. As this study investigates the work trip generation and as workers' status (full time / part time) could have a different impact on the number of work trips, the number of full time and part time workers will be included separately in the trip generation models of Chapter 6.

5.1.3 Number of children in the household

Table 5.4 presents the distribution of the number of children in the selected survey data. The table shows that 61.7% of households have no child. Households with one child represent about 16.2% of the sample and about 20.6 % of households have two or more children.

Table 5.4 Number of children in the selected sample data (n=1,979)

Number of Children	Percentage
0	61.7
1	16.2
2	15.6
3	5.0
4 or more	1.5

5.1.4 The distribution of car ownership

Table 5.5 presents the distribution of car ownership in the complete NTS dataset and in the sample used for model calibration. In the first case about 20.4 percent of households do not have access to a car, 46.8% have one car and about one third had two or more cars. While for those with at least one worker, only 10.2% has no car and over 40% have two or more cars. It should be noted that 8.4% of selected data has one or more company cars. Car ownership is one of the main variables which affect trip generation. It is well established that as car ownership increases, the number of trip generations increase.

Table 5.5 Car ownership for the whole dataset (n=20,214) and the selected data for model calibration (n=1,979)

Car Ownership	Complete NTS (2002/2004) Dataset (%)	Selected Data (%)
0	20.4	10.2
1	46.8	49.3
2	27.9	34.2
3+	4.9	6.4

5.1.5 The distribution of household income

Table 5.6 presents the distribution of household income. 30.7 percent of households' annual income is less than £19,999 and 27 percent over £40,000. The other 42.3% of households have income between £20,000 and £39,999.

Table 5.6 Household income

Household Income	Percentage
Less than £19,999	30.7
£20,000-£39,999	42.3
£40,000 and over	27.0

Similar to car ownership income contribute positively to the increase in the number of trips, and therefore is included in trip generation models which are calibrated in this study.

5.2 HOUSEHOLD SURVEY AND SHOPPERS' SURVEY IN EDINBURGH

Another source of data used in this study was gathered as part of the Household Survey and Shoppers' Survey (ECCM, 2004) by DTZ Pieda consultants, who investigated shopping trips in Edinburgh and the impacts of implementing a congestion charge on these trips.

The Edinburgh Household Survey (HS) (ECCM, 2004) included information on the socio economic data including age, gender, car ownership and social grade, mode of travel for shopping and location of residence. Respondents were also asked to report on their non-food shopping trip frequency into the city centre in a week and the parking costs. The Household Survey examined the effect of congestion charging on shopping behaviour in the city centre catchments' area of Edinburgh.

The Shoppers' Survey on the other hand was principally designed to provide a snapshot of spending patterns in the City Centre. As it was a survey of all visitors to the City Centre, it included tourists, day visitors and those who go there for work, as well as shoppers.

The Shoppers' Survey was conducted on weekdays between 7am and 6.30pm from 31 May to 11 June. A total of 1,000 randomly selected shoppers were interviewed on a sample of days and times throughout this period. The household survey was conducted by telephone interview from 14th June to 2nd July 2004 with a total of 1,199 interviews. The survey was a representative quota sample of households in three areas: 1) the Edinburgh city centre, 2) the area of Edinburgh between the two proposed cordons, and 3) the Edinburgh "hinterland", comprising of Midlothian, West Lothian, East Lothian and part of Fife.

The general statistics of the two surveys are presented in the following sections. For more details of the survey, see ECCM (2004).

5.2.1 Household survey

In the household survey, respondents were asked to report on their non-food shopping trip frequency into the city centre in a week. Also they were asked about the mode of transport for shopping, and their perception of the potential impacts of introducing congestion charge on shopping trips.

5.2.1.1 Shopping trip frequency

The frequency of shopping trips to the city centre was investigated in the household survey. Table 5.7 shows the frequency of shopping visits for all respondents and for car users only. About 10% of all the respondents in the survey reported that they shop in the city centre daily or at least 4-6 times a week. About 41% of all respondents stated that are regular shoppers (i.e. they shop at least once a week). On the other hand, about 35% of all respondents reported shopping trip frequency of fortnightly or monthly with 22% of respondents saying that they shop less than once a month in the city centre.

Table 5.7 Frequency of visits to the city centre for non-food shopping for all users (n = 895) and car users only (n = 240)

Frequency	All Users (%)	Car Users (%)
Daily	7.4	7.5
4-6 times a week	3.1	1.7
2-3 times a week	12.0	7.1
Weekly	19.4	19.1
Fortnightly	16.8	16.7
Monthly	18.9	19.2
Less than once a month	22.1	28.7

Similar percentages of car users and all users reported daily shopping trips to city centre (7.5%). A smaller percentage of car users than all users reported frequent shopping trips to city centre (1.7% of car users), while a higher percentage of shoppers reported less frequent shopping trips. These frequencies were then combined into three categories: very frequent, frequent and infrequent.

5.2.1.2 Gender of the respondent

The gender of respondents is another relevant variable to trip generation and has been included in the model analysis in this study.

Table 5.8 below represents the percentage distribution of the respondents according to gender. From the table, 57.2 % of those in the Household Survey are female and 42.8% are male.

Table 5.8 Gender of the respondents (n = 884)

Gender	Percentage
Female	57.2
Male	42.8

5.2.1.3 Age of the respondent

According to age, the respondents are divided into three groups as shown in Table 5.9: 29.2% of respondents are in the age group of 16-34. About 35% of respondents are in age group 35- 54% while 35.1% of them are in the age group of 55 and over. This factor has also been included in the trip generation models.

Table 5.9 Age of the respondent (n = 884)

Age Group	Percentage
16-34	29.2
35-54	35.7
55 and over	35.1

5.2.1.4 Car ownership of the respondent's household

Car ownership is one of the important variables in most trip generation models. From the survey data (Table 5.10) it appears that over one third of the respondents' households own no car (35.9%), 40.8% have one car and 23.3% own two or more cars. It should be noted here that those who own no cars would be expected to make more shopping trips to the city centre than those with cars because of the cost of parking, the traffic congestion time spent searching for a parking space, etc.

Table 5.10 Car ownership of the respondent' household (n = 884)

Car Ownership	Percentage
0	35.9
1	40.8
2+	23.3

5.2.1.5 Mode of transport for shopping into the city centre

As shown in Table 5.11, public transport is the main mode of transport for travelling into the city centre for shopping, with nearly 60% taking the bus or the train. However, about 27% of shoppers drive to the city centre and 15% walk. As discussed later on, it was found that those who drive are the most likely to reduce their shopping trips to the city centre if congestion charging was introduced, while those who use public transport show the least change in trip frequency.

Table 5.11 Normal mode of transport into the city centre (n = 895)

Mode of Transport	Percentage
Public transport (bus, train and taxi)	57
Car/van	26.6
Walking and cycling	15.8
Other	1.0

5.2.1.6 Factors affecting the level of accessibility of city centre for shopping trips

People were also asked in the survey to report on the reasons which would encourage them to shop more in the city centre. Table 5.12 shows respondents' preferences for various transport policies. Of the respondents surveyed, more than 38% of car users stated that cheaper and/ or more accessible parking spaces would encourage them to do more shopping trips into the city centre, while about 15% stated that public transport improvements would encourage more shopping trips. Only 7% of car users considered traffic congestion to be a major problem for them. Of those who do not visit the city centre for shopping (300 respondents), 16% stated that improved parking prices and accessibility would encourage more shopping trips to the city centre, while only 7% thought traffic congestion was a major concern. From the results, access to cheaper/ easier parking and good public transport seem to be important to encourage more shopping in the city centre. t is also clear that traffic congestion in the city centre does not appear to represent a major problem to the majority of users.

Table 5.12 Factors that would encourage people to shop in the city centre more often

Response Option (transport related)	Total (n = 1199) %	Car users (n = 240) %	PT users (n = 510) %	Those who don't visit (n = 300) %
Cheaper/free parking	11.2	27.9	5.7	10.0
More car parking / easier parking	11.0	27.1	4.9	10.7
Better public transport	12.7	14.6	13.7	11.0
Less traffic congestion	7.9	7.1	8.8	7.0
Nothing	47.0	35.4	46.3	57.3

Adopted from ECCM (2004).

5.2.1.7 Impact of the congestion charge and transport improvements

Interviewees were asked firstly to express their stated intention for their future shopping trip generation and the impacts of introducing congestion charging only in Edinburgh. Also they were asked to express the perceived impacts on shopping trips if congestion charging was combined with each of the following: a) improved public transport into and out of the city and b) more park and ride facilities would be provided and would be situated outside the cordon. This was to investigate the preferences of the users and to optimise any possible investments for improving the transport system. Table 5.13 shows the reported results from the survey.

It is clear from the table that the transport improvements, and public transport improvements in particular, would have a marked effect on people's spending compared to the baseline scenario (congestion charge with no transport improvement), with sizeable proportions of people saying that they would spend more in the city centre and a dwindling of the percentages of those who would have spent less or gone elsewhere. This shows the significance of these transport infrastructure improvements to people living in Edinburgh and the surrounding areas in encouraging them to visit the city centre on a more frequent basis. However about 80% of respondents said that they would visit the city centre and spend the same amount of money if congestion charging is introduced with slightly lower percentage when public transport is improved.

Table 5.13 Impact of the congestion charge with and without transport improvements (n = 895)

Response Option	With No Transport Improvement %	With Improved PT %	With More Park & Ride Facilities %
Would visit the city centre and spend more	1.6	15.5	9.2
Would visit the city centre and spend less	10.3	3.4	3.1
Would visit elsewhere	4.9	1.1	1.5
Would visit the city centre outside the charge period	2.3	1.1	0.9
Would visit the city centre and spend the same	80.7	78.1	84.8
Don't know/No answer	0.4	0.9	0.8

Source: ECCM (2004)

Table 5.14 replicates these results for those that normally travel into the city centre by car only. Compared with the results with no transport improvement, transport improvements would have a considerable impact on car users' decision to shop in the city centre, with fewer saying that they would visit the city centre and spend less and much more saying they would visit the city centre and spend more, or spend the same.

5.2.2 Shoppers' survey

In the Shoppers' Survey, respondents were asked to express the reasons for being in the city centre and to report on the frequency of non-food shopping trips in the city centre per week. The analysis of these two questions is given in the following sections.

Table 5.14 Impact of the congestion charge with transport improvements – those that travel in normally by car for shopping only (n = 238)

Response Option	With No Transport Improvement %	With Improved PT %	With More Park & Ride Facilities %
Would visit the city centre and spend more	0.8	17.2	11.3
Would visit the city centre and spend less	23.9	9.7	8.4
Would visit elsewhere	13.0	3.4	3.8
Would visit the city centre outside the charge period	6.3	4.2	3.4
Would visit the city centre and spend the same	56.3	64.3	72.3
Don't know/No answer	0.4	1.3	0.8

Source: ECCM (2004)

5.2.2.1 Reasons for being in the city centre

Table 5.15 provides details of the purpose of the journeys they were observed to be doing when they were interviewed in the Shoppers' Survey. Of all the people interviewed in the Shoppers' Survey (n=1000), just over 20% of them were shopping in the city centre for groceries or other items. Many people (28%) were in the city centre because they worked there and over one-third (34.4%) were visiting the city. If only the people from Edinburgh and Fife were included (n=624), 28.4% of them were for shopping purposes and about 40% worked there. Of the 208 shoppers who were shopping in the city centre, 132 of them who answered all the questions in the survey are used in the study and the following sections present some general statistics about them.

5.2.2.2 Gender of the shoppers

Of the 132 shoppers, 78.8% of them are female, while only 21.2% are male.

Table 5.15 Reasons for being in the city centre (n = 1,000)

Reason	All Respondents (n=1,000) %	Those from Edinburgh and Fife (n=624) %
Shopping for groceries	3.3	5.0
Shopping for other items	17.5	23.4
Using services, such as bank, travel agents, restaurant etc.	5.1	7.5
Passing through/ window shopping	5.0	6.7
Work	28.2	39.3
Visiting Edinburgh for the day	12.4	6.7
Visiting Edinburgh as a tourist (includes an overnight stay)	22.0	2.2
Other	6.9	9.6

5.2.2.3 Car ownership of the shoppers' households

Of the 132 shoppers, 54.5% do not own a car, while 30.3% own one car and only 15.2% of them own two or more cars.

5.2.2.4 Age group of the shoppers

25.0% of the shoppers are in age group 16-25, 37.9% are in 26-54 and 37.1% of them belong to age group 55 and more years old.

5.2.2.5 Expenditure per non-food shopping trip

Table 5.16 presents the expenditure of the non-food shopping trip at the city centre on the day of the interview. About 29.5 percent of the shoppers did not spend any money while one half of them spent over thirty pounds.

5.2.2.6 Shopping trip frequency

15.9 % of those in the Shoppers' Survey shop in the city centre daily or at least 4-6 times a week. 54.5% are regular shoppers (at least once a week) while only 17.4% of them shop less than once a month. Table 5.17 shows the frequency of visits.

Table 5.16 The expenditure per non-food shopping trip (n =132)

Expenditure (pounds)	Percent
0	29.5
1-30	20.5
31-90	27.3
over 91	22.7

Table 5.17 Frequency of visits to the city centre for non-food shopping in the shoppers' survey (n =132)

Frequency	Percent
Daily	7.6
4-6 times a week	8.3
1-3 times a week	38.6
Fortnightly	10.6
Monthly	17.4
Less than once a month	17.4

5.2.2.7 Investigation of public opinions of the public transport services

Respondents in the survey were asked to evaluate current public transport services to and from the city centre. As it is shown in Table 5.18, 58.3% of shoppers thought public transport services are very good or good and 12.3% thought the service is poor or very poor. About 30% of the respondents thought the service was adequate.

Table 5.18 Opinion of current public transport services (n=132)

Response Option	Percentage
Very good	18.9
Good	39.4
Adequate	29.5
Poor	6.1
Very poor	6.1

5.3 SUMMARY

This chapter describes the three surveys and some general analysis of each of them. The data from each survey is used in a different application as discussed in later chapters of this thesis. National Travel Survey data (the commuting trips of the household in a day) were used in Chapters 6, 7 and 8 to model the commuting trips of the household using different techniques of trip generation. The three techniques of trip generation (linear regression analysis, category analysis and logistic analysis) - are used to calibrate the models and the results from the estimations are compared. The Household Survey data in Edinburgh is used in Chapter 9 to model the frequency of non-food shopping trips of an adult in a week to the city centre in Edinburgh. Linear and logistic regression analyses are used to investigate how social-economic factors, transport policy factors and hence transport accessibility would affect the shopping trips to the city centre. The stated preference data from the Household Survey was also used to model and investigate impacts of introducing congestion charging in the city centre of Edinburgh, in Chapter 10. Finally Shoppers' Survey data were used to investigate the impact of perceived accessibility of transport on shopping trips in Chapter 11

CHAPTER 6 METHODOLOGY FOR MODELLING TRIP GENERATION USING LOGISTIC REGRESSION

6.1 INTRODUCTION

Chapter 5 described the three datasets which have been used for trip generation modelling: the National Travel Survey (NTS) data, the Household Survey (HS) and the Shoppers' Survey (SS) in Edinburgh. In this chapter, the methodology for modelling trip generation using logistic regression is firstly explained. Then, the NTS data are used to calibrate trip generation models for work trips using three techniques of logistic regression analysis, and these are: binary logit, multinomial logit and nested logit models. The results are assessed and compared with other models in the next chapters.

6.2 THE DATA SET

It should be reported here that initially the analysis was carried out using work trips per household in a weekday (Monday) for the modelling using three years NTS data (1996-1998). That data contained 5,125 households' records which had at least one worker in the household. However, that dataset represented the whole of the UK and therefore large variations in household characteristics were found. For example variations in car ownership, income, household structures between different regions and between urban and rural areas and other factors which affect overall average trip rates. For these reasons the resulting models were, mostly, very insignificant in terms of their statistical performance (i.e. the t-values and the overall statistical significance of the models).

Therefore, a new data set was acquired in order to improve the models' estimation. This was made possible by the release of an extra national travel survey data set; that is data for the years 2002 to 2004. Furthermore, this data was then disaggregated by geographical areas as presented in Table 6.1. An urban area is usually considered to be an area that is relatively built up and its residents are usually regarded as being town or city dwellers. Based on the definition of urban areas at the time of the 2001 Population Census we took that an urban area was any continuously built-up area, of at least 20 hectares and with at least 1,500 residents.

To obtain more homogeneous data with fewer variations, only urban areas of 50,000 –250,000 residents were used. Each household should have at least one worker in order to be selected for the analysis and investigation of trip generation models. That is, a total of 2,706 households were selected for the analysis which represents 19.2% of the total households in the NTS dataset. It should be noted here that although the selection of this dataset should reduce the variability observed in household characteristics, it will not eliminate all variations since these urban areas are spread out through the whole of the UK, and thus will have various types of households with different characteristics.

Table 6.1 NTS data (2002-2004) and types of areas

Geographical Areas	Number of Households	Percent
Inner London	643	4.6
Outer London b/u area	1,315	9.3
West Midlands b/u area	516	3.7
G. Manchester b/u area	570	4.0
W. Yorkshire b/u area	321	2.3
Glasgow b/u area	102	0.7
Liverpool b/u area	162	1.1
Tyneside b/u area	182	1.3
Urban over 250K	1,807	12.8
Urban 100K to 250K	**1,703**	**12.1**
Urban 50K to 100K	**1,003**	**7.1**
Urban 25K to 50K	1,010	7.2
Urban 10K to 25K	1,674	11.9
Urban 3K to 10K	1,100	7.8
Rural	1,983	14.1
Total	14,091	100.0

For the work on this chapter, 73.1% of the sample (1,979 households) were randomly selected and used to calibrate the models by each of the three techniques. The calibrated models were then used to predict trip generation for the rest of the 26.9% of the data set (i.e. 727 households). The factors considered in this analysis include the number of full time and part time workers in the household, car ownership, household income, number of company cars and number of children in the household. As discussed in the next section. These are some of the typical variables which have been used previously in the literature for modelling trip generation (see for example Ortúzar and Willumsen, 2001).

6.2 DESCRIPTION OF THE VARIALBES

In this section, a trip generation model for work trips per household in a typical working weekday (in this case Wednesday is selected) is calibrated using logistic regression analysis. The descriptions of the variables which are used in these models are given in Table 6.2. As shown in the table, the variables include the number of workers (full time and part time) in the household, car ownership, household income, the number of children and number of company cars in the household.

Table 6.2 Description of variables used in work trip generation models

Variables	Description
WORKER_FT	A continuous variable: describes the number of full-time workers in the household (see Section 5.2.2).
WORKER1_FT	A dummy variable: takes the values of 1 if there is one full-time worker in the household, 0 otherwise.
WORKER2+_FT	A dummy variable: takes the values of 1 if there are two or more full-time workers in the household, 0 otherwise.
WORKER_PT	A continuous variable: describes the number of part-time workers in the household (see Section 5.2.2).
CA_WORKER_PT	A dummy variable: takes the values of 1 if there part time workers in the household, 0 otherwise (included in MCA_3).
WORKER2+	A dummy variable: takes the values of 1 if there are two or more full time/part time workers in the household, 0 otherwise (included in MCA_3).
CAR	A continuous variable: describes the number of cars in the household.

CAR1	A dummy variable: takes the values of 1 if the household owns one car, 0 otherwise.
CAR2+	A dummy variable: takes the values of 1 if the household owns two or more cars, 0 otherwise.
COM_CAR	A dummy variable: takes the values of 1 if the household has one or more company cars, 0 otherwise.
INCOME_MH	A dummy variable: takes the values of 1 if the annual household income is £20,000-£39,999, or £40,000 and over.
CHILD	A continuous variable: describes the number of children in the household.

In general, the number of workers in the household is expected to have a positive relationship with the number of work trips in a trip generation model. The number of full time workers in the household was tested as a continuous variable (WORKER_FT) and as two dummy variables to represent the three categories of full time workers in the household (0, 1, and 2 or more full time workers in the household). The number of part time workers is entered as a continuous variable (WORKER_PT). Car ownership and household income have also been included in the models and are expected to have positive impacts on the trips to work.

The annual household income is a relevant and important variable in the analysis and prediction of household trip generation models. In this analysis household income has been tested as a dummy variable (Medium/High) to represent two income groups (Low or Medium/High).

Car ownership was tested as both a continuous variable and as a dummy variable in the models. In the multinomial logit and nested logit models, the number of cars was entered as a continuous variable.

The number of children (CHILD) is included as a continuous variable in the trip generation models. This variable is expected to have a negative impact on the number of work trips. Finally, the availability of company cars is included as a dummy variable which takes a value of 1 if the household has one or more company cars and 0 otherwise. The variable is expected to have a positive sign in the models.

The following section summarises the methodology for using the logistic regression analysis to model trip generation. The logit models have been used to predict the probabilities of making a certain number of trips (i.e. trip frequency) in a certain time period which would allow the calculations of the number of trips generated in each household as discussed in the following sections.

6.3 THE METHODOLOGY FOR MODELLING TRIP GENERATION USING LOGISTIC REGRESSION TECHNIQUES

In this section, the appropriateness of using logistic analysis modelling for trip generation is investigated. The probabilities of a household making *j* work trip(s) are modelled using the typical independent variables often used in trip generation models. Three different types of logistic regression models are calibrated in this section: three binary logit models, one multinomial logit (MNL) model and one nested logit (NL) model. The methodology of how to model trip generation using each of the three modelling approaches is discussed below. The models are analysed and compared in terms of statistical significance and their prediction of trip generation in later sections.

6.3.1 Modelling trip generation using binary logit models

6.3.1.1 Model specifications

As discussed earlier in Section 4.4.2, binary logit analysis is suitable to model individual level choice data, when two alternatives are available. Typically, the dependent variable is a choice while the independent variables are relevant factors which may affect that choice. In choice situations where the dependent variable is a discrete one, the process is straightforward. In trip generation analysis however, where the dependent variable is the trip generation, the model structure is different in this case.

Here we assume that the dependent variable is a binary variable to represent the household making work trips or not. In the models, alternative 1 represents no work trips in a household per day and alternative 2 represents one or more work trips in the household per day. This seems to be a logical manner to represent trip making using a binary logit modelling specifications.

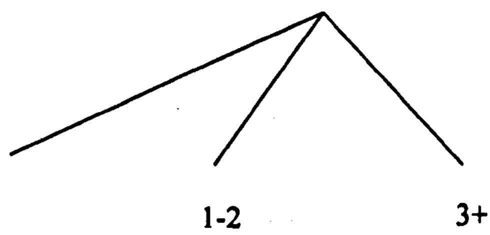
6.3.1.2 Utility function

The binary logistic regression models are calibrated as shown in Table 6.6. The variables used in these models are the number of workers in the household, car ownership, household income, the number of children in the household and the number of company cars in the household. In the first model (BLM_1) the number of full time workers and the number of cars in the household were included as continuous variables. In the second model (BLM_2) the number of full time workers in the household has been included as a continuous variable and number of cars as two dummy variables to represent the three levels of car ownership (0, 1, and 2+ cars). In the third model BLM_3 the number of full time workers in the household has been included as three dummy variables to represent the three levels of number of full time workers (0, 1, and 2+ workers). On the other hand, the number of cars in the household was treated as a continuous variable. In all the three models, alternative 1 was used as the reference, hence its utility $V_1 = 0$.

6.3.2 Modelling trip generation using multinomial logit (MNL) model

6.3.2.1 Specification of the model

The multinomial logit (MNL) model is one of the most popular choice models and it is used to analyse individual choices when the dependent variable is a discrete multi criteria variable which relates to a number of independent variables. In modelling trip generation using the MNL model, we assume that the probability of a household making a certain number of work trip(s) is a function of a number of independent variables. In this research, a number of trials for the structure of the model and for the allocation of variables to each utility have been carried out. The best fit of the models was obtained with the trips assigned as follows: {0 trips, 1-2 trips, 3 or more trips}, with the structure presented in Figure 6.1. This is compatible with Daly (1997) in his pioneering work on improved methods for trip generation which states that the change from 0 to making a trip (or more) is the most crucial choice, and the choices of making more than 1 trip are less important, which would suggest that the best structure is that such as in the stop-go mode as adopted in this analysis.



MNL model structure

Figure 6.1 The structure for the MNL trip generation model

The following logistic formula has been used for the MNL model:

$$P(Y=j|x)=\frac{e^{g_j(x)}}{\sum_{k=0}^{3+} e^{g_k(x)}} \tag{6.1}$$

$$g_k(x)=\beta_{k0}+\beta_{k1}x_x+\beta_{k2}x_2+... \tag{6.2}$$

Where

$P(Y=j|x)$ is the probability of household making j work trip(s), j = 0, 1-2, and 3+;

$g_k(x)$ is the utility equation of $j=k$;

x_1's include the number of workers (full time and part time), car ownership, household income, number of company cars, and the number of children as described in Table 6.1. Table 6.3 below shows the trip frequency distributions of the households.

Table 6.3 Trip frequency distributions

Trip Frequency	Number of Households	Percentage of Households	Percentage of Households in Accumulated Categories
0	354	17.9	25.0
1	140	7.1	
2	778	39.3	39.3
3	145	7.3	35.7
4	352	17.8	
5+	210	10.6	
Total	1,979	100	100

The results of the MNL model estimates which give the most statistically significant results are presented below. As shown in Figure 6.1, in the MNL model, the options are structured as 0, 1-2 and 3+ work trips per household.

6.3.2.2 The utility functions of the MNL models

The utility functions of the alternatives in the model are presented in Table 6.4. In the MNL model, the option '0 trips' has been assigned as the reference case ($V_0 = 0$ in Table 6.4). The number of full time workers in the household has been treated as a continuous variable which is included in the utility functions of options 2 and 3, each with an alternative specific coefficient. The number of part time workers in the household, number of children and number of cars have been treated as continuous variables and are included in the utility function of option 3. The availability of a company car and income has been treated as dummy variables and are included in the utility function of option 3. The income variable represents two categories (low and medium/ high).

Table 6.4 Utility functions for the MNL models

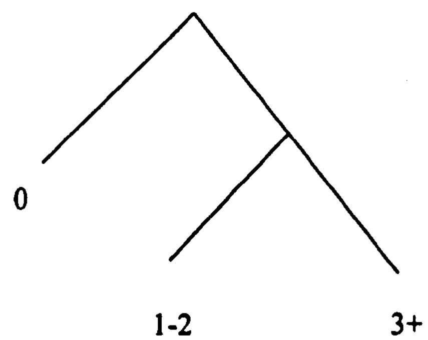
Utility Function	Variables	Coefficients to be Estimated
$V_0 = 0$ $V_{1-2} = \theta_{1-2} + \theta_{1-2}^{worker\_ft} WORKER\_FT$ $V_{3+} = \theta_{3+} + \theta_{3+}^{worker\_ft} WORKER\_FT$ $+\theta_{3+}^{worker\_pt} WORKER\_PT$ $+\theta_{3+}^{income\_mh} INCOME\_MH$ $+\theta_{3+}^{child} CHILD + \theta_{3+}^{car} CAR$ $+\theta_{3+}^{com\_car} COM\_CAR$	WORKER_FT WORKER_PT CHILD INCOME_MH CAR COM_CAR	θ_{1-2}, θ_{3+}, $\theta_{1-2}^{worker\_ft}$, $\theta_{3+}^{worker\_ft}$, $\theta_{3+}^{worker\_pt}$, $\theta_{3+}^{income\_mh}$, θ_{3+}^{child}, θ_{3+}^{car}, $\theta_{3+}^{com\_car}$

6.3.3 Nested logit (NL) model

6.3.3.1 Model specifications

When the IIA property of MNL is violated (i.e. when there are shared unobserved components associated with different choices or alternatives, the utilities of the elements of the corresponding multidimensional choice set cannot be independent), the modeller should consider alternative specifications such as the nested logit or multinomial probit models. Multinomial probit is an extension of probit models to more than two alternatives. Unfortunately, they are difficult to estimate when the number of alternatives is more than two. The nested logit model on the other hand allows subsets of alternatives to share unobserved components of utility, while using the MNL modeling specifications.

A nested logit model was also calibrated with the nested structure shown in Figure 6.2. In this case, trip makers are being assumed to be trading off between making no trips against making 1 or more trips. Then, at the second level, a trade off between 1-2 trips against 3 or more trips is assumed.



NL model structure

Figure 6.2 The structure for the NL trip generation model

6.3.3.2 The utility function

The utility functions of the alternatives in the NL model are presented in Table 6.5. As shown in the table, the number of full time workers in the household has been treated as a continuous variable which is included as the only common attribute inside the nest alternatives (that is the options of making 1 or more trips, see Figure 6.1). The number of part time workers in the household (continuous variable), number of cars (continuous variable), income (dummy variable) and availability of company cars (dummy) are all included as attributes that vary inside the nest and are included in the utility function of option 3. The number of children has been treated as a continuous variable and is included as the only attribute in the option at the higher level of the nested structure (i.e. making 0 trips). The results of the calibration for the NL model, i.e. the coefficient estimates, the t-values, the initial and final likelihood, the ρ^2 and the logit utility parameters are presented in the following section.

Table 6.5 Utility functions for the NL model

Utility Function	Variables	Coefficients to be Estimated
$V_0 = \theta_0^{child} CHILD$		
$V_{1-2} = \theta^{worker\_ft} WORKER\_FT$	WORKER_FT WORKER_PT CHILD INCOME_MH CAR COM_CAR	θ_{3+}, θ_0^{child}, θ^{worker_ft}, $\theta_{3+}^{worker\_pt}$, $\theta_{3+}^{income\_mh}$, θ_{3+}^{child}, θ_{3+}^{car}, $\theta_{3+}^{com\_car}$
$V_{3+} = \theta_{3+} + \theta^{worker\_ft} WORKER\_FT + \theta_{3+}^{worker\_pt} WORKER\_PT + \theta_{3+}^{income\_mh} INCOME\_MH + \theta_{3+}^{car} CAR + \theta_{3+}^{com\_car} COM\_CAR$		

6.4 RESULTS OF MODELING TRIP GENERATION USING LOGISTIC REGRESSION

6.4.1 Binary model

The results obtained from the calibration of the binary logistic regression models are shown in Table 6.6.

Table 6.6 Logistic regression model of work trip generation in a household

Variables (option)	Coefficient (t-test)		
	BLM_1	BLM_2	BLM_3
Constant (2)	-	-0.054 (-0.3)	-0.394 (-1.8)
WORKER_FT(2)	1.045 (10.0)	1.067 (8.2)	-
WORKER1_FT(2)	-	-	1.588 (7.7)
WORKER2+_FT(2)	-	-	2.254 (8.3)
WORKER_PT(2)	0.181 (1.7)	0.202 (1.7)	0.313 (2.3)
CHILD(2)	-0.159 (-2.9)	-0.161 (-2.8)	-0.201 (-3.4)
INCOME_MH(2)	0.325 (2.3)	0.335 (2.4)	0.300 (2.0)
CAR(2)	0.145 (1.6)	-	0.172 (1.8)
CAR1+(2)	-	0.224 (1.2)	-
COM_CAR(2)	0.301 (1.1)	0.338 (1.2)	0.248 (0.9)

Initial log-likelihood	-1371.7383	-1371.7383	-1371.7383
Likelihood constants only	-929.5130	-929.5130	-929.5130
Final log-likelihood	-840.6181	-840.9474	-836.8466
$\rho^2(0)$	0.3872	0.3869	0.3899
$\rho^2(c)$	0.0956	0.0953	0.0997
n	1,979	1,979	1,979

The options used in modelling:

1 = No work trip per household per day

2 = One or more work trips per household per day

From the table, the overall goodness of fit of these models is good with $\rho^2(0)$ being 0.3872, 0.3869 and 0.3899 respectively. However, some of the independent variables are not statistically significant at the 95% level of significance. For example the company car variable (COM-CAR) which might be due to correlation with income. However, it is decided to keep this variable in the model since it is a relevant one and also it shows statistical significance in the other models (i.e. linear regression and MCA_3 models). It should also be noted here that there might be a problem in the statistical significance of some of the variables because the proportion of households who are making 0 trips in a typical working day in the sample is much lower than that that are making one or more trips (see Table 6.3).

From Table 6.6, it can be seen that the number of workers and car ownership have positive impacts on households making work trips (positive coefficients of WORKER and CAR in utility two). Similarly, number of company cars in the household also has a positive coefficient in the model, as expected, although it has a lower t-value. The household income has a positive impact on households making work trips. On the other hand, number of children in the household has a negative impact on work trips as expected (negative coefficients of CHILD).

To further investigate the results from these models, the relative importance of each variable is obtained. The mean value (*m*) of each independent variable is calculated from the survey data (i.e. the average value of each variable). The mean value is then multiplied by the coefficient of the corresponding variable to work out a relative importance value for each variable.

The mean values, the relative importance values (*m*coefficient*) of BLM_1, BLM_2 and BLM_3 are presented in Table 6.7. It appears from this table that the number of workers in the household is one of the most important variables in the model. That is, relative values of 1.345 and 1.374 are obtained in the table below for models BLM_1 and BLM_2 respectively. In model BLM_3 a combined value of over 1.538 is resulted from both categories of the dummy variable representing number of full time workers in the household. Car ownership, income (BLM_1 and BLM_2) and number of children (BLM_2) come next as the most relatively important variables. Of the three binary logit models, BLM_3 has the best $\rho^2(0)$ and will be used in Section 8.5 for model estimation and comparison.

Table 6.7 Relative importance of each variable in the binary logit models

Variables (option)	Relative Importance of Variables (m * coefficient)		
	BLM_1	**BLM_2**	**BLM_3**
Constant	-	-0.054	-0.394
WORKER_FT(2)	1.345	1.374	-
WORKER1_FT(2)	-	-	0.871
WORKER2+_FT(2)	-	-	0.767
WORKER_PT(2)	0.076	0.085	0.132
CHILD(2)	-0.109	-0.110	-0.138
INCOME_MH(2)	0.225	0.232	0.208
CAR(2)	0.198	-	0.235
CAR1+(2)	-	0.201	-
COM_CAR(2)	0.025	0.028	0.021

6.4.2 MNL model

The results of the calibration for the MNL model are presented in Table 6.8. As shown in the table, all the variables have the correct signs and are statistically significant at the 95% level of significance with $\rho^2(0)$ being 0.215.

Table 6.8 MNL model of work trip generation in a household

Variables	0 trip	1-2 trips	3+ trips
Constant	-	0.286 (2.4)	-3.198 (-14.4)
WORKER_FT	-	0.681 (6.2)	2.258 (15.5)
WORKER_PT	-	-	1.134 (9.8)
CHILD	-	-	-0.334 (-5.4)
INCOME_MH	-	-	0.488 (3.2)
CAR	-	-	0.264 (3.1)
COM_CAR	-	-	0.443 (2.3)
Initial log-likelihood		-2174.154	
Log-likelihood with Constants only		-2042.140	
Final log-likelihood		-1706.560	
$\rho^2(0)$		0.215	
$\rho^2(c)$		0.164	
N		1,979	

Table 6.9 below shows the relative importance of each variable for each category of number of trips. From the table it appears that the variables used in this model are statistically significant and have impacts on the number of trips. However, the constant is also statistically significant and has a relatively high important role in prediction. A likelihood ratio test shows that the model with all the independent variables is more statistically significant than the model with constant only, i.e. -2*(-2042.140-(-1706.560)) = 671.16 > 14.067. Similar to the previous models, the number of full time workers has an important role to play in the prediction of number of trip generation in this model.

Table 6.9 The relative importance of each variable in the MNL model

Variables	0_1 trip	2 trips	3+ trips
Constant	-	(0.286)	(-3.198)
WORKER_FT	-	0.877	2.907
WORKER_PT	-	-	0.478
CHILD	-	-	-0.229
INCOME_MH	-	-	0.338
CAR	-	-	0.361
COM_CAR	-	-	0.037

6.4.3 Nested Logit Model

The results of the calibration for the NL model, i.e. the coefficient estimates, the t-values, the initial and final likelihood, the ρ^2 and the logit utility parameters are presented in Table 6.10.

Table 6.10 NL model of work trip generation in a household

Variables	0 trip	1-2 trips	3+ trips
Constant	-	-	-2.044 (-13.6)
WORKER_FT	-	1.021 (3.6)	
WORKER_PT	-	-	0.267 (3.1)
CHILD	0.156 (2.8)	-	-
INCOME_MH	-	-	1.107 (8.3)
CAR	-	-	0.558 (7.5)
COM_CAR	-	-	0.246 (1.4)
Theta		0.978 (5.0)	
Initial log-likelihood		-2174.154	
Final log-likelihood		-1851.105	
$\rho^2(0)$		0.149	
$\rho^2(c)$		0.094	
N		1979	

From Table 6.10, all the variables of the model are statistically significant at 95% level (except the company car variable as discussed before) and have the expected signs.

Table 6.11 shows the relative importance of each of the variables using the calibrated NL model. Similar to the results obtained from the MNL model, it appears that as in the previous models the number of full time workers has the highest relative importance amongst all independent variables, followed by income and car ownership.

Table 6.11 The relative importance of each variable in the NL model

Variables	0 trip	1-2 trips	3+ trips
Constant	-	-	(-2.044)
WORKER_FT	-	1.314	
WORKER_PT	-	-	0.113
CHILD	0.107	-	-
INCOME_MH	-	-	0.767
CAR	-	-	0.763
COM_CAR	-	-	0.021

6.5 DISCUSSIONS OF THE RESULTS AND SUMMARY

One of the main aims and novelties of this research has been to develop a methodology for adopting logistic regression analysis to model trip generation. The methodology for modelling trip generation using the three logistic modelling approaches has been explained in this chapter. Trip generation has been successfully modelled using the binary, MNL and NL modelling approaches and the results obtained are both statistically significant and logical.

While three modelling approaches provided an appropriate way to model trip generation for work trips in this analysis and their results were all statistically significant, the MNL structure performed much better than the nested logit model. This might be because mathematically, the nested structure allows subsets of alternatives to share unobserved components of utility, to overcome the problem of violating the IIA property in the MNL model. Because of the limited data available in this research, it was not very straight forward to identify shared or common unobserved components of the utilities. To further assess the results obtained from using logistic analysis in modeling trip generation, these results need to be compared with results obtained from conventional trip generation models.

Therefore, in Chapter 7 the NTS data are used to calibrate trip generation models for work trips using the conventional trip generation model; that is the linear regression and category analysis including multiple classification analysis

(MCA). The analysis and comparisons of all model results are also presented in Chapter 8. The performance of the trip generation models using logistic regression is compared with the conventional trip generation models (i.e. linear regression and category analysis).

Moreover, it should be noted that considering the overall performance of the model, the NL (as shown in Table 6.10) model does not make any improvements to the MNL model (as shown in Table 6.8) with $\rho^2(0)$ a reduction from 0.215 to 0.149. The theta parameter has an acceptable value of 0.978 which suggests that the MNL is most appropriate in this case.

CHAPTER 7 MODELLING WORK TRIP GENERATION USING CONVENTIONAL MODELLING APPROACHES

7.1 INTRODUCTION

Chapter 6 described the estimation of trip generation models using logistic regression. In order to be able to assess these models, they should be compared with trip generation estimates from conventional models. These are linear regression models and category analysis models. It is often argued in the literature that the linear regression models are superior to the category analysis results because of the known limitations of the later (Ortúzar and Willumsen, 2001). Techniques of multiple classification analysis however provide significant improvements of the results of trip generation over the classical category analysis (Guevara and Thomas, 2007). Therefore, an extensive investigation and analysis of the data using multiple classification analysis techniques has been conducted in this chapter to include the up to date methodological development in this method.

In this chapter, the NTS data are used to calibrate and compare trip generation models for work trips using linear regression analysis and category analysis techniques including multiple classification analysis.

The same data set which was used for the calibration of the logistic models (Section 6.2) are also used in the current analysis. The same variables which were used as independent variables in the logistic regression models were also used in the linear regression. Three linear regression models have been calibrated and compared in the following sections.

For the classical category analysis and the two multiple classification analysis (MCA) models, only income, car ownership and total number of workers in the household have been included in the models to maintain a manageable number of categories. For the third MCA model however, there was no problem with having as many categories as needed.

7.2 LINEAR REGRESSION ANALYSIS

In this section, a trip generation model for work trips per household in a typical working weekday (in this case Wednesday is selected) is calibrated using linear regression analysis. The descriptions of the variables which are used in the linear and logistic regression models are given in Table 6.2. As shown in the table, the variables include the number of workers (full time and part time) in the household, car ownership, household income, the number of children and number of company cars in the household.

Similar to the discussion in Section 6.2 of the expected impacts of independent variables on the dependent variable are discussed here. The number of full time workers in the household was tested as a continuous variable (WORKER_FT) as well as two dummy variables representing the three categories of full time workers in the household (0, 1, and 2 or more full time workers in the household). The number of workers in the household is expected to have a positive relationship with the number of work trips. The number of part time workers is entered as a continuous variable (WORKER_PT). Car ownership and household income have also been included in the models and are expected to have positive impacts on the trips to work. Household income has been tested as a dummy variable (Medium/High) to represent two income groups (Low or Medium/High). Finally, car ownership was tested as both a continuous variable and as a dummy variable in the models.

The number of children (CHILD) is included as a continuous variable in the trip generation models. This variable is expected to have a negative impact on the number of work trips. Finally, the availability of company cars is included as a dummy variable which takes a value of 1 if the household has one or more company cars and 0 otherwise. The variable is expected to have a positive sign in the models.

Three linear regression models have been calibrated from this data. In the first model, the number of full time workers, number of part time workers, and number of children were included as continuous variables. Income, car ownership and number of company cars in the household were included as dummy variables. The second model is similar to the first model, except that number of cars was tested as a continuous variable. In the third model the number of full time workers was tested as dummy variables to represent the three levels of number of full time workers (0, 1, and 2+ workers) and the rest of the variables are similar to the second model.

Table 7.1 shows the coefficient estimates and the t-values for the linear regression models estimated from the data set as discussed in the earlier section. All the models include the number of full time and part time workers in the household, car ownership, household income, the number of children in the household and the number of company cars in the household.

Table 7.1 Linear regression models of work trip generation by a household

Variables	LM-1	LM-2	LM-3
Constant	0.162 (1.2)	0.154 (1.6)	0.159 (1.1)
WORKER_FT	1.394 (24.5)	1.331 (22.4)	-
WORKER1_FT	-	-	1.269 (9.1)
WORKER2+_FT	-	-	2.657 (16)
WORKER_PT	0.803 (12.2)	0.752 (11.2)	0.692 (9.4)
CHILD	-0.203 (-5.7)	-0.204 (-5.7)	-0.229 (-6.2)
INCOME_MH	0.214 (2.4)	0.170 (2.0)	0.209 (2.3)
CAR	-	0.199 (3.8)	0.294 (5.4)
CAR1+	0.140 (1.2)	-	-
COM_CAR	0.358 (2.8)	0.290 (2.2)	0.242 (1.8)
R^2	0.322	0.326	0.272
n	1979	1979	1979

From the table, it appears that all the variables have the correct signs and most of them are statistically significant at the 95% level of significance. The R^2 values of the three models are 0.322, 0.326 and 0.272 respectively, which are reasonable. It should be noted here that the most significant R^2 value here is obtained in the model which has continuous variables for the number of full time workers and the number of cars in the household (LM-2). Therefore, this model will be the selected linear regression model to be used later on in Section 8.5 for the prediction and comparisons of trip generations using the three techniques.

The signs of the coefficients for full time and part time workers are positive as expected. As the number of each of these types of workers increases, households are observed to make more work trips. In fact the number of workers seems to be a statistically significant variable in all models; as a continuous variable and also as dummy variables. As the number of cars in a household increases, households are expected to make more work trips (positive coefficients of CAR in model LM-2 and LM-3). The dummy variables for car ownership (CAR1) in model LM-1 are not statistically significant at 95% level; this might be due to a possible correlation with income. The variable representing the presence of company cars in the household has a positive impact on households making work trips (positive and statistically significant coefficient of COM_CAR in model LM-1 and LM-2). The variable representing the presence of children has a negative impact on households making work trips (negative coefficient of CHILD) as expected. As expected, household income has a positive impact on work trips and is statistically significant in all the three models.

Similar analysis to that in Section 6.4.1, relative importance of variables is carried out for these models as well. This is worked out by multiplying the mean value by the coefficient of the corresponding variable and elasticities (i.e. the percentage change in the dependent variable with respect to a given percentage change in the relevant independent variable) have been carried out. The elasticity analysis is carried out as follows:

$$Elasticity = \left(\frac{T-T_0}{T_0}\right) \Big/ \left(\frac{m-m_0}{m_0}\right)$$

In the linear regression model, if only one independent variable changes, the change in the dependent variable ($T-T_0$) with respect to the change in the independent variable $(m-m_0)$ can be expressed as $(m-m_0)*coefficient$. So the above elasticity equation becomes:

$$Elasticity = \left(\frac{(m-m_0)*coefficient}{constant+\sum m_0*coefficient}\right) \bigg/ \left(\frac{m-m_0}{m_0}\right)$$
$$= \frac{m_0*coefficient}{constant+\sum m_0*coefficient}$$

The mean values, the relative importance values (*m*coefficient*) and elasticities of LM-1, LM-2 and LM-3 are presented in Table 7.2, Table 7.3 and Table 7.4 respectively.

Table 7.2 The relative importance of each variable in LM-1

Variables	LM-1 Coefficients	Mean Values of Variables (m)	Relative Importance of Variables (m * coefficient)	Elasticity
Constant	0.162 (1.2)	-	(0.162)	
WORKER_FT	1.394 (24.5)	1.288	1.795	0.729
WORKER1_FT	-	0.548	-	-
WORKER2+_FT	-	0.340	-	-
WORKER_PT	0.803 (12.2)	0.422	0.339	0.138
CHILD	-0.203 (-5.7)	0.684	-0.139	-0.056
INCOME_MH	0.214 (2.4)	0.693	0.148	0.060
CAR	-	1.368	-	-
CAR2+	0.140 (1.2)	0.898	0.126	0.051
COM_CAR	0.358 (2.8)	0.084	0.030	0.012
TOTAL	-	-	2.461	

In all the three models, the estimates for the constants (0.162, 0.154 and 0.159) are compared to the estimates for the rest of the variables in the model. As shown in Table 7.2, in LM-1, the number of full time and part time workers has the largest importance relative to the rest of the variables. In LM-2, the number of workers (full time and part time), number of children and car ownership has the most significant importance in the model (see Table 7.3). Finally in LM-3, Table 7.4 shows that income and the presence of company cars have relatively least importance in the model while the rest of the variables are more significant (e.g.

number of workers, car ownership and number of children has the most impacts on work trip generation).

Table 7.3 The relative importance of each variable in LM-2

Variables	LM-2 Coefficients	Mean Values of Variables (m)	Relative Importance of Variables (m * coefficient)	Elasticity
Constant	0.154 (1.6)	-	(0.1540	
WORKER_FT	1.331 (22.4)	1.288	1.714	0.696
WORKER1_FT	-	0.548	-	-
WORKER2+_FT	-	0.340	-	-
WORKER_PT	0.752 (11.2)	0.422	0.317	0.129
CHILD	-0.204 (-5.7)	0.684	-0.140	-0.057
INCOME_MH	0.170 (2.0)	0.693	0.118	0.048
CAR	0.199 (3.8)	1.368	0.272	0.111
CAR1+	-	0.898	-	-
COM_CAR	0.290 (2.2)	0.084	0.024	0.010
TOTAL	-	-	2.460	

Table 7.4 The relative importance of each variable in LM-3

Variables	LM-3 Coefficients	Mean Values of Variables (m)	Relative Importance of Variables (m * coefficient)	Elasticity
Constant	0.159 (1.1)	-	(0.159)	
WORKER_FT	-	1.288	-	-
WORKER1_FT	1.269 (9.1)	0.548	0.696	0.283
WORKER2+_FT	2.657 (16)	0.340	0.904	0.367
WORKER_PT	0.692 (9.4)	0.422	0.292	0.119
CHILD	-0.229 (-6.2)	0.684	-0.157	-0.064
INCOME_MH	0.209 (2.3)	0.693	0.145	0.059
CAR	0.294 (5.4)	1.368	0.402	0.163
CAR1+	-	0.898	-	-
COM_CAR	0.242 (1.8)	0.084	0.020	0.008
TOTAL	-	-	2.461	

From all the above results and discussions it appears that the most significant R^2 (the R^2 values of the three models are 0.322, 0.326 and 0.272 respectively) value here is obtained in the model which has continuous variables for the number of full time workers and the number of cars in the household (LM-2). Therefore, this model will be the selected linear regression model to be used later on in Section 8.5 for the prediction and comparisons of trip generations using the three techniques.

7.3 CATEGORY ANALYSIS / CROSS-CLASSIFICATION

7.3.1 Category Analysis - the classical model

The second model of trip generation in this study is category analysis or cross-classification model. As discussed in Section 2.3.2, category analysis is based on estimating the trip production rates per household for a given purpose as a function of household attributes. The method's basic assumption is that trip generation rates are relatively stable over time for certain household stratifications. Therefore the art of this method is in defining the categories although it is well recognised that it is not very easy to choose the best categorisations of the selected variables (see Ortúzar and Willumsen, 2001 for more discussions).

The NTS data has been used to carry out this analysis. Three variables have been identified to be included in the analysis: household income with three categories (see Table 7.5), car ownership with two categories (≤ 1, and 2+ cars) and the number of workers (including both full time and part time workers) in the household with two categories (1, and 2+ workers). These are the three most commonly used factors in studies of category analysis (see Wootton and Pick 1967 for more discussions on category analysis). It should be noted here that although more variables have been included in the regression analysis, it was deemed not very practical to use any more variables in this analysis since the number of categories would have increased radically. Extensive trials and errors procedures have been used to choose the best combinations or categorisations of the selected variables and their levels. In total this yields 12 categories of households as shown in Table 7.6. This categorisation has been adopted for the basic category analysis model as well as the MCA_1 and MCA_2 (see Sections 7.3.2 and 7.3.3). However, for the MCA_3 model, further categorisation of the data has been used (Section 7.3.4).

Table 7.5 Household income groups

Code	Household Income
A	Less than £19,999
B	£20,000 - £39,999
C	£40,000+

Table 7.6 Number of households in each category

No. of Workers	No. of Cars	Household Income		
		A	B	C
1	≤1	403	248	37
	2+	55	89	25
2+	≤1	96	258	134
	2+	53	243	338

In Table 7.6, of the 12 categories, 11 of them have more than 30 observations and only one of them has less than 30 observations (1 worker, 2+ cars, high income group). The lower number of observations of this category is due to the fact that there are fewer households with one worker owning two or more cars and with very high income, which is a common problem in category analysis models.

Despite all the efforts to construct best groupings of categories, it is still clear from the table that there are some variations between the categories. For example, for households with 1 worker and 2+ cars there is generally lower number of households in each income group than in other categories. Since the trip rate depends on the number of households in each category as well as on the number of trips made by each household, these variations will have impacts on the average number of trips or the trip rates for each household calculated using the category analysis method. In other words, these impacts might result in overestimation in some cases and underestimation in other cases of the trip rates and/or the total number of trips for these categories.

The work trip rates per household for each household category have been worked out from the NTS data as usual, i.e. the average trip rate within any specific category is equal to the observed number of trips in that specific category of households divided by the number of households in that category, or in equation format (Ortúzar and Willumsen, 2001):

$$t(h) = T(h) / H(h)$$

Where:

$t(h)$ is the trip rate per household by category h;

$T(h)$ is the total number of trips in cell h; and

$H(h)$ is the number of households in cell h.

Table 7.7 below presents the work trip rates by households' categories. It appears that in general the trip rate progression is logical and as expected (with the exception of a couple of cells indicated with a '*'). That is, in most of the cells the trip rates increase as income increases on one hand and as car ownership and number of workers per household increase on the other hand. It should be noted here that the cells indicated with a '*' have lower numbers of observations as discussed earlier.

Table 7.7 Work trip rates by households' categories (trips/HH/day)

No. of Workers	No. of Cars	Household Income		
		A	B	C
1	≤1	1.397	1.746	1.351*
	2+	1.345*	1.798	2.040
2+	≤1	2.281	2.783	3.216
	2+	2.660	3.305	3.630

(*) Trip rate which does not logically follow on with the rest of the table.

As well documented in the literature (see Stopher and McDonald, 1983; Ortúzar and Willumsen, 2001, etc) and as discussed earlier, in category analysis, unduly large samples are usually required in order to guarantee good reliability of the models. In addition, the unequal number of records in each cell could also lead to inefficient estimation of trip rates. In these cases the cell values will vary in reliability because of differences in the numbers of households being available for calibration. To overcome this possible problem a number of enhanced approaches known as multiple classification analysis have been applied and reported in the literature (Stopher and McDonald, 1983; Ortúzar and Willumsen, 2001; SECTRA 1998; Clark, 1996 and Guevara and Thomas, 2007). These approaches estimate the cell values based on a grand mean derived from the entire data set, and two or more class means which are derived from all data in each class relevant to the cell in question.

Three Multiple Classification Analysis (MCA) approaches which have been documented in the literature have been tested to investigate their impacts on trip generation estimation and are referred to here as MCA_1, MCA_2 and MCA_3. Some background discussions of these approaches are given in Section 2.3.2, but see also Guevara and Thomas (2007) for a very thorough discussion of these approaches. The three approaches have been applied in this study for estimating trip generation using the same data set used in the regression analysis methods, as discussed below.

7.3.2 Multiple Classification Analysis-1 (MCA_1)

As discussed, the method is based on estimating the cell values from a grand mean derived from the entire data set, and two or more class means which are derived from all data in each class relevant to the cell in question. In equation form the trip rate in each cell is calculated as follows:

$$\hat{t}_{twm} = \hat{t} + \left(\hat{t}_t - \hat{t}\right) + \left(\hat{t}_{wm} - \hat{t}\right)$$

Where

$\hat{t}_{iwm}$ is the trip rates for a income-worker-car category (iwm); $\hat{t}$ is the total average; $\hat{t}_i$ is the average number of trips of households of income i; $\hat{t}_{wm}$ is the average number of trips of households of worker w and car m.

Table 7.8 shows the trip rates which resulted from applying MCA_1.

Table 7.8 Work trip rates by household categories (trips/HH/day) using MCA_1

No. of Workers	No. of Cars	Household Income		
		A	B	C
1	≤1	0.702	1.582	2.354
	2+	0.868	1.748	2.520
2+	≤1	1.985	2.865	3.636
	2+	2.606	3.486	4.257

From the table, it is clear that the trip rate patterns produced from this approach are logical and positively proportionate to the increase in income, car ownership and number of workers per household. The problem of not having a sufficiently large number of observations in each cell of the classical category analysis method, i.e. as in Table 7.7 has apparently been overcome by using this analysis. However, as reported in the literature (Guevara and Thomas 2007) the results from this method might still have a problem of trips overestimation that occurs at the higher income groups (income group C in Table 7.8) and underestimation that occurs at the lower income groups of household categories (income group A category). See also Table 8.3 and Table 8.6 and the discussion in Section 8.5 later in the section which shows that some of the estimated trip rates vary by about 40% difference from the observed values.

7.3.3 Multiple Classification Analysis-2 (MCA_2)

To estimate the household work trip rates using the MCA_2, a weighted average factor is applied to correct for the biases which result from the unequal number of observations by each category (see more discussion of the method in Section 2.3.2). The trip rate for each category is calculated using the following formula adopted from Guevara and Thomas (2007):

$$\hat{t}_{iwm} = \sum_{h=1}^{H}\sum_{wm}^{W,M}\left(\frac{H_{wm}}{H}\right)1_{iwm}^{h}v^{h}\Big/H_{i} + \sum_{h=1}^{H}\sum_{i}^{I}\left(\frac{H_{i}}{H}\right)1_{iwm}^{h}v^{h}\Big/H_{wm} - \sum_{h=1}^{H}v^{h}\Big/H$$

Where

$1_{iwm}^{h}=1$ if household h belongs to category iwm and zero otherwise;

W, M, I and H correspond to the total number of worker clusters, car clusters, income clusters and household respectively;

H_{wm} corresponds to the number of households of worker w and car m;

H_{i} corresponds to the number of households of income i;

v^{h} corresponds to the observed trips generated by household h.

This method has also been applied to the same data set and the results have been compared and assessed. Table 7.9 shows the trip rates which resulted from applying MCA_2. From the table, it is clear that the trip rate patterns produced from this approach are also logical and positively proportionate to the increase in income, car ownership and number of workers per household. The problem of having an unequal number of observations in each cell has been partly overcome by using this analysis. In addition, the trip rate estimates from this method seem to slightly overcome the problem of overestimation and underestimation that occurs at the higher/lower income groups as discussed in Section 6.3.2 above.

Table 7.9 Work trip rates by household categories (trips/HH/day) using MCA_2

No. of workers	No. of cars	Household income		
		A	B	C
1	≤1	1.087	1.577	1.672
	2+	1.279	1.769	1.864
2+	≤1	2.301	2.791	2.885
	2+	2.749	3.239	3.334

7.3.4 Multiple Classification Analysis-3 (MCA_3)

Finally, the third modified method MCA_3, which is also well illustrated by Guevara and Thomas (2007), has been applied to the data in order to investigate and compare the resulting trip rates from applying this approach. The method is based on estimating the household trip rates using least square regressions where the independent variables are all dummy variable; one for each of the categories of the strata variables. It should be noted here that in this model, unlike the classic category analysis model, it was decided to use as many cells as there are that could be tested for groupings similar to the variables used in regression analysis. The equation used here, adopted from that of Guevara and Thomas (2007) is:

$$v_h = \beta_0 + \sum_{i \neq 1} \beta_i 1_i^h + \sum_{w\_ft \neq 0} \beta_{w\_ft} 1_{w\_ft}^h + \sum_{w\_pt \neq 0} \beta_{w\_pt} 1_{w\_pt}^h + \sum_{m \neq 0\_1} \beta_m 1_m^h$$
$$+ \sum_{com\_car \neq 0} \beta_{com\_car} 1_{com\_car}^h + \sum_{child \neq 0} \beta_{child} 1_{child}^h + \varepsilon'^h$$

Where β_0, β_i, β_{w_ft}, β_{w_pt}, β_m, β_{com_car}, and β_{child} are coefficients to be estimated; ε'^h is the error.

Table 7.10 The coefficients of Multiple Classification Analysis-3

Variables	MCA_3
Constant	0.394 (2.8)
WORKER1_FT	1.266 (9.1)
WORKER2+_FT	2.697 (15.7)
CA_WORKER_PT	0.835 (9)
CAR2+	0.297 (3.6)
COM_CAR	0.228 (1.7)
CHILD	-0.520 (-6.7)
INCOME_M	0.256 (2.7)
INCOME_H	0.346 (2.9)
R^2	0.263
n	1979

Table 7.11 Work trip rates by HH categories (trips/HH/day) using MCA_3

No. of PT workers	No. of FT workers	No. of cars	No. of company car	No. of children	Household income A	Household income B	Household income C
0	0	≤1	0	0	0.394	0.650	0.740
0	0	≤1	0	1+	-0.126	0.130	0.220
0	0	≤1	1	0	0.622	0.878	0.968
0	0	≤1	1	1+	0.102	0.358	0.448
0	0	2+	0	0	0.691	0.947	1.037
0	0	2+	0	1+	0.171	0.427	0.517
0	0	2+	1	0	0.919	1.175	1.265
0	0	2+	1	1+	0.399	0.655	0.745
0	1	≤1	0	0	1.660	1.916	2.006
0	1	≤1	0	1+	1.140	1.396	1.486
0	1	≤1	1	0	1.888	2.144	2.234
0	1	≤1	1	1+	1.368	1.624	1.714
0	1	2+	0	0	1.957	2.213	2.303
0	1	2+	0	1+	1.437	1.693	1.783
0	1	2+	1	0	2.185	2.441	2.531
0	1	2+	1	1+	1.665	1.921	2.011
0	2	≤1	0	0	3.091	3.347	3.437
0	2	≤1	0	1+	2.571	2.827	2.917
0	2	≤1	1	0	3.319	3.575	3.665
0	2	≤1	1	1+	2.799	3.055	3.145
0	2	2+	0	0	3.388	3.644	3.734
0	2	2+	0	1+	2.868	3.124	3.214
0	2	2+	1	0	3.616	3.872	3.962
0	2	2+	1	1+	3.096	3.352	3.442
1+	0	≤1	0	0	1.229	1.485	1.575
1+	0	≤1	0	1+	0.709	0.965	1.055
1+	0	≤1	1	0	1.457	1.713	1.803
1+	0	≤1	1	1+	0.937	1.193	1.283
1+	0	2+	0	0	1.526	1.782	1.872
1+	0	2+	0	1+	1.006	1.262	1.352
1+	0	2+	1	0	1.754	2.010	2.100
1+	0	2+	1	1+	1.234	1.490	1.580
1+	1	≤1	0	0	2.495	2.751	2.841
1+	1	≤1	0	1+	1.975	2.231	2.321
1+	1	≤1	1	0	2.723	2.979	3.069
1+	1	≤1	1	1+	2.203	2.459	2.549
1+	1	2+	0	0	2.792	3.048	3.138
1+	1	2+	0	1+	2.272	2.528	2.618
1+	1	2+	1	0	3.020	3.276	3.366
1+	1	2+	1	1+	2.500	2.756	2.846
1+	2+	≤1	0	0	3.926	4.182	4.272
1+	2+	≤1	0	1+	3.406	3.662	3.752
1+	2+	≤1	1	0	4.154	4.410	4.500
1+	2+	≤1	1	1+	3.634	3.890	3.980
1+	2+	2+	0	0	4.223	4.479	4.569
1+	2+	2+	0	1+	3.703	3.959	4.049
1+	2+	2+	1	0	4.451	4.707	4.797
1+	2+	2+	1	1+	3.931	4.187	4.277

The resulting trip rates from MCA_3 are presented in Table 7.11. From the table, it is clear that the trip rate patterns produced from this approach are also logical and positively proportionate to the increase in income, car ownership and number of workers per household.

Table 7.12 below shows the number of observed trips as well as the predictions using category analysis and the three MCA models. It also shows the overall percentage differences and the Residual Sum of Squares (RSS) for each model. From the table, it appears that the MCA_2 model produces the lowest overall differences between the predicted and observed number of trips. However, when considering the RSS of each model prediction the MCA_3 model appears to give the best results. Only the results obtained from the basic category analysis and MCA_3 models will be used in the final comparisons of the predictions of the models in Section 8.5 below. It should be noted here that a further model, called MCA_4 (Guevara and Thomas, 2007) has also been developed but not used in this study. Category Analysis is the conventional category analysis technique and this is taken as the base for the analysis of RSS in the table below. From the table, it seems that MCA_1 produces the largest sum of errors in the family of category analysis (11.1% higher than that obtained from the base CA technique). The MCA_2 does not provide any improvement of the RSS (0.1%) while the MCA_3 produces the least RSS values (-7.7%) than the base CA method. Therefore, the MCA_3 has been recommended to be used as the best technique in this family.

Table 7.12 Comparison of work trips estimated by CA and MCAs

Models	Work Trips		Difference (%)	RSS	RSS – Diff from CA %
	Predicted	Observed			
CA	1,785	1,726	59 (3.42)	1,904	--
MCA_1	1,786		60 (3.48)	2,116	11.1%
MCA_2	1,673		-53 (-3.07)	1,905	0.1%
MCA_3	1,790		64 (3.71)	1,758	-7.7%

7.3.5 Summary of the section

An extensive amount of analysis and modelling of trip generation using category analysis and the most up to date approaches of multiple classification analysis (three methods) have been carried out in this chapter. Firstly, the basic category analysis approach has been implemented. The resulted trip rates were in general logical and as expected. However, few of the resulted trip rates were illogical and did not follow the expected trend in trip rate progression. That is, trip rates increase as income increases on one hand and as car ownership and number of workers per household increase on the other hand. Three improved multiple classification analysis approaches have been tested to investigate their impacts on trip generation estimation. The first method which is based on estimating the cell values from a grand mean derived from the entire data set, and two or more class means which are derived from all data in each class relevant to the cell in question. The trip rate patterns produced from this approach are logical and positively proportionate to the increase in income, car ownership and number of workers per household. The problem of not having large number of observations in each cell in the classical category analysis method has been overcome by using this analysis. However, the results from this method still have a problem of trips overestimation/ underestimation that occurs at the higher/lower income groups.

To estimate the household work trip rates using the MCA_2, a weighted average factor is applied to correct for the biases which result from the unequal number of households in each category. The trip rates for each category were calculated which were logical and positively proportionate to the increase in income, car ownership and number of workers per household. The problem of having unequal number of observations in each cell was overcome by using this analysis. Finally, the third modified method MCA_3 has been applied to the data to investigate the resulting trip rates from applying this approach. The method is based on working out estimation of the household trip rates by estimating least squares regressions with the independent variables being all dummy variable; one for each of the categories of the strata variables. The trip rates resulting from this method are found to be superior to the values obtained by MCA_2.

7.4 DISCUSSIONS OF THE RESULTS

In this chapter, firstly trip generation models using conventional approaches have been calibrated using data from the National Travel Survey (NTS). These are the linear regression analysis and category analysis including the up to date methodological development of this approach (i.e. multiple classification analysis). Three linear regression models, a category analysis model and three multiple classification analysis models have been calibrated. In linear regression analysis, LM-2 (which includes number of workers and car ownership as continuous variables as well as number of part time workers, number of children, availability of company car and HH income) has shown the best performance amongst the linear regression models. This model is therefore selected to be used in the analysis and comparisons of model performance in Section 8.5. In multiple classification analysis, the MCA_3 (see Section 7.3.4) has shown the best performance amongst the approaches of this technique. Therefore, in the final analysis and comparisons of the models, results from category analysis and MCA_3 have been included. The results show that the most significant R^2 for the linear regression models (R^2 = 0.326) obtained in the model which has continuous variables for the number of full time workers and the number of cars in the household (LM-2). Therefore, this model will be the selected linear regression model to be used later on in Section 8.5 or the prediction and comparisons of trip generations using the three techniques.

From the analysis of the category analysis results, it appears that MCA_1 produces the largest sum of errors in the family of category analysis (11.1% higher than that obtained from the base CA technique). The MCA_2 does not provide any improvement of the RSS (0.1%) while the MCA_3 produces the least RSS values (-7.7%) than the base CA method. Therefore, the MCA_3 has been recommended to be used as the best technique in this family.

CHAPTER 8 PREDICTION OF TRIP GENERATION USING THE CALIBRATED MODELS

8.1 INTRODUCTION

As mentioned previously, the aim of this work has been to investigate the appropriateness of using logistic regression in trip generation modelling. The methodology adopted to apply these techniques (i.e. binary, MNL and NL models) to modelling trip generation as well as the results of models have been explained and presented in Chapter 6. Chapter 7 calibrated trip generation models using conventional techniques (i.e. linear regression and category analysis). In this chapter, the prediction of trip generations using all the calibrated models in Chapters 6 and 7 are analysed and compared.

About 73.1% of the NTS data set was used to calibrate each of the above models while the remaining 26.9% of the data was left as a validation sample to predict trip rates using the calibrated models as discussed in Chapter 6. The prediction techniques of the trip generation using each of the approaches (logistic regression, linear regression and category analysis) are discussed below. A comparison of the estimated predictions using each of the three approaches is then discussed in Section 8.5.

8.2 PREDICTION OF TRIP GENERATIONS USING LOGISTIC REGRESSION

To use the binary logit model for prediction, an overall weighted average of the trips ($\bar{j}$) is calculated. This weighted average of the trips is obtained using the total number of trips made by all households who make at least one work trip divided by the number of the households. In this case:

$$\bar{j} = \frac{1*140+2*778+3*145+4*352+5*58+...+10*10+11*1+11*2}{140+778145+352+58+93+24+18+4+10+1+2} = 2.997$$

Then this overall weighted average is multiplied by the probability of making 1+ work trips in the household which will give the expected number of work trips per household.

When using the MNL and NL models in prediction, to calculate the expected number of work trips per household T, a summation of the j trip(s) multiplied by their corresponding probabilities is carried out as below:

$$T = \sum_{j=0}^{3+} j * P(Y = j)$$

The categories used for trip frequencies in the MNL and NL models are 0, 1-2, and 3+ work trips per household as discussed above. The trip frequencies and their corresponding number of trips in the data set used for model calibration are shown in Table 8.1.

Table 8.1 Trip frequency distributions

Trip frequency	Number of households	Trip frequency	Number of households
0	354	7	24
1	140	8	18
2	778	9	4
3	145	10	10
4	352	11	1
5	58	12	2
6	93	Total	1,979

In this case, the number of households who make 1 trip is 140 and the number of households making 2 trips is 778, and so on. Therefore, for j=1-2, the weighted average number of trips is calculated as below:

$$\bar{j} = \frac{1*140 + 2*778}{140 + 778} = 1.847 \qquad \text{for j=1-2}$$

Similarly, for j=3+, the weighted averages number of trips are calculated as below:

$$\bar{j} = \frac{3*145+4*352+5*58+...+10*10+11*1+11*2}{145+352+58+93+24+18+4+10+1+2} = 4.489 \qquad \text{for j=3+}$$

To use the NL model for estimation, we need to work out the probabilities of making j trips. These are worked out by firstly, computing the conditional probabilities from the lower nests (see Figure 6.2) as below:

$$P_{1-2/3+} = \frac{e^{V_{1-2}}}{e^{V_{1-2}} + e^{V_{3+}}}$$

$$P_{0/1+} = \frac{e^{V_0}}{e^{V_0} + e^{V_{1+}}}$$

where

$$V_{1+} = \beta X_t + \theta \ln\left(e^{\left(\frac{1}{\theta}\right)V_{1-2}} + e^{\left(\frac{1}{\theta}\right)V_{3+}}\right)$$

Then, the modelled probabilities of each option can be computed as the product of the marginal probability of choosing the composite alternative and the conditional probability of choosing the option in the lower nest:

$$P_0 = P_{0/1+}$$

$$P_{1-2} = P_{1-2/3+}(1 - P_0)$$

$$P_{3+} = (1 - P_{1-2/3+})(1 - P_0)$$

For logit models, the total number of estimated trips is then obtained by the summation of the expected work trips of each household.

8.3 PREDICTION OF TRIP GENERATIONS USING LINEAR REGRESSION

Using linear regression analysis for predicting trip generation is a straightforward process. Using the calibrated equations of the linear trip generation models, the total number of predicted trips was calculated for the 26.9 % of the data using the values of the independent variables. It should be mentioned here that the trip generation prediction in this section is based on the model (LM-2) estimates since it was the best model obtained as discussed.

8.4 PREDICTION OF TRIP GENERATIONS USING CATEGORY ANALYSIS

For the category analysis and MCA_1 and MCA_2 models, this data (i.e. the 26.9% of the NTS data set) was categorised into the same 12 categories as presented in Table 8.2 and was used to predict trip rates using the calibrated models (CA, MCA_1 and MCA_2) in order to assess their performance. For MCA_3, the data was categorised into 144 categories (see Table 7.11) and was used to predict trip rates using the MCA_3 model.

Table 8.2 Number of households in each category in 26.9% of the NTS

No. of Workers	No. of Cars	Household Income			Total
		A	B	C	
1	≤1	135	104	12	251
	2+	22	36	10	68
2+	≤1	28	111	49	188
	2+	14	95	111	220
Total		199	346	182	727

8.5 COMPARISONS OF THE TRIP PREDICTIONS USING THE THREE TECHNIQUES OF TRIP GENERATION

Trip generation predictions using the three models were then investigated and compared. Table 8.3 presents a comparison of the observed number of trips with the trip prediction for the 727 households using the three types of methods (i.e. linear regression analysis, category analysis and logit models). It should be mentioned here that the trip generation prediction using linear regression was based on model LM-2 since this was the best model obtained as discussed. The predictions using the basic category analysis method provide a basis for the comparisons as well as the results from MCA_3. In terms of the logistic regression the predictions using the three techniques are included (the binary logit model, the MNL model as well as the nested logit model). The results are presented in the following table.

Table 8.3 Comparison of work trips estimated by the three sets of models

Models	Work Trips		Diff (%)	RSS	RSS – Diff from MNL %
	Predicted	Observed			
LM_2	1,798	1,726	72 (4.2)	1,731	1.1
CA	1,785		59 (3.4)	1,904	11.2
MCA_3	1,790		64 (3.7)	1,758	2.6
BLM_3	1,798		72 (4.2)	2,037	18.9
MNL model	**1,795**		**69 (4.0)**	**1,713**	-
NL model	1,800		84 (4.9)	1,942	13.4

As shown in the table, the total numbers of work trips predicted by all the models are quite similar and similar to the observed number of work trips. However, that does not necessarily indicate perfect predictions by the models. For example when a higher prediction than the observed value is added up to a lower prediction than the observed value, the overall difference in this case might be misleading. So even if the predicted total is very close to the observed total, it does not necessarily indicate perfect prediction. Therefore the residual sum of squares is calculated to further investigate the results.

The Residual Sum of Squares (RSS) (or Error Sum of Squares) for each model is calculated in order to test for the accuracy of the models. RSS can be obtained as $\sum(y_i - \hat{y}_i)^2$, where y_i is the observed value and $\hat{y}_i$ is the predicted value. Table 8.3 presents the predicted against observed number of trips by each model, the overall % difference and the RSS in each case. Based on the percentage difference between the observed and predicted, it seems that the category analysis model produces the lowest overall differences between the predicted and observed number of trips. However, when considering the RSS of each model prediction, the results show that the least RSS values have been obtained from MNL model with a value of 1,713, making it outperforms all the other models (Table 8.3). This is followed by the linear regression model (LM-2) and lastly, the MCA_3 models with their RSS value 1.1% and 2.6% higher than that of the MNL model. The RSS results of conventional category analysis, the binary logit model and NL model are 11.2%, 18.9% and 13.4% greater than that of the MNL (the best performing model) respectively. While the MNL model shows best performance amongst the logistic regression models, the binary logit model shows worst results. This might be because of the aggregation of travellers into {making 0 trips or making 1 or more trips} categories and the fact that the number of travellers who are making 0 trips are very low in the sample.

In addition to the above comparisons, disaggregate validation tests by several market segmentations, including household income groups, car ownership levels and number of full time workers were conducted. Tables 8.4 – 8.6 below present the observed and predicted work trips per household by household income, car ownership and number of full time workers respectively.

Table 8.4 Observed and predicted work trips rates per household by household income

		Household Income						Total (n=727)	
		Income_L (n=199)		Income_M (n=346)		Income_H (n=182)			
		Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff
Observed		**1.698**		**2.405**		**3.055**		**2.374**	
Predicted	LM_2	1.626	-4.3	2.537	5.5	3.279	7.3	2.473	4.2
	CA	1.604	-5.5	2.512	4.5	3.281	7.4	2.456	3.5
	MCA_3	1.598	-5.9	2.520	4.8	3.298	9.0	2.462	3.7
	BLM_3	2.102	23.7	2.572	7.0	2.695	-12	2.474	4.2
	MNL	1.729	1.8	2.535	5.4	3.155	3.3	2.470	4.0
	NL	1.673	-1.0	2.673	11.1	2.984	-2.3	2.477	4.3

Table 8.5 Observed and predicted work trip rates per household by car ownership

		Car Ownership								Total (n=727)	
		0 (n=80)		1 (n=359)		2 (n=253)		3+ (n=35)			
		Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff
Observed		**1.875**		**2.164**		**2.597**		**4.057**		**2.374**	
Predicted	LM_2	1.718	-8.4	2.208	2.0	2.857	10.0	4.146	2.2	2.473	4.2
	CA	1.849	-1.4	2.142	-1.0	2.978	14.7	3.295	-18.8	2.456	3.5
	MCA_3	1.854	-1.1	2.169	0.2	2.916	12.3	3.586	-11.6	2.463	3.7
	BLM_3	2.141	14.2	2.424	12.0	2.611	0.5	2.759	-32.0	2.474	4.2
	MNL	1.888	0.7	2.242	3.6	2.808	8.1	3.686	-9.1	2.470	4.0
	NL	1.707	-9.0	2.257	4.3	2.896	11.5	3.470	-14.5	2.477	4.3

Table 8.6 Observed and predicted work trip rates per household by number of full time workers

		Number of Full Time Workers								Total (n=727)	
		0 (n=73)		1 (n=396)		2 (n=219)		3+ (n=39)			
		Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff	Trip Rate	% Diff
Observed		**1.151**		**2.010**		**3.055**		**4.538**		**2.374**	
Predicted	LM_2	1.109	-3.6	2.007	-0.2	3.293	7.8	5.164	13.8	2.473	4.2
	CA	1.613	40.2	2.112	5.0	3.211	5.1	3.292	-27.5	2.456	3.5
	MCA_3	1.115	-3.1	2.031	1.0	3.471	13.6	3.700	-18.5	2.463	3.7
	BLM_3	1.533	33.2	2.478	23.3	2.728	-10.7	2.764	-39.1	2.474	4.2
	MNL	1.337	16.2	2.072	3.1	3.247	6.3	4.261	-6.1	2.470	4.0
	NL	1.444	25.0	2.358	17.3	2.904	-4.9	3.218	-29.3	2.477	4.3

Firstly, it is observed that in general the accuracy of the predictions seems to improve with increasing category sample size. For example, the least differences of observed and estimated trip rates per household when analysed for the different income groups, are obtained for the medium income group which has the highest number of households (n=346) in Table 8.4. Higher differences between observed and estimated values are obtained when sample sizes are lower. Similarly, category 2 of car ownership which has the largest sample size (n=359) show the least difference between observed and estimated trip rates per household by car ownership between all car ownership categories (Table 8.5). The same observations are obtained when investigating the difference of observed and estimated trip rates per household by number of full time workers; as the sample size decreases the predictions become less accurate (Table 8.6).

Secondly, Table 8.4 show that the linear regression and category analysis, as well as MCA_3 model results underestimate values of work trips in relation to the observed values at lower income categories and overestimate values at the higher income categories with the exception of BLM, see also the discussion by Guevara and Thomas (2007).

Investigating the differences between observed and predicted work trip rates per household by car ownership categories shows a similar picture to that using income groups, which is an overestimation of trip rates at lower car ownership categories and underestimation at higher car ownership categories, except for the highest car ownership category which has a very small number of observations, which might have affected the accuracy of prediction in this category.

Finally, when investigating the observed and predicted work trip rates per household by number of full time workers it is clear that the small sample size of some categories affect the accuracy of prediction of that category.

As shown in the tables above and in Figure 8.1 and Figure 8.2, the predicted number of work trips per household by the LM-2, MNL and the MCA_3 models are the closest to the observed ones.

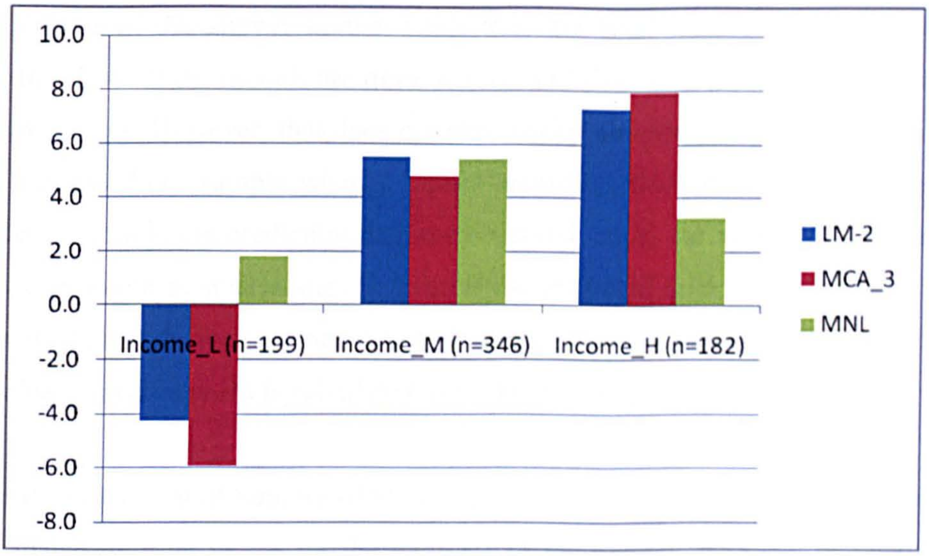


Figure 8.1 The percentage difference between observed and predicted work trip rates per household by household income

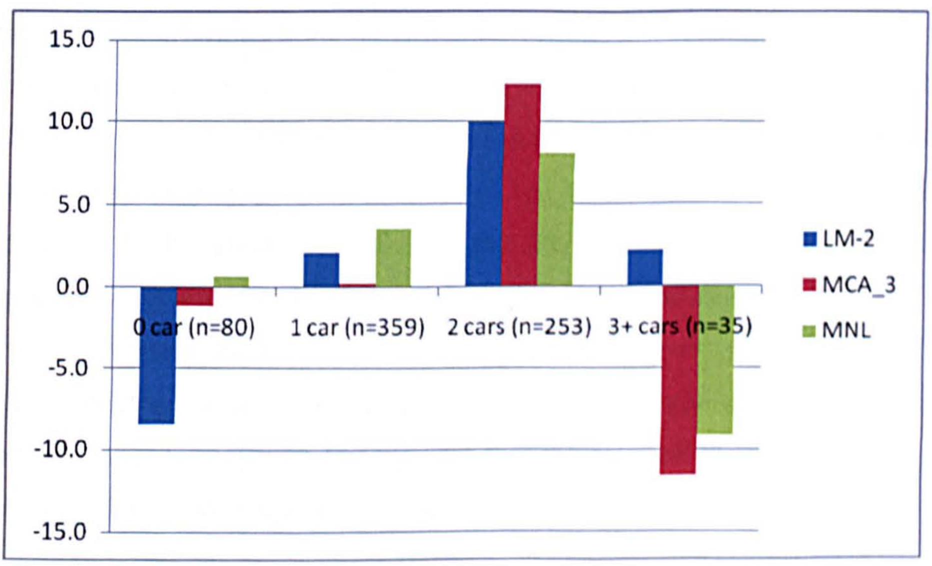


Figure 8.2 The percentage difference between observed and predicted work trip rates per household by car ownership

8.6 DISCUSSIONS OF THE RESULTS AND SUMMARY

In this chapter, the prediction of trip generations using all the calibrated models in Chapters 6 and 7 are analysed and compared. The resulting models are mostly statistically significant at 95% level with all the independent variables have the logical signs. As shown in the Table 8.3, the total numbers of work trips predicted by all the models are quite similar and similar to the observed number of work trips. However, that does not necessarily indicate perfect predictions by the models. For example when a higher prediction than the observed value is added up to a lower prediction than the observed value, the overall difference in this case might be misleading. So even if the predicted total is very close to the observed total, it does not necessarily indicate perfect prediction. Therefore the residual sum of squares is calculated to further investigate the results.

The Residual Sum of Squares (RSS) (or Error Sum of Squares) for each model is calculated in order to test for the accuracy of the models. RSS can be obtained as $\sum(y_i - \hat{y}_i)^2$, where y_i is the observed value and $\hat{y}_i$ is the predicted value. When considering the RSS of each model prediction, the results show that the least RSS values have been obtained from MNL model with a value of 1,713, making it outperforms all the other models (Table 8.3). This is followed by the linear regression model (LM-2) and lastly, the MCA_3 models with their RSS value 1.1% and 2.6% higher than that of the MNL model. The RSS results of conventional category analysis, the binary logit model and NL model are 11.2%, 18.9% and 13.4% greater than that of the MNL (the best performing model) respectively. While the MNL model shows best performance amongst the logistic regression models, the binary logit model shows worst results. This might be because of the aggregation of travellers into {making 0 trips or making 1 or more trips} categories and the fact that the number of travellers who are making 0 trips are very low in the sample.

In addition to the above comparisons, disaggregate validation tests by several market segmentations, including household income groups, car ownership levels and number of full time workers were conducted. Tables 8.4 – 8.6 present the observed and predicted work trips per household by household income, car ownership and number of full time workers respectively.

From the results of the models, it seems that in general the accuracy of the predictions seems to improve with increasing sample size of the category. The estimated trip generation rates for work trips are generally lower than the observed values at lower income categories and are overestimated at higher income categories. The only exception to this pattern is the estimations using the binary logit model which show reverse patterns. Other logistic regression models (i.e. the MNL models) show very moderate or small overestimation of work trips for all income groups, which constitutes an advantage of these models.

8.6.1 Potential improvements in trip generation modelling using logistic regression

One of the main objectives of this research has been to develop a methodology for adopting logistic regression analysis to model trip generation. The methodology for modelling trip generation using logistic regression is explained in Section 6.3. It is a considerable achievement to devise the methodology to use each of the three logistic modelling approaches to model trip generation.

Then, the NTS data are used to calibrate trip generation models for work trips using three techniques of logistic regression analysis, these are: binary logit, multinomial logit and nested logit models.

The ability to use logistic regression analysis to model trip generation would provide a way forward to overcome some of the strong assumptions implied by the other conventional techniques. For example, in linear regression analysis, the assumption of linearity of each of the independent variables with the dependent variables is a strong restrictive. Also, the lack of built-in upper and lower limits to the number of trips could potentially lead to unreasonable predictions, or could result in negative number of trips when the covariate values are relatively low. The assumption that the number of trips is approximately continuous can also be questioned especially where the number of trips are low. The lack of a behavioural justification in trip generation such as supported by the theory of random utility for example is also a drawback of this stage. All of these restrictions of linear regression techniques can be overcome by using logistic regression.

Although multiple classification analysis (MCA) methods provide improved techniques to overcome some of the shortcomings of category analysis approach, these methods are largely suffer from same limitations of category analysis. The use of logistic regression would provide a more flexible approach than MCA.

Logistic regression has been widely used to model other travel choices such as mode, route, departure time and other choices. However, not many applications in trip generation modelling have been reported. The problem is that typically in logistic regression analysis the dependent variable is a choice while the independent variables are relevant factors which may affect that choice. In choice situations where the dependent variable is a discrete one, the process is straightforward. In trip generation analysis however, where the dependent variable is the trip generation, the model structure is neither typical nor straight forward. The dependent variable has to be defined in a logical way as a probabilistic function of a number of independent variables.

8.6.2 Summary

The NTS data have been used to calibrate trip generation models for work trips using logistic regression, linear regression and category analysis and the results of model predictions are compared. The results provide strong evidence the appropriateness of using logistic regression analysis for trip generation modelling. Based on the RSS of each model prediction, it appears the results from the MNL model outperform that of all the other models. This is followed by the linear regression model (LM-2) and the MCA_3 model.

In addition, the results in this research support those obtained by Guevara and Thomas (2007) that MCA_1 method, which is most commonly used in applications of trip generation modeling, is the least accurate model in the family of MCA. MCA_2 method also produced no accurate results compared to MCA_3 which proved to be the most accurate method, and therefore should be recommended for use as the preferred category analysis method.

CHAPTER 9 MODELLING TRIP GENERATION WITH PARKING COSTS FOR SHOPPING TRIPS

9.1 INTRODUCTION

In Chapter 6, the NTS data was used to calibrate household work trip generation models using linear regression analysis, category analysis including multiple Classification Analysis (MCA) and logistic analysis. Parking costs are included in the models as a factor which is representing transport policies. In this Chapter the Edinburgh Household Survey (HS) data have been used to calibrate trip generation models for shopping trips also including parking costs. Models were calibrated using linear regression analysis and logistic regression analysis techniques. Logistic analysis techniques include binary logit, MNL and NL models. Results of modelling trip generations for different segments of the shoppers based on mode of travel are also presented.

The weekly non-food shopping trip frequencies in the household survey in the city centre were investigated. Firstly, the factors considered in the models are investigated in Section 9.2. Trip generation models are calibrated and presented in Section9.3 and Section9.4.

9.2 INVESTIGATION OF THE FACTORS AFFECTING SHOPPING TRIP GENERATION

Based on a general analysis of the survey data (Section 5.3), the following variables were defined as important factors which affect shopping trip generation:

1. Mode of travel into the city centre for non-food shopping: The mode of travel to the city centre for non food shopping trips is considered to be an important factor which affects the trip generation and its frequency. The different modes of travel were categorised into three groups (see Table 5.10): car or van, public transport (i.e. bus, train or taxi) and walking or cycling. This

categorisation is based on the fact that using the bus, train or the taxi to travel to the city centre would involve paying travel costs but not parking costs, while driving a private car/ van would involve paying parking cost but not fare. In addition, in the questionnaire, the bus, train and taxi costs were investigated as a one category. It should also be mentioned that there were only 5 respondents out of 884 in the survey data using taxis, therefore it seemed logical to exclude the taxi trips from the analysis (see Table 5.11 for the number of respondents in each category). Therefore, the train and bus were considered as one category in this study and referred to as public transport. While the private car/van was considered as a private mode.

2. Personal attributes: age, gender, car ownership and social grade. This set of socio economic variables has been widely investigated in the literature and identified for their impacts on trip generation (see Section 5.2 for discussion of the general analysis).

3. Location of residence: This variable has also been previously investigated in the literature and identified as an important variable to affect trip generation (see for example Sharpe *et. al.*, 1958, Goulias *et al.*, 1990, Cotrus *et al.*, 2005).
4. Characteristics of the transport system: These types of factors have generally been considered for their impacts on the mode choice *but not* on trip generation. In this study, accessibility of the transport system and its impacts on trip generation models has been identified as an under researched area. Therefore parking cost has been included to represent transport accessibility in the trip generation models. Parking cost is the only relevant variable in the data set which could have been used here to represent transport accessibility since the data set lacks level-of-service variables. Table 9.1 presents the variables that have been considered in this analysis.

Table 9.1 Description of the variables included in trip generation models

Variables	Description
CAR	A dummy variable: takes the value of 1 if the respondent normally travels into the city centre for non-food shopping by car or van, 0 otherwise.
PT	A dummy variable: takes the value of 1 if the respondent normally travels into the city centre for non-food shopping by bus or train, 0 otherwise.
CAR0	A dummy variable: takes the value of 1 if the respondent's household owns no car, 0 otherwise.
AGE1	A dummy variable: takes the value of 1 if the respondent's age is 16-34, 0 otherwise.
AGE2	A dummy variable: takes the value of 1 if the respondent's age is 35-54, 0 otherwise.
SOCI1	A dummy variable: takes the value of 1 if the respondent's social grade is upper middle class (A) or middle class (B), 0 otherwise.
SOCI2	A dummy variable: takes the value of 1 if the respondent's social grade is lower middle class (C1) or skilled worker (C2), 0 otherwise.
SOCI12	A dummy variable: takes the value of 1 if the respondent's social grade is upper middle class (A), middle class (B), lower middle class (C1) or skilled worker (C2), 0 otherwise.
LOCA1	A dummy variable: takes the value of 1 if the respondent's location is city centre, 0 otherwise.
LOCA2	A dummy variable: takes the value of 1 if the respondent's location is inter-cordon area, 0 otherwise.
LOCA12	A dummy variable: takes the value of 1 if the respondent's location is city centre or inter-cordon area, 0 otherwise.
GENDER	A dummy variable: takes the value of 1 if the respondent is a male, 0 female.
PARKCOST	A continuous variable: describes the parking cost of a non-food shopping visit travelling to the city centre.

It should be noted here that only those respondents who are over 16 years of age were included in the survey, so when both AGE1 and AGE2 are zero that would mean that the respondent falls in the age group of more than 54. Also, in the social grades there is no 'high class', so when both SOCI1 and SOCI2 are zero, it means that the respondent falls in the social class group of "unskilled worker".

Table 9.2 below shows the frequencies of shopping trips to the city centre by different modes of travel in the survey and the number of respondents in each category. As the trips considered here were non-food shopping trips to the city centre, those who walk (or cycle) are observed to make more frequent trips (an average of 2.25 shopping trips per week) than those by other modes (an average of 1.117 shopping trips per week for car users and 1.139 for public transport users). Therefore, in this case the private car/van and public transport modes are expected to have a relatively negative effect on the trip frequency.

Table 9.2 The weekly shopping trip frequencies to the city centre by mode

Mode of Transport	Average Frequency of Weekly Shopping Trips	Number of Respondents (%)
Car	1.117	237 (27.0)
Public transport	1.133	505 (57.5)
Walked/cycling	2.250	137 (15.6)
Total	1.307	879

People who live in the city centre and inter-cordon zone and who belong to the upper middle class and middle class are expected to make more shopping trips to the city centre than others in the same class who live outside the city centre. Parking cost is expected to have a negative impact on the trips by car. That is as the cost of parking increases the number of trips by car generated to the city centre would decrease.

From the survey data (see Chapter 5), it is clear that people in age group one (i.e. 16-34) were observed to make more shopping trips to the city centre than other age groups (see Table 9.3). Slightly more shopping trips were also observed for male than female (Table 9.4).

Table 9.3 Shopping trip frequencies to the city centre by age group

Age Group	Average Frequency of Weekly Shopping Trips	Number of Respondents (%)
16-34	1.522	258 (29.4)
35-54	1.291	315 (35.8)
55 and more	1.142	306 (34.8)
Total	1.307	879

Table 9.4 Shopping trip frequencies to the city centre by gender

Gender	Average Frequency of Weekly Shopping Trips	Number of Respondents (%)
Male	1.396	376 (42.8)
Female	1.241	503 (57.2)
Total	1.307	879

In the next sections the data from Edinburgh household survey is used to calibrate trip generation models.

9.3 LINEAR REGRESSION TRIP GENERATION MODELS

Trip generation models were calibrated using (i) logistic regression analysis techniques and (ii) linear regression analysis. As discussed earlier, the mode of travel to the city centre for a shopping trip has an influence on the trip generation of this trip. Therefore in this analysis, the mode of travel was firstly considered as a factor in the trip generation model for all the respondents (Model_1-a) for all users (including car, public transport, walking and cycling users). Secondly, a model with interaction effects of the location (LOCA1 & LOCA2) with the mode of travel (Car & PT) (Model-1-b). Separate models were then calibrated for each of the car users and public transport users since these two categories represent about 85% of all users. The modelling process was carried out using linear regression analysis then using logistic regression analysis. In all cases, the number of shopping trips to the city centre was modelled as a function of socio-economic variables, location, mode of transport used as well as some policy factors. The results were then discussed and compared.

Table 9.5 presents the estimated coefficients, their t-values, the R^2 and the number of observations in each model. As shown in the table, in total 879 observations were included in the analysis for the model for all users (Model-1-a & b), of which 237 were car users (Model-2) and 505 were public transport users (Model-3).

Table 9.5 Linear regression trip generation models

Variables	Coefficient (t-test)			
	Model-1-a (all users)	**Model-1-b (all users)**	**Model-2 (car users)**	**Model-3 (PT users)**
Constant	0.590 (2.0)	0.131 (0.7)	-0.238 (-0.7)	0.085 (0.4)
CAR	-0.344 (-1.4)	-	-	-
PT	-0.500 (-2.4)	-	-	-
CAR0	0.308 (2.1)	0.322 (2.2)	-	0.334 (2.0)
AGE1	0.430 (2.7)	0.446 (2.8)	1.009 (3.2)	0.445 (2.7)
AGE2	0.214 (1.4)	0.225 (1.5)	0.701 (2.5)	-
SOCI1	0.202 (1.1)	0.218 (1.2)	-	0.441 (2.0)
SOCI2	0.386 (2.3)	0.396 (2.4)	-	0.555 (2.9)
SOCI12	-	-	0.292 (1.0)	-
LOCA1	1.257 (6.8)	1.719 (7.9)	-	1.310 (5.6)
LOCA2	0.603 (4.0)	0.487 (3.0)	-	0.419 (2.5)
LOCA12	-	-	1.167 (5.0)	-
PARKCOST	-0.022 (-0.5)	-0.022 (-0.6)	-0.028 (-0.7)	-
CAR*LOCA1	-	-0.590 (-1.7)	-	-
CAR*LOCA2	-	0.585 (2.4)	-	-
PT*LOCA1	-	-0.430 (-1.5)	-	-
R^2	0.113	0.117	0.135	0.094
n	879	879	237	505

From the above table it appears that all coefficients have correct (i.e. as expected signs). However the values of R^2 are very low, suggesting that the relation might not be linear. The negative signs of the CAR and PT variables indicate that car users and public transport users make relatively less shopping trips as discussed above. In addition, from the table there are evidences to suggest that there are significant variations and differences between car users and public transport's users' attitudes and behaviour (different values of the coefficients).

It also appears that people living in the city centre and inter-cordon zone make more shopping trips to the city centre (the positive sign of the LOCA1, LOCA2 and LOCA12 variables in the three models (Model-1-a, Model 2 & Model 3). Moreover, from Model-1-b it is also clear that those who live in central locations (LOCA1) make less shopping trips by each of the car or public transport (-ve sign of CAR*LOCA1 and PT*LOCA1) in Model-1-b. On the other hand, those who reside outside the city centre tend to make more trips by car (+ve sign of CAR*LOCA2). People in the age group of 16-34 make more shopping trips to the city centre than people in other age groups (positive sign of AGE1 in all the models). People in the age group between 35 and 54 have a positive impact on making shopping trips to the city centre for all users and car users (positive sign of AGE2).

From the results it emerges that the upper middle class and middle class respondents make more trips (positive signs of SOCI1, SOCI2 and SOCI12).

PARKCOST is the only variable in the model which reflects impacts of transport policies as discussed earlier. The negative sign of the coefficient is logical and as expected. This is encouraging to suggest that more transport policy measures should be investigated and included in trip generation models.

To further analyse these results the values to indicate the importance of each variable (i.e. the product of the coefficient and the mean value of the variable as discussed in Section 7.2) and their elasticities have been calculated for three models (Model-1-a, Model-2 & Model-3) and presented in Table 9.6 below.

From the table it appears that for all users locations play an important role in the trip generation model. Also those who use public transport seem to make more frequent shopping trips. People in social class 2 tend to make higher number of trips too. It should be noted here that the relative value of the constant is relatively high which suggests some deficiencies of the model.

Table 9.6 Relative importance of each variable in linear trip generation models

Variables	Model-1-a (all users)		Model-2 (car users)		Model-3 (PT users)	
	M*Coeff.	Elasticity	M*Coeff.	Elasticity	M*Coeff.	Elasticity
Constant	0.590		-0.238		0.085	
CAR	-0.093	-0.071	-	-	-	-
PT	-0.288	-0.220	-	-	-	-
CAR0	0.110	0.084	-	-	0.152	0.134
AGE1	0.126	0.097	0.259	0.232	0.146	0.129
AGE2	0.077	0.059	0.308	0.276	-	-
SOCI1	0.067	0.051	-	-	0.131	0.116
SOCI2	0.163	0.124	-	-	0.216	0.191
SOCI12	-	-	0.240	0.215	-	-
LOCA1	0.324	0.248	-	-	0.197	0.174
LOCA2	0.245	0.187	-	-	0.206	0.182
LOCA12	-	-	0.615	0.551	-	-
PARKCOST	-0.015	-0.012	-0.068	-0.061	-	-
TOTAL	1.306	-	1.116	-	1.132	-

9.4 LOGISTIC REGRESSION TRIP GENERATION MODELS

In this section, we present the trip generation models for shopping trips in Edinburgh calibrated using logistic regression analysis. Binary logit models as well as MNL and NL models were calibrated. In this analysis, the frequency of weekly shopping trips was used to form the discrete options of the choice sets available to the shoppers. Table 9.7 shows the shopping trip frequency of all users and for car users only respectively. From the table it appears that of all respondents, 22.6% make very frequent trips and 57.8% make infrequent trips, while for car users only, the percentages are 16.3% and 64.6% respectively. This categorisation of the trip frequencies has been used as the basis to construct the discrete options in the logit models.

Firstly, binary logistic models were calibrated for trip generation models with two discrete options: respondents who make less than one shopping trip a week and respondents who make one or more shopping trips per week. Secondly, MNL and NL models were calibrated with three options, i.e., respondents making infrequent trips (less than once a week), respondents making frequent trips (weekly trips) and respondents making very frequent trips (2-7 trips a week). The models are presented and discussed in the following sections.

Table 9.7 Frequency of visits to the city centre for non-food shopping for all users (n = 879) and car users only (n = 237)

Frequency		All Respondents %		Car Users %	
Very Frequent	Daily	7.4	22.8	7.5	16.3
	4-6 times a week	3.2		1.7	
	2-3 times a week	12.2		7.1	
Frequent	Weekly	19.4	19.4	19.1	19.1
Infrequent	Fortnightly	16.8	57.8	16.7	64.6
	Monthly	18.8		19.2	
	Less than once a month	22.2		28.7	

9.4.1 Binary logit models for shopping trips

As discussed, binary logit models were calibrated for trip generation models. Three models were calibrated; a model for all users (Model-4), a model for car users (Model-5) and a model for public transport users (Model-6). The utility functions for Model-4, Model-5 and Model-6 are as presented in Table 9.8.

The coefficient estimates for the above models were calibrated using the ALOGIT software (Daly, 1992) as shown in Table 9.9. As shown in Table 9.9, all coefficients have the correct signs and there are evidences that car users have different attitudes and behaviour than public transport users (i.e. different coefficients of the variables used in the model). The positive sign of AGE1 (people of age 16-34) in utility 2 indicate that this age group is more likely to make more trips (Model-4, Model-5 and Model-6). The negative sign of the PARKCOST in utility 2 indicates that fewer trips are expected as parking costs increase. Moreover, from the models it is confirmed that car users and public transport users make relatively less shopping trips (positive signs of CAR and PT in utility 1 in Model-4) as discussed before.

Table 9.8 The utility functions for Model-4, Model-5 and Model-6

Model	Utility Function	Variables (see Table 9.1)	Coefficients to be estimated
Model 4	$V_1 = \theta_{car} CAR + \theta_{pt} PT$	CAR, PT	θ_{pt}, θ_{car}
	$V_2 = constant_2 + \theta_{age1} AGE1 + \theta_{car0} CAR0 + \theta_{soci12} SOCI12 + \theta_{loca1} LOCA1 + \theta_{loca2} LOCA2 + \theta_{parkcost} PARKCOST$	AGE1, CAR0, SOCI12, LOCA1, LOCA2, PARKCOST	$constant_2$, θ_{age1}, θ_{car0}, θ_{soci12}, θ_{loca1}, θ_{loca2}, $\theta_{parkcost}$
Model 5	$V_1 = 0$		
	$V_2 = constant_2 + \theta_{age1} AGE1 + \theta_{soci2} SOCI2 + \theta_{loca1} LOCA1 + \theta_{loca2} LOCA2 + \theta_{gender} GENDER + \theta_{parkcost} PARKCOST$	AGE1, SOCI2, LOCA1, LOCA2, GENDER, PARKCOST	$constant_2$, θ_{age1}, θ_{soci2}, θ_{loca1}, θ_{loca2}, θ_{gender}, $\theta_{parkcost}$
Model 6	$V_1 = 0$		
	$V_2 = constant_2 + \theta_{age1} AGE1 + \theta_{car0} CAR0 + \theta_{soci1} SOCI1 + \theta_{soci2} SOCI2 + \theta_{loca1} LOCA1 + \theta_{loca2} LOCA2$	AGE1, CAR0, SOCI1, SOCI2, LOCA1, LOCA2,	$constant_2$, θ_{age1}, θ_{car0}, θ_{soci1}, θ_{soci2}, θ_{loca1}, θ_{loca2}

People in social groups 1 and 2 are more likely to make one or more shopping trips to the city centre (positive coefficients of SOCI1 and SOCI2 in utility 2 for Model-5 and Model-6). People who live in the city centre or inter-cordon zone are more likely to make one or more shopping trips to the city centre (positive coefficients of LOCA1 and LOCA2 in utility 2 for Model-4, Model-5 and Model-6). The respondents from households with no cars would make more frequent shopping trips to the city centre (positive coefficient of CAR0 in the utility 2 in Model-4 and Model-6). This is possibly because people with cars might decide to go shopping at other locations than the city centre to avoid parking charges, while non-car owners would more frequently go to the city centre for their shopping trips. Male respondents are observed to make more frequent shopping trips than female respondents (positive coefficient of GENDER in utility 2 in Model-5).

In order to further investigate these results, the relative importance of each of the variables, in a similar way to the previously presented approach in Table 9.6, has been calculated here and presented in Table 9.10. These values have been calculated as the product of the coefficient and the mean value of the variable. From the table, it appears that constant has a relatively high value to the rest of the variables. As expected the location, the public transport mode of travel variables have positive influence on the frequency of trip generation to the city centre.

Table 9.9 Binary logit models of shopping trip generation to the city centre

Variables (option)	Coefficient (t-ratio) Model-4 (all users)	Model-5 (car users)	Model-6 (PT users)
Constant (2)	-0.925 (-2.8)	-1.734 (-4.6)	-1.538 (-5.4)
CAR (1)	0.319 (1.1)	-	-
PT (1)	0.593 (2.5)	-	-
AGE1 (2)	0.579 (3.6)	0.395 (1.2)	0.799 (3.9)
CAR0 (2)	0.280 (1.6)	-	0.345 (1.6)
SOCI1 (2)	-	-	0.478 (1.7)
SOCI2 (2)	-	0.344 (1.1)	0.316 (1.3)
SOCI12(2)	0.240 (1.3)	-	-
LOCA1 (2)	1.343 (6.1)	1.854 (4.6)	1.335 (4.5)
LOCA2 (2)	0.710 (4.0)	1.548 (4.4)	0.350 (1.6)
GENDER(2)	-	0.369 (1.2)	-
PARKCOST (2)	-0.115 (-2.1)	-0.125 (-2.1)	-
Initial log-likelihood	-609.276	-164.276	-350.039
Final log-likelihood	-546.120	-134.978	-316.274
$\rho^2(0)$	0.104	0.178	0.097
$\rho^2(c)$	0.088	0.124	0.062
n	879	237	505

The options used in the models:

1 = less than once a week

2 = One and more trips a week

Table 9.10 The relative importance of the variables in the binary logit models

Variables (option)	Model-4 (all users)	Model-5 (car users)	Model-6 (PT users)
Constant (2)	-0.925	-1.734	-1.538
CAR (1)	0.086	-	-
PT (1)	0.341	-	-
AGE1 (2)	0.170	0.102	0.261
CAR0 (2)	0.100	-	0.157
SOCI1 (2)	-	-	0.142
SOCI2 (2)	-	0.164	0.123
SOCI12(2)	0.181	-	-
LOCA1 (2)	0.346	0.367	0.200
LOCA2 (2)	0.288	0.509	0.172
GENDER(2)	-	0.168	-
PARKCOST (2)	-0.079	-0.305	-

9.4.2 MNL and NL models for shopping trips

Shopping trip generation models using three options: infrequent shopping trips (i.e. less than once a week); frequent (weekly) and very frequent (2-7 trips a week) were also calibrated. It might be argued however, that the frequent and very frequent shoppers are more similar and that they are different than those who are infrequent travellers. For this reason, two models forms were tested; firstly the standard MNL model, where the three options were considered as independent and then the Nested Logit model (NL) to investigate any correlation between the frequent and very frequent users. The structures of the two models are shown in Figure 9.1. It is noted here that the best NL model was obtained by nesting the two groups of respondents (frequent and infrequent) together at the lower level while the 'very frequent' group is considered at the higher level. This is interesting since the trips' frequencies are more similar for respondents in the first two groups of travellers than those who make very frequent trips (see also Table 9.7).

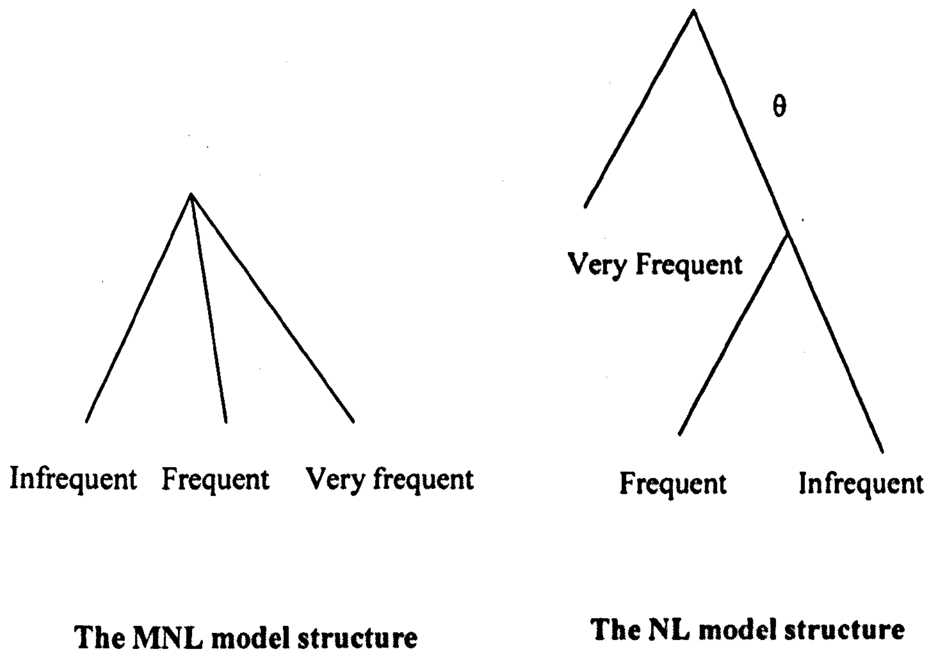


Figure 9.1 The structures for the MNL and NL trip generation models

The MNL coefficient estimates of the variables were calibrated using the ALOGIT software (Daly, 1992). Furthermore the coefficients of the NL and the theta parameter for the model were also calibrated using the ALOGIT.

The utility functions, the variables used in the models and their coefficients for Model-4, Model-5 and Model-6 are as presented in Table 9.11. It is noted here that the allocation of these variables to each utility function has been mainly done based on the statistical significance of the model outcomes. Therefore there were a number of trials and errors before deciding on the final models structure presented here.

Table 9.12 shows the estimates of the coefficients for the MNL model (Model-7) and the NL model (Model-8). As shown in Table 9.12, all the coefficients have the correct signs and the ρ^2 values have improved from those calibrated from the binary logit model (Table 9.9). The negative sign of the AGE1 (people of age 16-34) indicates that this age group is less likely to make more trips (Model-7 and Model-8). The negative sign of the PARKCOST indicates that fewer trips are expected to be made to the city centre as parking costs increase. Moreover, from the models it appears that car users make relatively less shopping trips than the users of the other modes (negative sign of CAR in both models).

Table 9.11 The utility functions for Model-7 and Model-8

Utility Function	Variables (see Table 9.1)	Coefficients to be estimated
$V_1 = \theta_1^{pt} PT + \theta_1^{age1} AGE1$	PT, AGE1,	θ_{pt}, θ_{age1}
$V_2 = constant_2 + \theta_2^{car0} CAR0 + \theta_2^{local} LOCA1 + \theta_2^{loca2} LOCA2 + \theta_2^{gender} GENDER + \theta_2^{parkcost} PARKCOST$	CAR0, LOCA1, LOCA2, GENDER, PARKCOST	$constant_2$, θ_2^{car0}, θ_2^{local}, θ_2^{loca2}, θ_2^{gender}, $\theta_2^{parkcost}$
$V_3 = constant_3 + \theta_3^{car} CAR + \theta_3^{car0} CAR0 + \theta_3^{soci1} \theta_{soci1} SOCI1 + \theta_3^{soci2} SOCI2 + \theta_3^{local} LOCA1 + \theta_3^{loca2} LOCA2 + \theta_3^{parkcost} PARKCOST$	CAR, SOCI1, SOCI2	$constant_3$, θ_3^{car}, θ_3^{car0}, θ_3^{socl1}, θ_3^{soci2}, θ_3^{local}, θ_3^{loca2}, $\theta_3^{parkcost}$

People in social groups 1 and 2, and those who live in the city centre or inter-cordon zone, are more likely to make one or more shopping trips to the city centre (positive coefficients of SOCI1, SOCI2, LOCA1 and LOCA2 in the two models). Similar to the results which were obtained from the binary logit model, the respondents whose households have no cars seem to make more frequent shopping trips to the city centre (positive coefficient of CAR0 in the models), since they make all or most of their shopping within the city centre. On the other hand the car owners would probably drive to out with the city centre to other locations for their shopping in order to avoid parking charges. Male respondents are observed to make more frequent shopping trips than female respondents (positive coefficient of GENDER).

From the table, the coefficients estimates and the final likelihood values are very similar in each of the MNL and NL models. The one difference here is the CAR0 variable which is incorporated as the common factor in the NL model in both options of the nest (options 2 and 3). Moreover the Theta parameter is close to 1 and not statistically significant at the 95% level (i.e. not statistically different from 0). This would suggest that the MNL structure is sufficient and there is no added value in this case for suggesting the nested structure. The results from the Likelihood ratio tests also support these findings (i.e. Final Likelihood values for each of the models are -792.5656 and -792.5655 respectively with 1 degree of freedom).

Table 9.12 MNL and NL models of shopping trip generation the city centre

Variables (option)	Coefficient (t-ratio)	
	Model-7 (MNL, all users)	Model-8 (NL, all users)
Constant (2)	-1.437 (-6.1)	-1.438 (-6.1)
Constant (3)	-2.022 (-5.9)	-2.129 (-5.4)
PT (1)	0.540 (2.9)	0.539 (2.8)
AGE1 (1)	-0.560 (-3.5)	-0.597 (-2.9)
CAR0(2)	0.311 (1.6)	**0.269 (1.6)**
LOCA1 (2)	0.826 (3.3)	0.830 (3.3)
LOCA2 (2)	0.531 (2.5)	0.530 (2.5)
GENDER(2)	0.173 (1.0)	0.179 (1.0)
PARKCOST (2)	-0.162 (-2.4)	-0.166 (-2.5)
CAR (3)	-0.499 (-1.8)	-0.439 (-1.5)
CAR0(3)	0.243 (1.2)	**0.269 (1.6)**
SOCI1 (3)	0.366 (1.5)	0.384 (1.6)
SOCI2 (3)	0.502 (2.2)	0.519 (2.3)
LOCA1 (3)	1.869 (7.0)	1.852 (7.0)
LOCA2 (3)	0.988 (4.0)	0.971 (3.8)
PARKCOST (3)	-0.081 (-1.3)	-0.076 (-1.1)
THETA	-	0.8583 (1.8)
Initial log-likelihood	-965.6802	-965.6802
Final log-likelihood	-792.5656	-792.5655
$\rho^2(0)$	0.1793	0.1793
$\rho^2(c)$	0.0737	0.0737
n	879	879

The options used in the models:

1 = infrequent (less than once a week)

2 = frequent (weekly)

3 = very frequent (2-7 trips a week)

9.5 SUMMARY

Linear regression analysis and logistic analysis (binary, MNL and NL models) have been used to calibrate shopping trip generation models including parking costs to represent a transport policy measure. The coefficient estimates of the variables used, the statistical significance and the overall goodness of fit of MNL and NL models are very similar. The nested logit model structure did not seem to provide any improvements of the goodness of fit over those obtained from the MNL model. Hence it has been concluded that there is no obvious evidence of correlation between frequent and very frequent travellers in this data set, as implied by the nested logit structure.

The results from the models presented in this chapter suggest that policy measures which would be implemented in the city centre should have an impact on the frequency of the shopping trips. For example, in this case the increase in parking costs result in people making less frequent trips to the city centre. While this type of measures seems logical and obvious to be included in trip generation models, there is still a lack of including such measures explicitly in current trip generation models, hence this analysis. In this data set, there are no other policy measures/variables for further investigations. For example parking duration, parking supply, bus lanes and other measures would present interesting transport policy measures which could be investigated, compared and included in trip generation models. Therefore, further investigations and inclusion of such measures would be recommended. Also there is evidence that socio economic variables such as age and social class also have impacts on the frequency of shopping trips in the city centre.

CHAPTER 10 MODELLING TRP GENERATION WITH PARKING COSTS AND CONGESTION CHARGING

In this chapter, the potential impacts of congestion charging as well as parking costs on trip generation of shopping trips in Edinburgh are investigated using logistic regression. Although the introduction of congestion charging seems to have mostly negative impacts on shopping trips, because of the inconvenience and the increase in the overall cost of shopping, the results show that there might be some positive impacts of congestion charging. This is mainly because the introduction of congestion charging would result in less congestion as well as improvements of the public transport system and hence, an increase in some shopping trips.

In this chapter, two sets of models were calibrated by segmenting the shoppers according to the mode they use. Firstly, models were calibrated for all users and secondly models for car users. Stated Preference (SP), Revealed Preference (RP) and mixed RP/SP models were investigated and assessed.

10.1 CONGESTION CHARGE SCHEME IN EDINBURGH

Congestion charging as well as parking management measures are increasingly considered as traffic demand management (TDM) tools in the UK as well as in most world cities (Litman, 2004; European Commission, 2004). In London, a congestion charging scheme was implemented in February 2003 to control traffic congestion into the city (Banister, 2003). Under this scheme vehicles inside a 22-square kilometre zone enclosing the core shopping, government, entertainment and business districts between 7:00 and 18:30 on weekdays have been charged a £5 daily fee (£8 since July 2005), unless they are eligible for a resident discount or are exempted from the charges (Schmöcker, 2006).

Recently, the City of Edinburgh had plans to introduce congestion charging in the form of a double cordon as a policy to reduce traffic in the central areas.

Although the scheme was abandoned following a public referendum (CEC, 2005), a number of research studies and investigations have been carried out to assess the appropriateness of the scheme and the related policies (MVA Consultancy, 2004; Farrell, 2005).

The continual increase of car ownership and usage has lead to increased traffic congestion and associated problems in Edinburgh. Although traffic levels have stabilised in the city centre due to a variety of reasons, such as the transport policies pursued in recent years (e.g. Greenways, parking controls and the closing of traffic on Princes Street) and other reasons, such as the location of business and activities away from the city centre, traffic levels have worsened in areas outside of the centre (Farrell, 2005). Traffic forecasts based on current trends and current levels of public transport investment show that traffic levels will increase by over 20% in Edinburgh between 2001 and 2021 (City of Edinburgh Council, 2002). It was recognised that there was a need for some form of traffic restraint if this forecasted increase in traffic was to be avoided.

The purpose of the congestion charge in Edinburgh was primarily to reduce congestion in the city and, secondly, to fund transport infrastructure improvements. It was planned to introduce the congestion charge in 2006 if the support of the local population was achieved in the public referendum and Scottish Ministers had approved the scheme.

Based on the plans, the cost to motorists coming into the city during the period at which the congestion charge would be operational was £2. This would be a one off daily charge irrespective of how many times a motorist crossed a cordon during a day. The congestion charge would apply during weekdays (Monday to Friday) only. Motorists would pay for crossing either of the two proposed cordons in the inbound direction only. There would be a city centre cordon operating between 7am and 6.30pm and an outer cordon, inside the Bypass, operating between 7am and 10am only (see ECCM, 2004 for more details of the scheme).

Another study, of the impacts of congestion charging in Edinburgh on departure time choice (Farrell, 2005), investigated and modelled departure time patterns as a result of the proposed scheme. The proposed scheme for Edinburgh was different from the London scheme in some aspects (Farrell, 2005). The London Scheme is an area-licensing scheme, which means that a charge is applied if a vehicle is within the charging zone even if it is moved only a short distance. For the Edinburgh scheme, a charge would only be applied if a vehicle crossed into the charging zone. Another difference between the two schemes is the level of charging; the charge was £5 (now £8) in London but would have been only £2 in Edinburgh. Saunders (2004) recognised that £2 was a modest charge that was not high relative to the overall cost of travel. Nevertheless, it was also claimed that the charge would be adequate in terms of affecting congestion and making available revenues for public transport. Interestingly, there have been a large number of studies and data collected in Edinburgh to investigate various impacts of the proposed congestion charging scheme (for example Farrell, 2005, Saunders, 2004 and ECCM, 2004).

10.2 POTENTIAL IMPACTS OF CONGESTION CHARGE ON SHOPPING FREQUENCY

In this section the data collected during the ECCM study has been further investigated to assess the impacts of congestion charging in Edinburgh on the frequency of shopping trips. For further discussion on this survey see section 5.3.

Figure 10.1 and Figure 10.2 show the stated current frequency of shopping trips for all users and car users "before" and "after" the introduction of congestion charge. The frequency of visits to the city centre for non-food shopping for the shoppers have been reported (see Section 5.3) and categorised in this section as three categories: not frequent, frequent or very frequent. From Figure 10.1 it is clear that about 58% of respondents were observed to make not very frequent shopping trips in the before case. This percentage would have increased to over 62% if congestion charging would have been introduced in Edinburgh. The very frequent, as well as the frequent shoppers (i.e. for shoppers who make 2-7 shopping trips per week and those who make weekly trips) would have dropped in the after case to 20.4% and 17.3% from 22.6% and 19.7% in the before case respectively.

For all users, the changes in frequency are not very significant as discussed above (i.e. the change in frequency of shopping trips for all users range from 2.2% to 4.4%). However, as shown in Figure 10.2, the changes in frequency of shopping trips for car users are more significant with "very frequent shoppers" reducing the frequency of shopping trips by about one third (from 16.3% to 10.5%) and with "frequent shoppers" reducing their shopping trips frequency from 19.2% to 15.1%.

Therefore it appears that on one hand, the car users are less frequent shoppers into the city centre than other groups. On the other hand, they would have been more affected by the introduction of congestion charging into the city centre and hence more perceptibly responding to it.

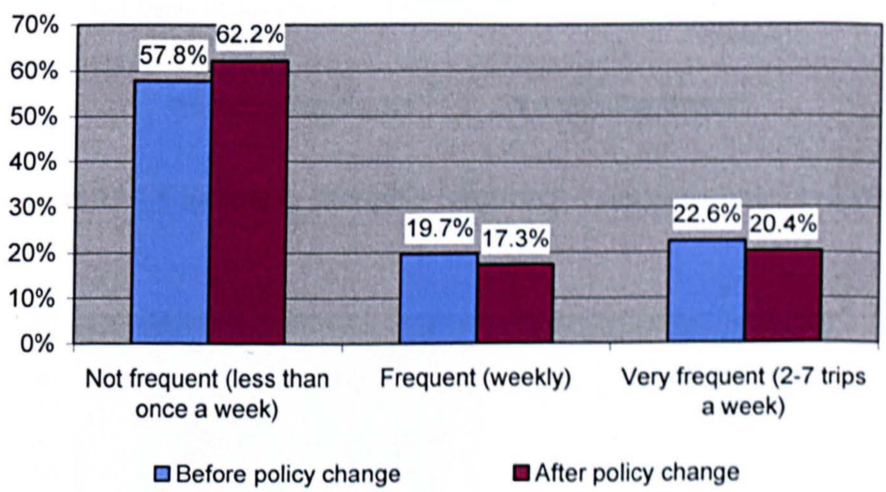


Figure 10.1 Frequencies of shopping trips before and after– all users (n=890)

Moreover, in the survey, respondents were asked to report on their perceived attitudes towards the introduction of congestion charging in the city centre in terms of shopping trips (i.e. reduction or increase in trip frequency). Figure 10.3 shows the perceived impacts of congestion charging on shopping trips. As shown in the figure, most shoppers appear not to be affected by the introduction of congestion charging (about 83%). However, of those who use the car (27.0% of the shoppers), about 37% said they would spend less or go elsewhere, whereas for over 90% of public transport users, the charge would make no difference. Therefore, the public transport users would be far less affected by the scheme as expected. Therefore, the segmentation of the data based on the mode used would be reasonable in this case.

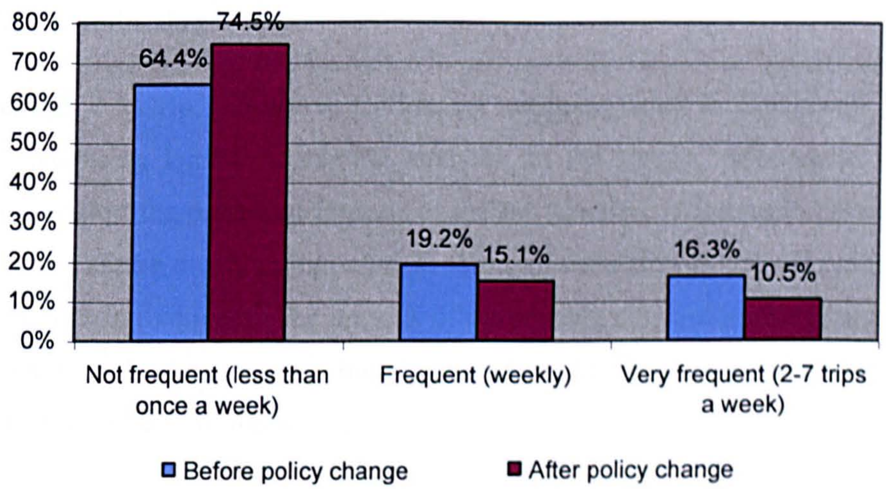


Figure 10.2 Changes in shopping frequency – car users only (n=239)

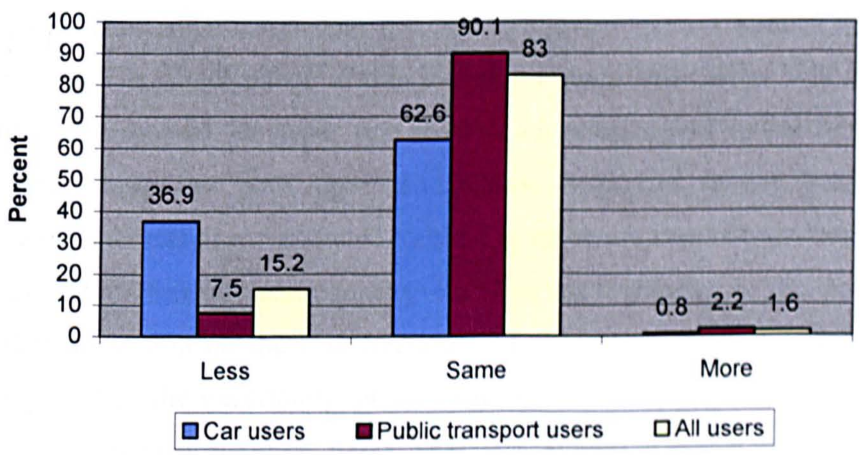


Figure 10.3 Changes in shopping frequency – all users (n=895), car users (n=237) and public transport users

Although this is not a case of multiple responses from each respondent on some future policies (i.e. as in a typical SP scenario), these reported responses have been used in this study as "stated preference" information on the likely impacts of the introduction of congestion charging on the frequency of trips into the city centre as discussed below.

10.3 MODELLING SHOPPING TRIP GENERATION AFTER INTRODUCING THE CONGESTION CHARGE USING MIXED RP/SP MODELS

10.3.1 Introduction

In section 7.3, trip generation models for shopping trips in Edinburgh were calibrated using logistic regression analysis. In the survey, respondents were asked to report the perceived impacts of the introduction of congestion charging in the city centre on shopping trips, in terms of reduction in trip frequency or increase in trip frequency. The impacts of introducing congestion charging in the city centre on the frequency of shopping trips have therefore been modelled in an SP and a mixed RP/SP models.

The reported responses have been used in this study as "stated preference" data to indicate the potential frequency of shopping trips to the city centre after the introduction of congestion charging. In this case however, there is only one response from each respondent (i.e. not multiple responses as in a typical SP exercise). The disadvantage of this is that not much information will be gained (only one response). However, one possible advantage could be that there will be no errors associated with repeated responses. Moreover, in this specific case, there is no effects of incorporated "state" or reference dependence between data types and preference heterogeneity on observed attributes in the model (see further discussion on these in Section 4.5). These are the two sources which cause most of the uncertainty/ errors in the joint RP/SP models as discussed in literature (see Hensher *et al*, 2008). Therefore, at least in theory the calibration of the mixed (RP and SP) models could well be implemented using NL trick model (see discussions in Section 4.5.4 and Hensher and Bradley, 1993).

10.3.2 Joint estimation of RP/SP trip generation models

In this section, modelling trip generation of shopping trips is carried out which includes the potential impact of introducing a road pricing scheme. The data used in this section was obtained from the ECCM Household Survey (see Section 5.3). The data includes a revealed preference section which contains information about shopping trips, socio economic and location characteristics of the respondents. It also contains information on the perceived or reported shopping trips patterns before and after congestion charges are introduced in the city.

As discussed above, there is only one SP response from each individual, which is not the typical SP design. However, in the absence of any other more appropriate SP data, it was decided to use this single statement as to represent potential behaviour regarding shopping trips with congestion charging and to calibrate mixed RP/SP models. These models have been calibrated to investigate the potential impacts of congestion charging on the frequency of shopping trips to the city centre of Edinburgh, for all users and for car users respectively as shown in Table 10.3.

Table 10.1 presents the description of the variables. In these models it was assumed that the congestion charging value was £2.00, applicable to car users who were not residents of the central area. This congestion charging value was added to the parking charging costs that was reported by the users.

Table 10.1 Variable description for the shopping trip generation models

Variables	Description
CAR	Dummy variable: takes the value of 1 if respondent normally travels into the City Centre for non-food shopping by car or van, 0 otherwise.
PT	Dummy variable: takes the value of 1 if respondent normally travels into the City Centre for non-food shopping by bus and train, 0 otherwise.
CAR0	Dummy variable: takes the value of 1 if respondent's household owns no car, 0 otherwise.
AGE1	Dummy variable: takes the value of 1 if respondent's age is 16-34, 0 otherwise.
AGE2	Dummy variable: takes the value of 1 if respondent's age is 35-54, 0 otherwise.
SOCI1	Dummy variable: takes the value of 1 if respondent's social grade is upper middle class (A) or middle class (B), 0 otherwise.
SOCI2	Dummy variable: takes the value of 1 if respondent's social grade is lower middle class (C1) or skilled worker (C2), 0 otherwise.
LOCA1	Dummy variable: takes the value of 1 if respondent's location is City Centre, 0 otherwise.
LOCA2	Dummy variable: takes the value of 1 if respondent's location is inter-cordon area, 0 otherwise.
LOCA12	Dummy variable: takes the value of 1 if respondent's location is City Centre or inter-cordon area, 0 otherwise.
GENDER	Dummy variable: takes the value of 1 if respondent is a male, 0 female.
PARKCOST	Continuous variable: describes the parking cost of a non-food shopping visit travelling to the City Centre.
PARK_CC	Continuous variable: describes the parking cost of a non-food shopping visit travelling to the City Centre, plus the £2 congestion charge for those car users who lives outside the central area.
CCOSTLY	Dummy variable: takes the value of 1 if respondent says the congestion charge is very costly, 0 otherwise.
INCONVEN	Dummy variable: takes the value of 1 if respondent says the congestion charge is inconvenient, 0 otherwise.
LESSCONG	Dummy variable: takes the value of 1 if respondent says it would be less congested if congestion charge is applied, 0 otherwise.
EASIGF	Dummy variable: takes the value of 1 if respondent says it would be easier to go and from the city centre if congestion charge is applied, 0 otherwise.
PTIMPROV	Dummy variable: takes the value of 1 if respondent says public transport would improve if congestion charge is applied, 0 otherwise.

The expected impacts of these variables (Table 10.1) are as discussed earlier in Section 9.2. The variable which combines the parking cost and the £2 congestion charge for car users who live outside the central area, is used to reflect the introduction of congestion charge and is expected to have a negative impact on shopping trips by car.

Figure 10.4 shows the artificial tree structure used in this mixed RP/SP model. For more details about this estimation method see discussions in Section 4.5.

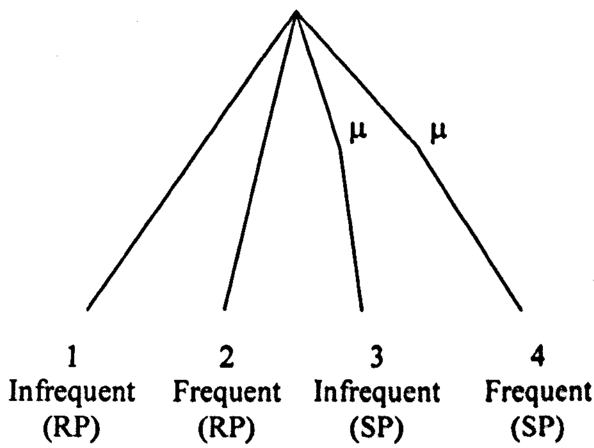


Figure 10.4 Artificial tree structure for mixed RP and SP estimation

In a mixed RP/SP model, we can have the following utility functions for a certain alternative A_i (Ortúzar and Willumsen, 2001):

$$U_i^{RP} = \theta X_i^{RP} + \alpha Y_i^{RP} + \varepsilon_i$$
$$\mu U_i^{SP} = \mu\left(\theta X_i^{SP} + \phi Z_i^{SP} + \eta_i\right)$$

where α, ϕ and θ are parameters to be estimated; X^{RP} and X^{SP} are common attributes (of both alternatives and individuals) at the RP and SP levels respectively; Y^{RP} and Z^{SP} are attributes which only belong to the RP or SP sets respectively; μ is the scale coefficient; and ε, η are errors.

Prior to estimate the mixed RP/SP model, the RP only and SP only models need to be estimated each of which includes all the independent variables to decide which attributes to be included as specific or common (the X set, as opposed to the Y and Z sets), see further discussion of this in Section 4.5. In this case, an RP only (Model-1) and an SP only (Model-2) models were calibrated and assessed. Table 10.3 shows the estimation results of the two individual models (RP and SP). These two models were then tested for the allocation of the independent variables in the combined model, using a procedure to investigate parameter equality in the two data sets suggested by Louviere et al (2000) and discussed in Section 4.5.

In this procedure a graph is plotted for the parameters' vectors obtained from the RP against those estimated from SP models (Figure 10.5). In this case, the graph of the RP parameter vector against the other (i.e. SP) produces a cloud of points passing through the origin of the graph with positive slope equal to the ratio of error variance of set 2 to set 1). From the figure, we can assume that the two sources of data produce the same utilities but potentially different scale. In this case, a combined model will have the variables (AGE1 SOCI2, GENDER) included as common variables, while the variables (CAR, PT, CAR0, SOCI1, LOCA1, LOCA2 and PARKCOST) are best included as RP specific and the variables (CAR, PT, CAR0, SOCI1, LOCA1, LOCA2 and PARK_CC) are SP specific.

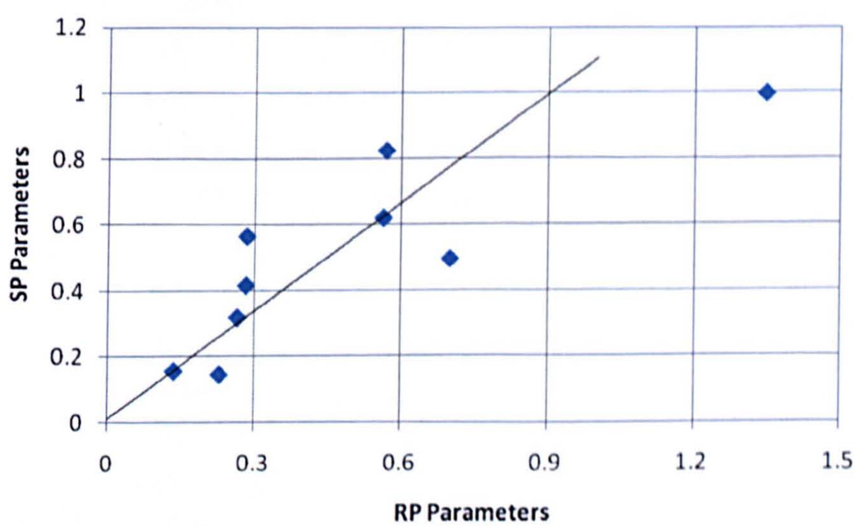


Figure 10.5 Parameter plot for data combination (all users)

Table 10.2 The utility functions for RP, SP and mixed RP/SP models for all users

Model	Utility Function	Variables (see Table 10.1 for definition of variables)	Coefficients to be Estimated
Model 1	$V_1 = \theta_1^{car} CAR + \theta_1^{pt} PT$ $V_2 = constant_2 + \theta_2^{age1} AGE1 + \theta_2^{car0} CAR0 + \theta_2^{soci1} SOCI1 + \theta_2^{soci2}{}_{soci2} SOCI2 + \theta_2^{loca1} LOCA1 + \theta_2^{loca2} LOCA2 + \theta_2^{gender} GENDER + \theta_2^{parkcost} \theta_{parkcost} PARKCOST$	CAR, PT AGE1, CAR0, SOCI1, SOCI2, LOCA1, LOCA2, GENDER, PARKCOST	$constant_2$, θ_1^{car}, θ_1^{pt}, θ_2^{age1}, θ_2^{car0}, θ_2^{soci1}, θ_2^{soci2}, θ_2^{loca1}, θ_2^{loca2}, θ_2^{gender}, $\theta_2^{parkcost}$
Model 2	$V_3 = \theta_3^{car} CAR + \theta_3^{pt} PT$ $V_4 = constant_4 + \theta_4^{age1} AGE1 + \theta_4^{car0} CAR0 + \theta_4^{soci1} SOCI1 + \theta_4^{soci2}{}_{soci2} SOCI2 + \theta_4^{loca1} LOCA1 + \theta_4^{loca2} LOCA2 + \theta_4^{gender} GENDER + \theta_4^{parkcost\_cc} PARKCOST\_CC$	CAR, PT AGE1, SOCI1, SOCI2, LOCA1, LOCA2, GENDER, PARKCOST_ CC	$constant_4$, θ_3^{car}, θ_3^{pt}, θ_4^{age1}, θ_4^{car0}, θ_4^{soci1}, θ_4^{soci2}, θ_4^{loca1}, θ_4^{loca2}, θ_4^{gender}, $\theta_4^{parkcost\_cc}$
Model 3	$V_1 = \theta_1^{car} CAR + \theta_1^{pt} PT$ $V_2 = constant_2 + \theta_2^{age1} AGE1 + \theta_2^{car0} CAR0 + \theta_2^{soci1} SOCI1 + \theta_2^{soci2}{}_{soci2} SOCI2 + \theta_2^{loca1} LOCA1 + \theta_2^{loca2} LOCA2 + \theta_2^{gender} GENDER + \theta_2^{parkcost} \theta_{parkcost} PARKCOST$ $V_3 = \theta_3^{car} CAR + \theta_3^{pt} PT$ $V_4 = constant_4 + \theta_2^{age1} AGE1 + \theta_4^{car0} CAR0 + \theta_4^{soci1} SOCI1 + \theta_2^{soci2}{}_{soci2} SOCI2 + \theta_4^{loca1} LOCA1 + \theta_4^{loca2} LOCA2 + \theta_2^{gender} GENDER + \theta_4^{parkcost\_cc} PARKCOST\_CC$	CAR, PT AGE1, CAR0, SOCI1, SOCI2, LOCA1, LOCA2, GENDER, PARKCOST PARKCOST_ CC	$constant_2$, $constant_4$, θ_1^{car}, θ_1^{pt}, θ_2^{age1}, θ_2^{car0}, θ_2^{soci1}, θ_2^{soci2}, θ_2^{loca1}, θ_2^{loca2}, θ_2^{gender}, $\theta_2^{parkcost}$ $\theta_4^{parkcost\_cc}$, θ_3^{car}, θ_3^{pt}, θ_4^{car0}, θ_4^{soci2}, θ_4^{loca1}, θ_4^{loca2} $\theta_4^{parkcost}$

The utility functions for Model-1 (RP model), Model-2 (SP model) and Model_3 (mixed RP/SP model) are given in Table 10.2. The coefficients of the RP, SP and mixed RP/SP models for all users are presented in Table 10.3.

As shown in Table 10.3, all the variables have the logical signs and most of them are statistically significant at 95% level. It appears that car users and public transport users make relatively less frequent shopping trips (positive signs of CAR and PT) as discussed before in Section 9.3. The coefficient of "congestion charging plus parking costs" is statistically significant at 95% level with a logical sign (negative sign). This implies that as the value of congestion charging plus parking costs increases, lower frequencies of shopping trips at the central area are expected; a result which is mainly applicable to the car users. The t-values in the three models are comparable, although it is difficult to draw specific conclusions on these values since the number of observations is different in the joint model. For the mixed RP/SP model, the results show a statistically significant scaling parameter of 1.099 suggesting that the SP data have less random noise than the RP data. This result could also be reinforced by the higher $\rho^2(0)$ of the SP model.

These results are encouraging in terms of the utilisation of logistic regression techniques and mixed logit in trip generation modelling Further investigations and applications however are still needed in this area. It should be mentioned here that the quality of data is a crucial factor for obtaining good quality models. That is in particular important when combining more than one type or source of data, for example in the joint estimation of RP/|SP models.

Similarly, trip generation models (i.e. RP, SP and Joint RP/SP models) for the car users were calibrated. A graph is plotted for the parameters estimated from the RP against the parameters estimated from SP models for the car users (Figure 10.6). From the figure, it appears that the variables (AGE1, SOCI2, LOCA1 and PARKCOST) are best included as RP specific, the variables (AGE1, SOCI2, LOCA1 and PARK_CC) are SP specific while the variables (LOCA2 and GENDER) are included as common variables.

Table 10.3 Mixed RP/SP models for shopping trip generation for all users

Variables (option)	Model-1 (RP)	Model-2 (SP)	Model-3 (Mixed RP/SP)
Constant (2)	-1.009 (-3.0)	-	-0.807 (-3.1)
Constant (4)	-	-0.813 (-2.4)	-1.006 (-3.2)
CAR (1)	0.283 (1.0)	-	0.214 (0.9)
CAR (3)	-	0.566 (1.7)	0.327 (1.5)
PT (1)	0.567 (2.4)	-	0.646 (3.7)
PT (3)	-	0.825 (3.5)	0.630 (2.8)
AGE1 (2,4)	0.562 (3.4)	0.621 (3.7)	0.568 (4.2)
CAR0 (2)	0.282 (1.6)	-	0.230 (1.6)
CAR0 (4)	-	0.416 (2.4)	0.426 (2.5)
SOCI1 (2)	0.228 (1.0)	-	0.195 (1.1)
SOCI1 (4)	-	0.145 (0.7)	0.161 (1.0)
SOCI2 (2,4)	0.264 (1.3)	0.318 (1.6)	0.278 (2.0)
LOCA1 (2)	1.347 (6.2)	-	1.225 (6.6)
LOCA1 (4)	-	0.998 (4.5)	1.113 (3.4)
LOCA2 (2)	0.699 (3.8)	-	0.621 (3.7)
LOCA2 (4)	-	0.496 (2.6)	0.537 (2.6)
GENDER(2,4)	0.134 (0.9)	0.156 (1.0)	0.138 (1.3)
PARKCOST (2)	-0.120 (-2.2)	-	-0.225 (-4.3)
PARK_CC (4)	-	-0.159 (-2.5)	-0.103(-2.2)
μ	-	-	1.099 (3.9)
Initial log-likelihood	-608.583	-608.583	-2434.333
Likelihood with constants only	-	-	2396.913
Final log-likelihood	-545.310	-521.651	-2280.213
$\rho^2(0)$	0.104	0.143	0.063
$\rho^2(c)$	0.088	0.103	0.049
n	878	878	1,756

The options used in modelling:

1 = infrequent (RP) - less than once a week

2 = frequent/ very frequent (RP) - one and more trips a week

3 = infrequent (SP) - less than once a week

4 = frequent/ very frequent (SP) - one and more trips a week

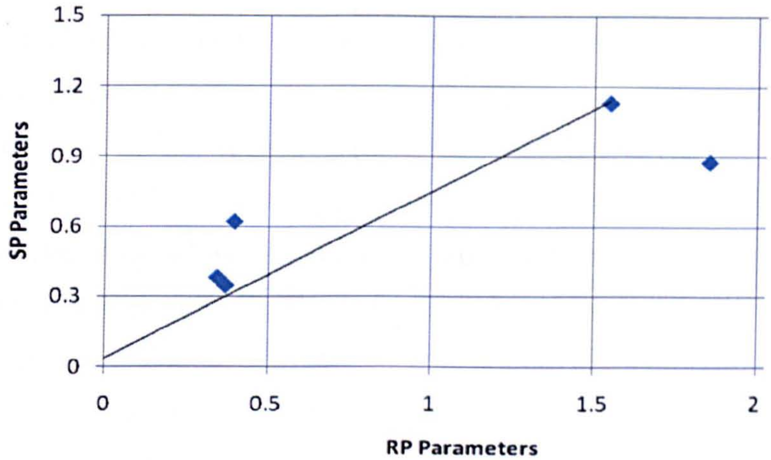


Figure 10.6 Parameter plot for data combination (car users)

The coefficients of the RP, SP and mixed RP/SP models for car users are presented in Table 10.4.

Table 10.4 Mixed RP/SP models for shopping trip generation for car users

Variables (Option)	Model-4 (RP)	Model-5 (SP)	Model-6 (Mixed RP/SP)
Constant (2)	-1.734 (-4.6)	-	-1.281 (-3.6)
Constant (4)	-	-1.519 (-3.4)	-2.499 (-3.8)
AGE1 (2)	0.395 (1.2)	-	0.491 (1.6)
AGE1 (4)	-	0.619 (1.8)	0.636 (1.5)
SOCI2 (2)	0.344 (1.1)	-	0.422 (1.5)
SOCI2 (4)	-	0.380 (1.2)	0.365 (1.0)
LOCA1 (2)	1.854 (4.6)	-	1.783 (4.9)
LOCA1 (4)	-	0.877 (2.0)	1.612 (3.0)
LOCA2 (2, 4)	1.548 (4.4)	1.129 (3.0)	1.484 (4.6)
GENDER(2,4)	0.369 (1.2)	0.345 (1.1)	0.411 (1.7)
PARKCOST (2)	-0.125 (-2.1)	-	-0.327 (-5.2)
PARK_CC (4)	-	-0.191 (-2.7)	-0.046 (-0.7)
μ	-	-	0.822 (3.2)
Initial likelihood	-164.276	-164.276	-657.104
Final likelihood	-134.978	-121.828	-576.289
$\rho^2(0)$	0.178	0.258	0.123
$\rho^2(c)$	0.124	0.092	0.066
n	237	237	474

The options used in modelling:

1 = infrequent (RP) - less than once a week

2 = frequent/ very frequent (RP) - one and more trips a week

3 = infrequent (SP) - less than once a week

4 = frequent/ very frequent (SP) - one and more trips a week

All the variables which have been included in the three models have the logical signs and most of them are statistically significant at 95% level, apart from the coefficient of the variable representing parking plus congestion costs in the combined model. However, the results show that the coefficient of congestion charging plus parking costs is statistically significant at 95% level with a logical sign (negative sign) for car users. As before, this implies that as the value of congestion charging increases, lower frequency of shopping trips at the central areas are expected for car users. For the mixed RP/SP model, the results show a scaling parameter of 0.822 suggesting that the SP data have more random noise than the RP data.

10.4 FACTORS AFFECTING THE CHANGE OF SHOPPING FREQUENCY TO THE CITY CENTRE

In this section a MNL model is used to investigate how people's social economic status will impact on the change of shopping frequency if a congestion charge was applied.

In the model, the dependent variable is the change of shopping frequency defined as below:

1 = to reduce shopping frequency (may shop less or go somewhere else);

2 = not to change; and

3 = to increase their shopping frequency (may shop over other time or change mode).

See Table 10.1 for the variables used in this section. In the Household Survey, a question was asked why the respondents would increase or decrease their shopping trips if the congestion charge was introduced. The results from this question are included in this analysis. Those who say the congestion charge is very costly and that the congestion charge is inconvenient would be expected to decrease their trips. On the other hand, for those who say it would be less congested, it would be easier to go to and from the city centre, and as public transport would improve if a congestion charge is applied this would be expected to increase their trips. The utility functions in the MNL model (MNL-7) are shown in Table 10.5.

Table 10.5 The utility functions in MNL-7

Utility Function	Variables (see Table 10.1 for definition of variables)	Coefficients to be estimated
$V_1 = constant_1 + \theta_{car}CAR + \theta_{car0}CAR0 + \theta_{local}LOCA12 + \theta_{ccostly}CCOSTLY + \theta_{inconven}INCONVEN$ $V_2 = 0$ $V_3 = constant_3 + \theta_{lesscong}LESSCONG + \theta_{easigf}EASIGF + \theta_{ptimprov}PTIMPROVE$	AGE1, CAR0, LOCA12, GENDER, PARKCOST, CCOSTLY, INCONVEN LESSCONG, EASIGF, PTIMPROVE	$constant_1$, $constant_3$, θ_{car}, θ_{car0}, θ_{local2}, $\theta_{ccostly}$, $\theta_{inconven}$ $\theta_{lesscong}$, θ_{easigf}, $\theta_{ptimprov}$

Table 10.6 shows the results of the MNL model. All the coefficients in the model have the correct signs with high values of ρ^2. As shown in the table, people who own no cars are less likely to reduce their shopping frequency (negative coefficient for CAR0 in utility one in Model-7) and car users are more likely to reduce their shopping frequency which is logical (positive coefficient for CAR in utility one in Model-7).

From the model results, it appears that shoppers who live in the city centre and inter-cordon zones are more likely to increase shopping trips (positive coefficients for LOCA12 in utility one in Model 7). This indicates that congestion charge would impact more on people living in the city centre and inter-cordon zone. Age and social group have not been found to be statically significant in the model.

Table 10.6 MNL model – shoppers who would change their shopping frequency

Variables	Model-7 (MNL - all users)		
	Coefficients (t-ratios)	Mean of the Variables	Mean*Coefficient
Constant (1)	-4.261 (-10.2)	-	-
Constant (3)	-5.608 (-9.5)	-	-
CAR (1)	0.533 (1.5)	0.270	0.144
CAR0 (1)	-1.088 (-2.4)	0.361	-0.393
LOCA12 (1)	1.419 (3.8)	0.665	0.944
CCOSTLY (1)	3.661 (10.4)	0.108	0.396
INCONVEN (1)	2.598 (6.3)	0.064	0.166
LESSCONG (3)	6.123 (7.0)	0.011	0.070
EASIGF (3)	3.055 (1.9)	0.007	0.021
PTIMPROV (3)	4.544 (4.1)	0.008	0.036
Initial likelihood	-964.582		
Likelihood with constants only	-385.308		
Final likelihood	-194.867		
$\rho^2(0)$	0.798		
$\rho^2(c)$	0.494		
N	878		

The options used in modelling:
1 = to decrease
2 = same
3 = to increase

As well as the cost incurred by the congestion charging, from Table 10.6, it seems that the shoppers to the city centre are also dissatisfied with the inconvenience of the congestion charging system (positive coefficients for CCOSTLY and INCONVEN in the utility 1 in Model-7).

On the other hand, it seems that there could be some positive impacts of introducing congestion charging on the frequency of shopping trips to the city centre since there will be less crowded / less congestion, it could be easier to go to and from the City Centre. This is evident from the model (positive coefficients of LESSCONG AND EASIGF COEFFICIENTS in utility 3 in Model-7). Moreover, public transport will have improved levels of service which might contribute to increasing frequency of shopping trips to the city centre. This is also evident from the model (positive coefficients of PTIMPROV in utility 3 in Model-7).

To further investigate the results, the values of the relative importance of each variable have also been worked out (Table 10.6). It is clear that the two constants in this model are relatively large, and statistically significant. Moreover, the location of the shoppers seems to have strong influence on shoppers' willingness to change their shopping frequency. Similarly, car ownership and the introduction of congestion charging in the city centre will also affect frequency of shopping trips. These results again support the use and further investigations of logistic regression in modelling trip generation and its applications.

10.5 SUMMARY

This chapter presents a further investigation of the utilisation of logistic regression in trip generation modelling. Two sets of models were calibrated using logistic regression techniques in this chapter to investigate impacts of the introduction of transport policies (congestion charging and parking costs) in the city centre; firstly, models for all users and secondly for car users. Revealed Preference (RP), Stated Preference (SP) and mixed RP/SP models were assessed and compared. A variable to represent the congestion charge as well as parking costs in the city centre is included in the models.

The results of the model estimations are mostly statistically significant at 95% level. The calibrated models show that as a result of the introduction of congestion charging, car users would tend to reduce the frequency of their shopping trips to the city centre. Shoppers who are living outside the outer cordon are less likely to reduce their shopping trips.

Although the introduction of congestion charging would have negative impacts on shopping trips to the city centre as a result of the costs incurred as well as the inconvenience experienced by the shoppers, it seems that there might be positive impacts of congestion charging since it would result in less congestion as well as improvements of the public transport system hence more shopping trips.

CHAPTER 11 MODELLING TRANSPORT ACCESSIBILITY IN TRIP GENERATION MODELS

In Chapter 3, various transport accessibility measures that had been previously used in models of trip generation were reviewed. As discussed, in most of those studies, the characteristics of the transport system have been included in the models but only in terms of the "observed" characteristics of the public transport services as well as transport infrastructure/network (for example, time or generalised cost functions). However, how people really think of the transport system, their perceptions and experiences that underlie attitudes, beliefs and the consequent behaviour were not considered in previous models.

In this chapter, measures of transport accessibility have been investigated for inclusion in trip generation models taking into account not only the characteristics of the transport system but also the perceived level of service of the system experienced by the individual users. The measures have only been investigated in the case of the public transport services but the approach could be similarly applicable to private transport. A limited disaggregate data set, which was collected from the Shopper Survey (SS) for shopping trips in Edinburgh as described in Section 5.2, has been used to calibrate a trip generation model which includes the accessibility parameters. The results are encouraging although the very small sample size and the fact that the data was not collected for this type of analysis prevented further investigations of the proposed methodology.

11.1 TRANSPORT ACCESSIBILITY MEASURES IN THIS STUDY

11.1.1 Introduction

Most of the transport accessibility measures reviewed in Chapter 3 included opportunities and a deterrence function in the forms of time or generalized costs. For example, an accessibility measure given by Hanson (1959) for a location (*i*) is calculated as the sum of the opportunities available at locations (*j*) factored by a deterrence function based upon the travel time between *i* and *j*. Another example is an accessibility measure for public transport given by Leake and Huzayyin (1979) which uses service frequency and zonal coverage by bus routes. More recent work in this area includes that of Daly (1997) who proposed as accessibility measure for trip generation the logsum of the distribution model, and Ortúzar *et al.* (2000) who applied stated preference tools and developed an access model using multinomial logit modelling techniques. Further discussions of those studies are given in Chapter 3.

In most of these models, transport system characteristics have only been included in terms of the "observed" characteristics of the public transport services as well as transport infrastructures/network, for example, travel time, cost of travel etc. The perception of the users of the transport system has not been reflected/included in these models. It might be possible to calibrate models which include perception of the users of the transport systems to reflect the level of transport accessibility. One main problem of using the users' perception as a factor in the model however, is that how to use the model for future prediction. In other words, how the forecasting of the perception in the future would be estimated.

11.1.2 Public transport accessibility to/from city centre

In this section, an illustration of public transport accessibility measures has been developed in an attempt to reflect the level of service of public transport as experienced and perceived by the users. The two factors that have been considered here are the distance travelled from the origin to the city centre as well as the perceived level of service of public transport as reported by the users. The distance is included in order to represent the separation between all the origins and the city centre and the perceived level of service of public transport is included to represent the users' preferences.

An investigation of the distances travelled and the frequencies of shopping trips to the city centre was carried out. Firstly, the investigation used the whole data set. Then, the respondents were split into two groups based on the frequency of shopping trips: respondents who make less than a weekly shopping trip and those who make one or more trips per week. The distances were categorised into seven categories (0-1.0, 1.1-2.0, 2.1-3.0, 3.1-4.0, 4.1-6.0 and 6.1+ miles) which were then combined into three categories (0-2.0, 2.1-4.0 and 4.1+ miles) because of the very low number of respondents in each category. It should be mentioned here that although this analysis of distances travelled and frequencies of trips is based on these categories, what we used in the trip generation model was the actual distance travelled (see Section 11.2). Table 11.1 shows the frequency of the shopping trips for each category of the distances and the number of respondents in that category (given in brackets in the table).

Table 11.1 Distance travelled and frequency of shopping trips

Distance Travelled (miles)	Trip Frequency (number of respondents)					
	All data (n=132)		Less than once a week trips (n=60)		One or more trips/week (n=72)	
	Seven categories	Three categories	Seven categories	Three categories	Seven categories	Three categories
0-1.0	2.811 (13)	2.362 (33)	0.257 (6)	0.248 (12)	5.000 (7)	3.571 (21)
1.1-2.0	2.072 (20)		0.238 (6)		2.857 (14)	
2.1-3.0	2.164 (24)	2.240 (41)	0.243 (8)	0.257 (15)	3.125 (16)	3.385 (26)
3.1-4.0	2.348 (17)		0.274 (7)		3.800 (10)	
4.1-6.0	1.647 (20)	1.241 (58)	0.357 (11)	0.242 (33)	3.222 (9)	2.560 (25)
6.1-15.0	1.456 (18)		0.173 (7)		2.273 (11)	
>15.1	0.642 (20)		0.189 (15)		2.000 (5)	
Total	1.832		0.247		3.153	

In general, it seems that lower shopping trip frequencies are observed as distance increases which is logical. When investigating the whole data set, it seems that the trip frequencies decrease as the distances increase in the case of three distance categories. However, when looking at the detailed categories, the pattern is not very clear, possibly because of the very small sample size.

When looking at the two groups (respondents who make less than a weekly shopping trip and those who make one or more trips/week) again the pattern is not very clear when investigating the detailed categories. Although the general pattern remains (i.e. lower frequencies of the shopping trips as the distances increase) for those who are observed to have made more than or equal to weekly trips (n=72 respondents), it is not the case for the first group, those who are less frequent shoppers. In that case the pattern is not consistent again maybe due to the sample size which is 60 respondents split to three classes of 12, 15 and 33 respondents respectively.

Similarly, as the perceived level of service of the public transport increases, the number of shopping trips increases (Table 11.2) for the whole data set and for the higher frequencies of shopping trips data set. However, when the number of respondents is low the pattern is inconsistent (for example there is a higher trip frequency observed with a very poor perceived level of service because of the small sample size of just eight respondents). On the other hand, for those who are making less frequent trips (less than weekly), it seems that the pattern is not clear, which is understandable.

These two variables; the distance travelled and the perception of public transport have been investigated in the trip generation models to represent the accessibility of the transport system and the perception of the users as discussed in Section 9.2. Two models were calibrated for the two groups of data sets as discussed above, despite the small sample size.

Table 11.2 Perception of public transport and frequency of shopping trips

Perception of Public Transport	Trip Frequency (number of respondents)					
	All data (n=132)		Less than once a week trips (n=60)		Weekly or more trips (n=72)	
Very poor	1.813 (8)		0.500 (1)		2.000 (7)	
Poor	1.636 (8)	1.673 (55)	0.218 (5)	0.271 (26)	4.000 (3)	2.931 (29)
Adequate	1.653 (39)		0.273 (20)		3.105 (19)	
Good	1.906 (52)	1.906 (52)	0.189 (22)	0.187 (22)	3.167 (30)	3.167 (30)
Very good	2.026 (25)	2.026 (25)	0.305 (12)	0.305 (12)	3.615 (13)	3.615 (13)
Total	1.832		0.247		3.153	

11.1.3 A measure of public transport accessibility

It is assumed here that the public transport accessibility to/from the city centre is a function of the characteristics of the transport system to/from the city centre, such as distance, fare, travelling/waiting times, etc., as well as the perceived level of service of the public transport system.

A simple form of this function could be that transport accessibility is directly proportional to how the public transport service is perceived by the users and inversely proportional to the distance to/from the city centre. Therefore, this can be expressed as:

$$acc_{ik}^{pt} \propto pt_k \tag{11.1}$$

$$acc_{ik}^{pt} \propto \frac{1}{d_{ij}} \tag{11.2}$$

where

acc_{ik}^{pt} is the public transport accessibility measure for individual k at origin i;

pt_k is the perceived level of service of public transport by individual k; and

d_{ij} is the distance between the respondent's home and the shopping location at the city centre j.

If we consider a particular destination (for example, the city centre) and distances from origins (i) which could be a zone or a household etc., using equation (11.1) and (11.2), this combined transport accessibility measure for public transport services could be expressed as:

$$acc_{ik}^{pt} = \frac{pt_k}{\exp(\lambda * d_{ij})} \tag{11.3}$$

Where:

λ is a parameter to be estimated. This parameter could be thought of as a deterrence factor to represent the separation between all the origins and the city centre. A higher value of λ indicates that distance is more of a deterrent. The exponential function is used for convenience since, unlike the power function for example, it is bounded (Kanafani, 1983). That is, acc_{ik}^{pt} does not approach infinity when d_{ij} approaches zero or increase quickly as d_{ij} decreases. The value of the parameter λ would be jointly calibrated in the model as well as other parameters of the equation for trip generation. It should be noted that the above relation could also be investigated using other forms.

11.1.4 Example illustrating transport accessibility measures in trip generation models

To illustrate the calibration of the transport accessibility measures with an example, let us assume an area with a number of origins and distances (e.g. 0, 5, 10 miles etc.) from the City Centre (which in this case represents the shopping location). Further, assume that the perception of the level of service of public transport is indicated using a 5 points scale ranging from 1 (lowest perceived level of service) to 5 (highest perceived level of service). Table 11.3 shows the calculation of the transport accessibility measures as discussed in Section 11.1.3 above using:

(1) The distance from each origin to the City Centre (d_{ij});

(2) The perceived level of service of public transport (pt_k); and

(3) A combined transport accessibility measure (acc_{ik}^{pt}) as discussed earlier in this section.

Table 11.3 The calculation of public transport accessibility measures

Indicator of Public Transport Services pt_k	Distance between Zones j and i (miles) d_{ij}	$acc_{ik}^{pt} = \dfrac{pt_k}{\exp(\lambda * d_{ij})}$ (λ=0.10)
1	0	1.00
1	5	0.61
1	10	0.37
5	0	5.00
5	5	3.03
5	10	1.84

It should be noted that the perceived level of service of public transport by individuals is considered, in the model, as a continuous variable ranging from 1 to 5, and then it is combined with the distances from each origin to the shopping location. The lowest possible value for transport accessibility measure in this case approaches 0 where the perceived level of service of public transport is at its lowest value (i.e. pt_k=1) and the distance between the origin and the city centre is very large (i.e. d_{ij}>10 miles or so). On the other hand, the maximum value of this measure is 5 where the perceived level of service of public transport reaches the highest value (i.e. pt_k=5) and the distance between the origin and the city centre is very small (i.e. the origin is within the city centre and d_{ij}=0). In this way it is seen that, in general, as distance from the origin to the city centre increases, the combined transport accessibility measure decreases. Similarly, as the perceived level of service decreases, also the transport accessibility measure decreases.

These accessibility measures were included in the linear trip generation models that will be presented in Section 11.2 and the results show that they are statistically significant at the 95% level of confidence.

11.1.5 Other possible accessibility measures

It should be noted here that distance is not only a possible measure to represent the characteristics of a transport system but obviously it is an easy factor to measure. It is also possible, however, to use other factors instead of or as well as distance. These factors could include costs, time, and so on. For example, one can use costs and time instead of the distance as in the following equation:

$$acc_{ik}^{pt} = \frac{pt_k}{\exp(\lambda_c * C_{ij} + \lambda_t * t_{ij} + \cdots)} \tag{11.4}$$

where

C_{ij} is the costs (e.g. public transport fare) from the respondent's home to the shopping location;

t_{ij} is the journey time from the respondent's home to the shopping location; and

λ_c and λ_t are specific coefficients associated with cost and time respectively.

But the calibration of this model would require data on costs and/or time of travel between the origin and the shopping location (i.e. city centre) which were not available in this current survey.

These forms of models could be further investigated in future research. The following section discusses the results for the transport accessibility measure.

11.2 SHOPPING TRIP GENERATION MODELS WITH TRANSPORT ACCESSIBILITY MEASURES FOR PUBLIC TRANSPORT SERVICES

The data used in this analysis were obtained from the Shopper Survey for shopping trips in Edinburgh as discussed in Section 5.3 and Section 11.1.2. It should be noted here that this data was not collected for the purpose of the current investigation. Therefore, the results obtained are not too solid as discussed earlier in this section. Moreover, the small sample size was also a contributing factor to the less than ideal quality of the results. Data from individuals whose main reason for the journey was shopping for groceries or other items were used to calibrate the models. Tourists, and respondents who work or use other services, were excluded from the data, since shopping in the city centre was not their main journey purpose. In total, there were 132 respondents included in the dataset, which is not a large sample for model calibration to start with, but it was further grouped into two groups based on the frequency of shopping trips as discussed in Section 11.1.2. The independent variables in these models are presented in Table 11.4

Table 11.4 Independent variables in the shopping trip generation models

Variables	Description
EXPEND	A continuous variable which describes the respondent's expenditure per shopping trip.
AGE2	A dummy variable which takes the value of 1 if respondent's age is in the 26-54 category, 0 otherwise.
AGE3	A dummy variable which takes the value of 1 if respondent's age is in the over 55 category, 0 otherwise.
GENDER	A dummy variable which takes the value of 1 if respondent is a male, 0 female.
CAR1	A dummy variable which takes the value of 1 if respondent's household owns one car, 0 otherwise.
INV_DIST	A continuous variable which calculated as the inverse of the distance between the respondent's home and the shopping location.
PT	A continuous variable which describes the respondent's perception of current public transport services to and from the City Centre.
ACC	A continuous variable which corresponds to the accessibility as discussed in Section 9.1.3 and Equation 9.3.

The respondent's expenditure per shopping trip would be expected to have a negative effect on the trips. This is because the more any individual spends on one shopping trip the less number of trips would be expected to be made by him/her to the city centre. People in age group two (26-54) would be expected to have a negative impact on the shopping trips as most of them should be in the work force. On the other hand, people in age group three (55 and more) would be expected to make more shopping trips. An increase in the distance between home and shopping area should make the respondent make fewer shopping trips. When the respondent's perception of public transport services is higher, the respondent is expected to make more trips. The accessibility measure which combines the distance and the respondent's perception of public transport is expected to have a positive impact on the trips. However, and as discussed, probably because of the small sample size of the data and after many attempts, it was not possible to calibrate a statistically significant model which included this combined function. Therefore it was decided to only investigate the distance and the perception as two independent variables in the model but not the combined function.

Two set of models were calibrated using the shopping trip data of this survey. Firstly, basic trip generation models with the basic variables (i.e. expenditure, gender, age and car ownership) as shown in Table 11.5. The first model in this case (Model-1) was calibrated for the whole data set. Then the data was classified into two sets based on the frequency of shopping trips (i.e. less than weekly trips and equal to or more than weekly trips). Secondly, trip generation models with the above variables as well as two extra variables (distance and perception of public transport), which represent the transport accessibility in the same way as discussed earlier in this chapter, were calibrated. In this case, also three models were used: one for the whole data set, another for the less than weekly shopping trips and another for the equal to or more than weekly shopping trips. However, for the accessibility function itself, it was not possible to successfully calibrate using this data set (Table 11.6).

Table 11.5 Linear regression trip generation models

Variables	Model-1	Model-2 (<weekly)	Model-3 (>=weekly)
Constant	1.884 (4.4)	0.193 (4.1)	3.267 (5.6)
EXPEND	-0.001 (-0.4)	2.90E-04 (1.2)	3.49E-03 (0.9)
AGE2	-0.467(-1.0)	-0.010 (-0.2)	-0.523 (-0.8)
AGE3	0.187(0.4)	0.051 (0.9)	-0.200 (-0.3)
GENDER	-0.195 (-0.5)	0.050 (1.0)	-0.184 (-0.3)
CAR1	0.620 (1.6)	0.052 (1.1)	0.048 (0.1)
R^2	0.044	0.068	0.023
n	132	60	72

Table 11.6 Linear regression trip generation models with accessibility variables

Variables	Model-4	Model-5 (<weekly)	Model-6 (>=weekly)
Constant	0.876 (1.2)	0.246 (2.5)	1.394 (1.4)
EXPEND	-0.001 (-0.5)	3.10E-04 (1.3)	1.38E-03 (0.4)
AGE2	-0.489 (-1.1)	-0.028 (-0.5)	-0.513 (-0.9)
AGE3	0.065 (0.1)	0.049 (0.9)	-0.381 (-0.7)
GENDER	-0.202 (-0.5)	0.053 (1.1)	-0.184 (-0.3)
CAR1	0.539 (1.4)	0.049 (1.0)	0.112 (0.2)
PT	0.096 (0.6)	-0.019 (-0.8)	0.328 (1.6)
INV_DIST	2.052 (3.1)	0.084 (1.1)	2.167 (2.3)
R^2	0.119	0.093	0.134
n	132	60	72

From Table 11.5 it appears that most of the independent variables are not statistically significant at 95% level. This may be due to the small sample size of the data set used in this section. In addition, recall the discussion in Section 5.3, there were some missing data in the survey and some assumptions about trip patterns were made which might have affected the results.

However, from Table 11.6, it is shown that the three models including some kind of accessibility measure (i.e. the distance and the perception of the public transport services) have coefficients with logical signs as well as a slightly improved R^2 values. The R^2 values are all too low, suggesting that a linear relation may not warranted. This also could be partly due to other relevant factors, such as income or cost of travel has not been included in the models. It is noted here that with all the efforts it was not possible to obtain a more statistically significant model with or without the accessibility function which has been discussed in this chapter due to data problems.

From general inspection of the results in Table 11.5 and Table 11.6, it seems that shoppers whose age group is 25-54 appear to be making less shopping trips than other groups (negative sign for AGE2) in all the models. This could be because people in this group are in employment and have less time for shopping. Car owners (CAR1) seem to be more frequent shoppers than those who own no cars, which could reflect the socio-economic status of the households.

From Table 11.6, the inverse of the distance from home to city centre (INV_DIST) has a positive sign and is statistically significant at the 95% level; this indicates that as distance decreases individuals make more trips to the city centre. The respondents' opinion of public transport services has a positive influence on the number of shopping trips made (positive sign of PT in Table 11.6). This indicates that people makes more shopping trips to city centre by public transport as the level of satisfaction increases.

The overall statistical performance of the models is poor. The signs of some of the variables are not logical and would acquire further investigation using a different data set. For example the EXPEND variable which appears to have a positive sign where it is expected to have a negative sign. However the positive outcome from this analysis is that there are evidences that the factors which represent the accessibility of the transport system such as the distance from the origin to the shopping centre as well as the perception of the users of the transport system are both statistically significant and seem reasonable to include in the model.

11.3 SUMMARY

Transport accessibility measures for public transport have been investigated and included as independent variables in trip generation models using disaggregate data. The approach appears to be logical and interesting. In this case, the distance to the city centre has been the only relevant variable which could be used to represent the accessibility of the transport system. The perceived quality of the public transport services has also been included in the models to represent the perception of the users.

Two sets of models have been calibrated. Firstly a set of models were calibrated with the basic and conventional factors of trip generation models only. In this case, a model was calibrated using the whole data set and two models were calibrated classifying the frequency of shopping trips to less than weekly and equal to or more than weekly trips. The second set of models includes variables which are related to the accessibility of the transport system, in this case the distance and the perception of the users of the transport system.

Although the approach seems rational and appealing the data set which has been used to investigate this concept is not the most appropriate data, hence the results are not statistically significant. Further work in this area is definitely needed.

CHAPTER 12 DISCUSSIONS

12.1 WHY USING LOGISTIC ANALYSIS TO MODEL TRIP GENERATION?

There are a number of reasons which justify the investigation and adoption of logistic regression to model trip generation and also the inclusion of policy factors in trip generation models. These include:

1. The main approaches which are used in modelling and predicting trip generations to date have had the least attention from modellers and analysts of travel demand forecasting. Whilst there has been a huge amount of research and investigations in the literature and methodologies of mode, route, destinations and departure time choice modelling (see for example Garson, 2002, Ortúzar, 1983; Bhat, 1995; Bhat, 1998a; Ortúzar and Willumsen, 2001, Bhat, 1998b; Saleh and Farrell, 2005, Yai *et al.*, 1997, Daly, 1997) there have been very little, if any advances on the techniques and approaches of modelling trip generation.
2. Moreover, since the four stages models are all dependent and related, it does make sense to use similar techniques and principles of modelling of the four stages. In reality, while mode choice modelling, destination choice and route choice mostly employ logistic regression modelling, trip generation still only employs category analysis and linear regression analysis techniques despite all the well recognised and documented drawback of such techniques.
3. It has been well recognised and documented that policy and accessibility factors do not only affect mode, destination and route choice but also trip productions and attractions (see for example Hanson (1959), Freeman (1976), Leake and Huzayyin (1979), Cohn *et al.*, 1996; Daly, 1997 and others). The inclusion of policy factors in trip generation therefore is badly needed. Moreover, it has been acknowledged in the literature that the main drawback of trip generation models is the lack of policy and accessibility measures.

4. Trip generation is the first stage of the conventional four stage transport model. Any errors in the prediction at this stage will therefore be carried over to other stages and affecting their accuracies. Therefore it is important to investigate and improve the prediction and modelling of trip generations.

The main aim of this research has been to investigate possible methodologies to improve performance of trip generation modelling. In order to achieve this aim a number of objectives have been defined and investigated as discussed in Chapter 1 and concluded in Chapter 13. This chapter presents a discussion of the main findings and investigations of this thesis.

12.2 GAPS IDENTIFIED IN PREVIOUS RESEARCH

As discussed in Chapter 1, limitations in trip generation techniques and analysis have been widely recognised in the literature, there have been various investigations of alternative techniques. Logistic regression analysis, which has been extensively used in other stages of travel demand modelling (mode, route, destination and departure time choices), can overcome some of the limitations of linear regression analysis (i.e. the assumption of linearity of independent variables with the dependent variable) and category analysis (i.e. the requirement of large sample size). It can bring a potential improvement in the performance over the conventional techniques and provide a behavioural framework that directly links the number of trips to utility-based consumer and decision-making theory.

In the meantime, fewer investigations have been focused on including variables that represents transport policies in trip generation models which can affect the trips generated. As to the data used, most trip generation models are calibrated from aggregate revealed preference (SP) data despite the growing applications of other sources of data such as disaggregate stated preference data. SP techniques offer the opportunity to modellers to test impacts of policy measures on travel behaviour. Finally, this study will attempt to include both the physical characteristics of the transport system and the perceived level of service of the system in the trip generation models.

12.3 DEVELOPING METHODOLOGY FOR USING LOGISTIC ANALYSIS TO MODEL TRIP GENERATION

The logistic regression analysis for work trip generation using NTS data is presented in Chapter 6. This includes binary, multinomial and nested logit models. The results show that in principle logistic regression modelling can be used to model trip generation. This approach will overcome some of the limitations of linear regression and category analysis methods as discussed. In the binary model, it is assumed that the dependent variable is a binary variable to represent the household making work trips or not. The MNL model assumes that the probability of a household making a certain number of work trip(s) is a function of a number of independent variables. The best fit of the models was obtained with the trips assigned as 0 trips, 1-2 trips, and 3 or more trips. A nested logit (NL) model was calibrated which assumes trip makers trade off between making no trips against making 1 or more trips at the first level and at the second level between 1-2 trips against 3 or more trips.

12.4 PERFORMANCE OF THE MODELS

12.4.1 The performance of logit models

The results in Chapter 6 show that all the calibrated logit models are all statistically significant at a reasonable level of significance with an overall goodness of fit. Bearing in mind the limitations of data (as discussed in Chapter 5), all the independent variables in the logit models have logical signs and most of them are statistically significant. The MNL model shows the best $\rho^2(0)$ result than the NL model with $\rho^2(0)$ value equals to 0.215 while it is equal to 0.149 in the NL model. The theta parameter in the NL model has an acceptable value of 0.978 which suggests that the MNL is most appropriate in this case. It was also possible to model trip generation using binary logit models as discussed above. Of the three binary models calibrated, BLM-3 has the best $\rho^2(0)$ with a value of

0.389 where the number of full time workers has been included as three dummy variables and the number of cars was treated as a continuous variable.

12.4.2 The performance of category analysis and MCA models

Four techniques of the category analysis, including MCA have been investigated using NTS data as discussed in Chapter 7. The results of the MCA analysis show statistically significant models with logical signs of the independent variables. Taking category analysis as the base for the comparison, and using the Residual Sum of Squares (RSS) or Error Sum of Squares to assess the overall performance of the models (Table 12.1), it appears that the results obtained from the MCA_1 model produce the largest sum of errors in the family of category analysis. That is 11.1% higher than that obtained from the base CA technique. The MCA_2 model does not provide noticeable improvement of the RSS (0.1% higher than that obtained from the base CA model) over the basic category analysis model. However, the MCA_3 produces the most reliable model with least RSS values (7.7% lower than that resulting from the base CA model). Therefore, the MCA_3 has been recommended to be used as the best technique in this family of models.

Table 12.1 Comparison of RSS of category analysis techniques

Models	RSS	RSS – Diff from CA %
CA	1,904	-
MCA_1	2,116	11.1%
MCA_2	1,905	0.1%
MCA_3	1,758	-7.7%

12.4.3 The performance of linear regression analysis

Three linear regression models (LM-1, LM-2 and LM-3) have been calibrated from the NTS data and the R^2 values of the three models are 0.322, 0.326 and 0.272 respectively. The most significant R^2 value here is resulted in LM-2 which has continuous variables for the number of full time workers and the number of cars in the household. Therefore, this model was the selected as the best linear regression model to be used in Section 8.5 for the prediction and comparisons of trip generation models using the three techniques.

12.4.4 Comparison of linear regression, category analysis and logistic analysis models

In this section a comparison of the results obtained from the three modelling approaches are discussed. The compared models are: the best linear regression model (LM-2), the basic category analysis model (CA), the best multiple classification analysis (MCA_3), the best binary logit model (BLM-3), multinomial logit (MNL) and nested logit (NL) models. The analysis of the predictions using these models (linear, category analysis and logit) is presented in Chapter 8. Table 12.2 below shows the comparison of RSS of models from all the three techniques. The results show that the least RSS values have been obtained from the MNL model with a value of 1,713, making it best performing model of all (Table 12.2). This is followed by the linear regression model (LM-2) and lastly, the MCA_3 models with their RSS value 1.1% and 2.6% higher than that of the MNL model. The RSS results of conventional category analysis, the binary logit model and NL model are 11.2%, 18.9% and 13.4% greater than that of the MNL (the best performing model) respectively.

Table 12.2 Comparison of RSS of models from all the three techniques

Models	RSS	RSS - Diff from MNL %
LM-2	1,731	1.1%
CA	1,904	11.2%
MCA_3	1,758	2.6%
BLM_3	2,037	18.9%
MNL model	**1,713**	-
NL model	1,942	13.4%

The results provide some evidence to support the appropriateness of using logistic regression analysis for trip generation modelling. The results indicate that LM2 and MNL are almost identical in their predictions of numbers of trips, and that the difference in the RSS is 1%. As the linear regression analysis is the best known techniques so far for trip generation predictions, the result is promising in consideration of the limitation of use of the logistic regression analysis such as data suitability etc. Presumably if the data used was collected specifically to calibrate this type of models the results might have been even more convincing. Further research and investigations are still needed still to establish whether this improvement is worthwhile for its use in trip generation prediction or not.

As mentioned before, using logistic regression would also have the added value of allowing the prediction of the trip frequency as well as the number of trips.

The three MCA methods have been investigated and compared using NTS data. The results of this research support those results obtained by Guevara and Thomas (2007) that MCA_1 model, which is most commonly used in applications of trip generation modeling, is the least accurate in the family of MCA. MCA_2 model did not produce accurate results compared to MCA_3 which showed the most accurate results. Therefore, MCA_3 has been recommended for use in practical applications as the preferred category analysis method.

12.5 TYPES OF VARIABLES

To ignore the impacts of transport measures and policies at the trip generation stage and only consider them at later choice decisions would be resulting in inaccurate predictions at this, and all subsequent stages. This has been one of the main criticisms of trip generation models. While there are a lot of empirical evidences that these schemes have resulted in a reduction of number of shopping and other trips to the central areas, most current trip generation models still do not include these types of variables. In Chapter 9, linear and logistic regression models of trip generation (shopping trips) have been calibrated using the Edinburgh Household Survey data, taking into account parking costs as a transport policy measure.

To assess the improvements of the models as a result of including policy measures (parking costs in this case), the liner regression trip generation model for car users shows a 6% (Table 9.6) improvement in the prediction of trip generation than the models without the parking costs. In the binary logit model for car users (Table 9.9) this variable (parking costs) also shows statistical significance (a negative sign and t-value = -2.1).

The results from the models suggest that policy measures which would be implemented in the city centre should have an impact on shopping trip generation. In this case, an increase in parking costs results in people making less frequent shopping trips to the city centre.

12.6 DATA TYPES

Most trip generation models are calibrated from aggregate revealed preference (RP) data despite the growing applications of other sources of data such as disaggregate stated preference (SP) especially in travel demand forecasting, mainly because of the nature of trip generation models. SP techniques offer the opportunity to modellers to test impacts of policy measures on travel behaviour. Therefore, in principle there is no reason why these techniques cannot be used in trip generation modelling, especially if logistic regression analysis is used. It would be very useful to use stated preference techniques to investigate impacts of transport policies on trip generations as well as other choice models.

In order to achieve this, the SP data from Edinburgh Household Survey is used to calibrate mixed RP/SP logistic regression models for trip generation taking account of introducing road user charging as a policy measure as presented in Chapter 10.

The results show that the model calibrated using SP data improves the $\rho^2(0)$ results by 72% than the model calibrated using RP data ($\rho^2(0)$ increases from 0.258 to 0.178), which is a significant improvement.

In addition, in this research the technique of mixed RP/SP in modelling trip generation has been investigated. For the mixed RP/SP model, the results show a scaling parameter of 0.822 suggesting that the SP data have more random noise than the RP data. Although the results are not very statistically significant here but again this has been a challenging achievement and further research should be developed in this area.

The calibrated models show that as a result of the introduction of congestion charging, car users would tend to reduce the frequency of their shopping trips to the city centre, which is logical. Moreover, shoppers who are living outside the outer cordon are less likely to reduce their shopping trips. However, the introduction of congestion charging would have negative impacts on shopping trips to the city centre as a result of the costs incurred. The results of the model estimations confirm the potential of using stated preference data in trip generation models.

12.7 THE ACCESSIBILITY FUNCTION

Accessibility of the transport system has been recognised and investigated in the literature but has always been limited to variables representing the characteristics of the transport system. Variables which represent the perceived level of service of that system have not been investigated in previous research. In this research, a public transport accessibility measure is calibrated as a function of the distance from the city centre and the perceived level of service of the public transport system by the users using the Shoppers' Survey data. These results are presented in Chapter 11.

The proposed accessibility measure (Model-4b in Table 11.6) shows a logical sign, i.e. when accessibility increases more trips are expected. Although the approach seems rational and appealing, the results in this case are not statistically significant. However it seems that at least there is evidence for the importance of representing the accessibility of the transport system as well as the perception of the users of the transport system. Further work in this area is therefore recommended.

12.8 SUMMARY

This research shows that logistic regression analysis is an appropriate technique to model trip generation and underlines the importance and relevance of including transport policy measures and accessibility in trip generation models. These two areas have been identified in the literature but not much researched. In this research logistic regression analysis has been used to calibrate trip generation models which also include policy measures. The results also confirm the potential of using stated preference data in trip generation modelling.

As mentioned earlier, the results from logistic regression analysis only improve slightly in RSS from that of linear regression model. Although logistic regression analysis provides an alternative methodology to trip generation modelling, with the limitations of the drawbacks of using the method such as data suitability, further research and investigations are still needed to establish the level of improvement of logistic regression analysis over linear regression analysis in trip generation prediction.

In addition, the investigations in this thesis confirm that MCA_1 method, one of the most commonly used techniques in trip generation models, is the least accurate model in the family of MCA and that MCA_3 proved to be the most accurate method. Therefore, MCA_3 should be recommended for use as the preferred category analysis method. Next Chapter concludes the work.

CHAPTER 13 CONCLUSIONS

13.1 INTRODUCTION AND GENERAL CONCLUSIONS

Trip generation is defined as the number of individual trips generated in a given period of time. The purpose of this stage is to predict the total number of trips which are generated from and attracted to each zone, as a function of its land-use and socio-economic characteristics. Trip generation analysis, however, has limitations in terms of the techniques, data used and type of variables. These limitations have been recognised in the literature and acknowledged that they limit the efficiency of trip generation models to produce accurate predictions.

Firstly, trip generation analysis has been mostly carried out using linear regression analysis and category analysis. Both approaches have their strengths and weaknesses. Linear regression analysis is easy and simple techniques and there are statistical tests for the goodness of fit of the model. However, the assumption of linearity of each of the independent variables with the dependent variables is restrictive. Unreasonable predictions from the models can be obtained as a result of the lack of built-in upper and lower limits to the number of trips, or could result in negative number of trips when the covariate values are relatively low. In addition, the assumption that the number of trips is approximately continuous can be questioned when typical values for the number of trips are relatively low. The link between number of trips and covariates in a linear regression lacks a behavioural justification such as supported by the theory of random utility (e.g. Ben-Akiva and Lerman, 1985).

In category analysis on the other hand, the large sample size required to calibrate the trip rates as well as the absence of statistical tests for the overall goodness of fit of the models undermines this method reliability. Multiple classification analysis (MCA) methods provide improved techniques to overcome some of the shortcomings of category analysis approach, however still the main limitations of category analysis methods apply.

Another main criticism of trip generation models is the absence of any variables that represent transport policies and measures that affect the trips generated (e.g. public transport, pricing and parking policies). The impacts of these policies are always considered in mode, route, destination and departure time choices. However, not many investigations of their impacts on trip generations have been reported. Failing to include effects of transport measures and policies at the first stage (TG), would certainly result in inaccurate predictions at all subsequent stages.

Type of data used in trip generation models are mainly revealed preference data despite the growing applications of other sources of data such as stated preference. Stated preference techniques offer the opportunity to modellers to test impacts of policy behaviour.

Logistic regression analysis which has been used in modelling other travel choices such as mode, route and destination provides an appropriate approach which could overcome many of the restrictive limitations of the current trip generation techniques. However, to the knowledge of the author, not many applications in trip generation modelling using logistic regression have been reported to date.

13.2 ACHIEVING THE AIM AND OBJECTIVES OF THIS RESEARCH

The aim of this research is to investigate possible methodologies to improve performance of trip generation modelling. In order to achieve this aim a number of objectives have been defined as discussed below.

The first objective of this research has been to investigate appropriateness of logistic regression analysis for modelling trip generation.

In order to do that, a number of data sets have been identified and analysed to carry out the investigations. National Travel Survey (NTS) data has been used to calibrate trip generation models using logistic analysis. National Travel Survey is a household survey of travel covering residents of Great Britain (GB) and includes information on the purpose of each trip made, the modes of transport, the timing of the trip, the origin and destination and demographic data. The logistic regression models considered include binary logistic models, multinomial logit (MNL) and nested logit (NL) models as presented in Chapter 6. In the binary model, it is assumed that the dependent variable is a binary variable to represent the household making work trips or not. In the MNL model, it is assumed that the probability of a household making a certain number of work trip(s) is a function of a number of independent variables. A number of trials for the structure of the model and for the allocation of variables to each utility have been carried out. The best fit of the models was obtained with the trips assigned as follows: {0 trips, 1-2 trips, 3 or more trips}. A nested logit (NL) model was also calibrated with the nested structure. In this case, trip makers are being assumed to be trading off between making no trips against making 1 or more trips. Then, at the second level, a trade off between 1-2 trips against 3 or more trips is assumed.

The results of this analysis are very encouraging as an appropriate methodology has been devised to model trip generation using each of the three approaches of logistic regression. The results show all the independent variables in the calibrated models have logical signs and most of them are statistically significant.

The second objective of this research is to investigate, analyse and compare trip generation models using logistic regression, linear regression and category analysis including multiple classification analysis.

The same data set has been used to calibrate trip generation models using the conventional (linear regression and category analysis) and presented in Chapter 7. A number of multiple classification analysis techniques which have been recently developed (Guevara and Thomas, 2007) but not widely empirically tested. Trip generation analysis of work trips have also been calibrated and analysed using MCA (Chapter 7). The results also show statistically significant models with logical signs of the independent variables and a reasonable overall goodness of fit of the model.

The real test of the models however, would be the accuracy of the predictions. As discussed in Chapter 5, about 73% of the data was used to calibrate the models and the 27% was used for model prediction. A comparison of the model predictions using all techniques (that is linear regression, category analysis including multiple classification analysis and logistic regression models (binary, MNL and NL models)) were performed and the results are presented in Chapter 8. The results show that the MNL model outperformed all the other models, followed by the linear regression model (LM_2) and MCA_3 models. These three modelling approaches performed better than the other techniques (i.e. binary logit and nested logit models).

These results provide strong evidence for firstly, the appropriateness of using logistic regression analysis for modelling trip generation and secondly, the prediction of trip generation is best using the MNL model and linear regression analysis. Using logistic regression would also have the added value of allowing the prediction of the trip frequency as well as the number of trips.

The three MCA methods have been investigated using NTS data. The results in this research support those obtained by Guevara and Thomas (2007) that MCA_1 method, which is most commonly used in applications of trip generation modeling, is the least accurate model in the family of MCA. MCA_2 method also produced no accurate results compared to MCA_3 which proved to be the most accurate method, and therefore should be recommended for use as the preferred category analysis method.

The third objective of this research is to investigate the impacts of including factors to represents transport policy in the trip generation models on their performance.

In order to investigate that, the Edinburgh Household Survey (HS) data has been analysed to carry out the investigations. The survey included information on the socio economic data including age, gender, car ownership and social grade, mode of travel for shopping and location of residence. Respondents were also asked to report on their non-food shopping trip frequency into the city centre in a week and the parking costs.

Linear regression and logistic analysis have been utilised to calibrate shopping trip generation models. The results from the models (Chapter 9) suggest that policy measures which would be implemented in the city centre should have an impact on shopping trip generation. For example, in this case an increase in parking costs results in people making less frequent shopping trips to the city centre.

In this research, the fourth objective is to investigate the use of stated preference data for calibrating trip generation models.

In order to achieve this, the SP data from Edinburgh Household Survey is used to calibrate mixed RP/SP logistic regression models for trip generation taking account of introducing road user charging as a policy measure. These results are presented in Chapter 10. The calibrated models show that as a result of the introduction of congestion charging, car users would tend to reduce the frequency of their shopping trips to the city centre. Shoppers who are living outside the outer cordon are less likely to reduce their shopping trips. Although the introduction of congestion charging would have negative impacts on shopping trips to the city centre as a result of the costs incurred. The results of the model estimations confirm the potential of using stated preference data in trip generation models.

Finally, in this research therefore, the inclusion of transport accessibility measure in trip generation models is investigated and analysed.

A public transport accessibility measure is calibrated as a function of the distance from the city centre and the perceived level of service of the public transport system by the users using the Shoppers' Survey data. These results are presented in Chapter 11. Although the approach seems rational and appealing the data set which has been used to investigate this concept is not the most appropriate data, hence the results are not statistically significant. However it seems there is evidence that the factors which represent the accessibility of the transport system such as the distance from the origin to the shopping centre as well as the perception of the users of the transport system can affect trip generation and hence seem reasonable to include in the trip generation models. Further work in this area is needed.

13.3 RESEARCH NOVELTIES: ADDITION TO KNOWLEDGE

In this thesis a number of novel investigations and additions to the knowledge in trip generation analysis and modelling have been carried out. Trip generation analysis has been under researched; most of recent research efforts in travel demand forecasting have been concentrated in the other stages (mode, route, etc.). Therefore, the techniques of trip generation modelling and the data types have not been developed a lot over the past few decades. Despite the known limitations of linear regression analysis and category analysis, limitations of variable types as well as limitations of revealed preference data, not much attempts in using other techniques or data types have been made. A number of additions to knowledge are reported in this thesis and are summarised below:

1) This research defines a framework for modelling trip generation using logistic analysis. This is an interesting research matter, and could also achieve improvements in trip generation predictions.
2) A number of multiple classification analysis techniques which have been recently developed but not widely empirically tested, are used to calibrate and analyse work trip generation models. The results are assessed and conclusions on the best techniques are derived.
3) Trip generation models including independent variables that represent transport policies (such as parking pricing) have been calibrated. This is another shortcoming of current trip generation models which have been recently strongly recognised.
4) The use of stated preference data in investigating preferences and attitudes in other stages (mode, route, etc.) has shown great improvements. However, trip generation models mostly rely on the use of

revealed preference data. In this thesis trip generation models have been calibrated using mixed SP/RP techniques.

5) Finally, the research also investigates modelling transport accessibility into trip generation models by including a public transport accessibility measure, which reflects the transport users' perceived levels of service of public transport.

13.4 IMPLICATIONS OF THIS RESEARCH FOR POLICY AND PRACTICE

From policy point of view, the main message of this work is that trip generation models should include the impacts of policies implemented in order to obtain realistic results. Results of models which do not include these policies should be taken with care.

From practice point of view, the research shows that there are further opportunities to improve trip generation models by using different types of data such as stated preference data. Also, some techniques have shown better performance in terms of the overall statistical significance of the models, and these should be considered by the practitioners. Most specifically here, the MCA_3 has been recommended to be used as the best technique in the family of category analysis.

In addition, this research shows gaps in current techniques of modelling trip generation. This underlines the importance of investigating the appropriateness of modelling techniques in general. It should be noted that most modelling approaches are developed for certain specific studies and situations, and they are usually adopted to be used in other situations. Policy and decision makers have to be careful when they are using and interpreting results from various models and also when they are selecting modelling techniques and approaches.

Finally, logistic regression could provide an appropriate tool to trip generation modelling. The applications of category analysis should be further enhanced to take account of recent development (MCA_3 and MCA_4) which shows more statistically significant results.

13.5 LIMITATIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

In this research, logistic regression analysis techniques have been investigated for modelling a number of trip generation models. A number of data sources have been used including National Travel Survey and Edinburgh Household Survey data. There were some limitations with this data. Further investigations for the appropriateness of logistic analysis in trip generation modelling using other sources of data and journey purposes would be recommended.

While the policy conclusions which can be drawn from such an analysis are clearly limited, however, the inclusion of policy factors and accessibility measures in trip generation models are clearly important and deserves further research. It is not very clear how this method of trip generation would be fully adopted in practice, however, it is always the case that new applications and commercial software packages become available much later than the theory. That might explain why such methodology has not yet been adopted in practice so far since there has been recognition of the limitations of the conventional trip generation models for the last ten years or so.

Perceived transport accessibility measures have been limitedly investigated for shopping trip generation models using Shoppers' Survey in the city of Edinburgh. It was not possible to calibrate a statistically significant trip generation model which includes an accessibility function because of the limitation of data. In future research, it is recommended that further transport accessibility measures to be investigated for inclusion in trip generation models. Distance and perceived level of service of public transport are two possible factors to represent accessibility. Other relevant variables may include time and cost of the journey.

The data used in this study, mostly National Travel Survey, is limited in terms of quality and quantity. It is recommended therefore that further surveys and data collection to be carried out for the calibration of trip generation models in order to improve model performance.

Three methods of MCA analysis have been calibrated and analysed in this thesis. There is however a further method in this family of techniques (i.e. MCA_4) which could be investigated using the same data set and the results to be compared.

Impacts of limited number of transport policies on accessibility and trip generation have been investigated. Further research and investigations of other transport policies on trip generation are also recommended.

This investigation of using logistic regression model in trip generation is very attractive in principle, as it handles generation and frequency of trips simultaneously. The approach can further be enhanced to combine distribution (i.e. choice of destination) and mode choice in one model. Further forms of accessibility measures could also be investigated. The results of this investigation however indicate that LM2 and MNL are almost identical in their predictions of numbers of trips, and that the difference in the RSS is 1%. While this is a good result given the limitations of data suitability to this type of analysis, it is encouraging enough to carry out further research in this direction to investigate appropriateness of logistic regression to trip generation modelling.

Finally, the continuing challenges which are faced with travel demand models are derived mainly from the quality of the data. Data usually consists of a sample of observations taken from a certain population on a limited number of their attributes or characteristics. The less relevant the data to the investigated problem the less reliable the results would be. In this research, Nation Travel Survey (NTS) and Household Survey (HS) data in Edinburgh were used to calibrate trip generation models for work trips and shopping trips. The NTS data was very aggregate with large variations (e.g. in income and car ownership) which would hinder the capture of greater amount of true behavioural variability in travel choices. In the HS on the other hand, data was very general (e.g. no information on income or employment status was available).

REFERENCES

Abkowitz, M. (1981) An analysis of the commuter departure time decision. *Transportation* **10**, 283-297.

Agresti, A. (1990) *Categorical Data Analysis*. John Wiley& Sons, New York.

Agyemang-Duah, K. and Hall, F.L. (1997) Spatial transferability of an ordered response model trip generation. *Transportation Research* **31A**, 389 – 402.

Agyemang-Duah, K., Anderson, W.P. and Hall, F.L. (1995) Trip generation for shopping travel. *Transportation Research Record* **1493**, 12-20.

Allaman, P.M., Tardiff, T.J. and Dunbar, F.C. (1982) *New approaches to understanding travel behaviour.* National Cooperative Highway Research Program (NCHRP) Report 250, Transportation Research Board (TRB), National Research Council, Washington, D.C.

Allison P.D. (1999) *Logistic regression using the SAS® system: theory and application.* SAS Institute Inc., Cary.

Anderson M.D. and Malave, D.N. (2003) Dynamic trip generation for a medium-sized urban community. *Transportation Research Record* **1858**, 118-123.

Atherton, T. and Ben-Akiva, M.E. (1976) Transferability and updating of disaggregate travel demand models. *Transportation Research Record* **610**, 12–18.

Badoe D.A. and Chen C.C. (2004) Unit of analysis in conventional trip generation modelling: an investigation. *Canadian Journal of Civil Engineering* **31**, 272-280.

Bagley M.N., Mannering J.S. and Mokhtarian P.L. (1994) *Telecommuting Centres and Related Concepts: A Review of Practice. Research Report No. UCD-ITS-RR-94-4.* Institute of Transportation Studies, University of California, Davis, CA. [online], Available from: http://repositories.cdlib.org/itsdavis/UCD-ITS-RR-94-04. [Accessed 1st February 2006].

Wilbur Smith and Associates (1964). Baltimore Metropolitan Area Transportation Study (BMATS). Maryland State Roads Commission.

Banister, D. (2003) Critical pragmatism and congestion charging in London. *International Social Science Journal* **176**, 249–264

Barber, G. (1986) Aggregate characteristics of urban travel. In S. Hanson (ed.), *The Geography of Urban Transportation*. The Guilford Press, New York.

Barmby, T. and J. Doornik (1989) Modelling trip frequency as a Poisson variable. *Journal of Transport Economics and Policy* **23**, 309–315.

Baxter, R.S. and Lenzi, G. (1975) The measurement of relative accessibility. *Regional Studies* **9**, 15-26.

Ben-Akiva, M.E. (1974) Structure of passenger travel demand models. *Transportation Research Record* **526**, 26-42.

Ben-Akiva, M.E. and Lerman, S.R. (1979) Disaggregate travel and mobility-choice models and measures of accessibility. In D.A. Hensher and P.R. Stopher, (eds.), *Behavioural Travel Modelling.* Croom Helm, London.

Ben-Akiva, M.E. and Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT, Cambridge, Ma.

Ben-Akiva, M. and Morikawa T. (1990) Estimation of switching models from revealed preferences and stated intentions. *Transportation Research* **24A**, 485-495.

Ben-Akiva, M.E., Daly, A. and Gunn, H. (1987) Destination choice models: design and appraisal. *Planning and Transport Research and Computation* Proceedings of Seminar C held at the PTRC 15th Annual Meeting, London, 99-116.

Bewley, R., and Fiebig, D.G. (1988) A flexible logistic growth model with applications to telecommunications. *International Journal of Forecasting* **4**, 177-192.

Bhat, C.R. (1995) A heteroscedastic extreme-value model of intercity mode choice. *Transportation Research* **29B**, 471-483.

Bhat, C.R. (1998a) Accommodating flexible variations in responsiveness to level-of-service measures in travel mode choice modelling. *Transportation Research* **32A**, 495-507.

Bhat, C.R. (1998b) Accommodating flexible substitution patterns in multidimensional choice modelling: Formulation and application to travel mode and departure time choice. *Transportation Research* **32B**, 455-466.

Bhat, C., 1999. An analysis of evening commute stop-making behavior using repeated choice observations from a multi-day survey. *Transportation Research* 33B, 495–510.

Bhat, C.R. (2000). Flexible model structures for discrete choice analysis. In D.A. Hensher and K.J. Button (eds) *Handbook of Transport Modelling*, Elsevier Science, Amsterdam.

Bhat, C.R. and Castelar, S. (2002) A unified mixed logit framework for modelling revealed and stated preferences: Formulation and application to congestion pricing analysis in the San Francisco Bay Area. *Transportation Research* **36B**, 593-616.

Bierlaire, M., Lotan, T. and Toint, P. L. (1997). On the overspecification of multinomial and nested logit models due to alternative specific constants. *Transportation Science* **31**, 363-371.

Bilt, K.A. (1997) *Analysis of Urban Trip Generation Rates Using Segmentation Modelling (in Portuguese)*. M.S. Thesis. Escola Politécnica da Universidade de São Paulo, São Paulo.

Bradley, M.A., and Daly, A.J. (1997) Estimation of logit models using mixed stated preference and revealed preference information. In P. Stopher and M. Lee-Gosselin (eds) *Understanding Travel Behaviour in an Era of Change*. Pergamon Press, Oxford.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth, Belmont.

Brownstone, D., Bunch, D.S. and Train, K. (2000) "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles," *Transportation Research B* 34(5), 315-338.

Bruinsma, F.R. and Rietveld, P. (1998) The Accessibility of European Cities: Theoretical Framework and Comparison of Approaches. *Environment and Planning* **30A**, 499–521.

Brunton, P.J. and Richardson A.J. (1998) A cautionary note on zonal aggregation and accessibility. Paper presented at the 77th Annual Meeting of the Transportation Research Board, Washington, D.C.

Bruton, M.J. (1985) *Introduction to Transportation Planning*. Hutchinson, London.

Bruzelius, N. (1979) *The Value of Travel Time*. Croom Helm, London.

Buchanan, C. and Partners (1965) *Traffic in Guilford.* Penguin Books, Baltimore.

Bucklin, L.P. (1967) *Shopping Patterns in an Urban Area*. Institute of Business and Economic Research, Berkeley, CA.

Cairns, M.R. (1997) The development of park and ride in Scotland. *Journal of Transport Geography* **6**, 295-307.

Cambridge Systematics Europe (1981). *Zuidvleugel Study, Report 7: Travel Demand Models*. Rijkswaterstaat, Dienst Verkeerskunde.

Cameron, A.C. and Trivedi, P.K. (1986) Econometric models based on count data: comparisons and applications of some estimators and tests. *Applied Econometrics* **1**, 29-53.

Cameron, A.C. and Trivedi, P.K. (1990). *Regression Analysis of Count Data.* Cambridge University Press, Cambridge.

Carrasco, J.A. and Ortúzar, J. de D. (2002) Review and assessment of the nested logit model. *Transport Reviews* **22**, 197-218.

Caudill, M. (1987) Neural networks primer part 1. *AI Expert* **2**, 46-51.

Chamberlain, G. (1980) Analysis of covariance with qualitative data. *Review of Economic Studies* **47**, 225-238.

Chapin, F.S. (1974) *Human Activity in the City.* John Wiley & Sons, New York.

Cherchi, E., Meloni, I. and Ortúzar, J.de D. (2005) Policy forecasts using mixed RP/SP models: some new evidence. In A. Jaszkiewicz, M. Kaczmarek, J. Zak and M. Kubiak (eds) *Advanced OR and AI Methods in Transportation.* Publishing House of Poznan University of Technology.

Cherchi, E. and Ortúzar, J. de D. (2002) Mixed RP/SP models incorporating interaction effects: modelling new suburban train services in Cagliari. *Transportation* **29**, 371-395.

Cherchi, E. and Ortúzar, J. de D. (2006a) On fitting mode specific constants in the presence of new options in RP/SP models. *Transportation Research* **40A**, 1-18.

Cherchi, E. and Ortúzar, J. de D. (2006b) Income, time effects and direct preferences in a multimodal choice context: application of mixed RP/SP models with non-linear utilities. *Networks and Spatial Economics* **6**, 7-23.

Chicoine, J.E. and Boyle, D.K. (1984) Life-cycle concept: a practical application to transportation planning. *Transportation Research Record* **987**, 1-7.

City of Edinburgh Council (CEC) (2002). *Integrated Transport Initiative for Edinburgh and South East Scotland,* [online], Available from: http://www.tie.ltd.uk/pdf/suppl_report.pdf. [Accessed 15th March 2005].

City of Edinburgh Council (CEC) (2005) *Council Positive on Transport Future,* [online], Available from: http://www.edinburgh.gov.uk/transportedinburgh/Referendum/. [Accessed 15th March 2005].

Clark, S. (1996) National multi-modal travel forecasts: literature review of aggregate models. Working Paper 465, Institute of Transportation Studies, University of Leeds.

Cohen, C.A. (1961) Estimating the Poisson parameter from samples that are truncated to the right. *Technometrics* **3**, 433-438.

Cohn, N.D., Daly, A.J., Rohr, C., Dam, A.F., Oosterwijk, W.; van der Star, T. (1996). *ProMiSe: Policy-Sensitive Rail Passenger Forecasting for the Netherlands Railways.* PTRC European Transport Forum.

Copley, G. and Lowe, S.R. (1981) The temporal stability of trip rates: some findings and implications. *Proceedings 9th PTRC Summer Annual Meeting,* University of Warwick, July 1981, England.

Cotrus, A.V., Prashker, J.N. and Shiftan, Y. (2005) Spatial and temporal transferability of trip generation demand models in Israel. *Journal of Transportation and Statistics* **8**, 37-56.

Dalvi, M.Q. and Martin, K.M. (1976) The measurement of accessibility: some preliminary results. *Transportation* **5**, 17–42.

Daly, A.J. (1982) Estimating choice models containing attraction variables. *Transportation Research* **16B**, 5-15.

Daly, A.J. (1987) Estimating 'tree' logit models. *Transportation Research* **21B**, 251-268.

Daly, A.J. (1992) *ALOGIT 3.2 User's Guide.* Hague Consulting Group, The Hague.

Daly, A.J. (1997) Improved methods for trip generation. *Proceedings 25th European Transport Conference*, Vol. P415, pp 207-222, PTRC, London.

Daly, A.J. and Miller, S. (2006) Advances in modelling traffic generation. *European Transport Conference*, Strasbourg.

Daly, A.J. and Zachary, S. (1978) Improved multiple choice models. In D.A. Hensher and M.D. Dalvi (eds), *Determinants of Travel Choice*. Saxon House, Westmead.

Daly, A.J., van der Valk, J. and van Zwam, H.P.H. (1983) Application of disaggregate models for a regional transportation study in the Netherlands. In P. Baron and H. Nuppnau (eds), *Research for Transport Policies in a Changing World.* SNV Studiengesellschaft Nahverkehr, Hamburg.

Daor, E. (1981) The transferability of independent variables in trip generation models. *Proceedings 9th PTRC Summer Annual Meeting*, University of Warwick, England.

Davidson, K.B. (1977) Accessibility in transport/land-use modelling and assessment. *Environment and Planning* 9A, 1401-1416.

Department of Transport (1979) *National Travel Survey: 1975/1976 Report.* H.M. Stationary Office, London.

de Jong, G., Daly, A.J., Pieters, M. and van der Hoorn, A.I.J.M. (2005) The logsum as an evaluation measure: review of the literature and new results. *45th Congress of the European Regional Science Association*, Amsterdam, 23rd-27th August 2005.

Derek Halden Consultancy (2003) Accessibility: Review of Measuring Techniques and their Application. Scottish Executive Social Research: Edinburgh.

DKS Associates (1994). *Travel Model Development and Refinement - Trip Generation.* Puget Sound Regional Council. Available from: http://ntl.bts.gov/DOCS/dks.html. [Accessed 1st August 2006].

Dobson, R. (1976) The generalized linear model analysis of variance: its relevance to transportation planning and research. *Socioeconomic Planning Science* **10**, 231-235.

Dobson, A.J. (2001) *An Introduction to Generalized Linear Models,* Second Edition. Chapman Hall, London.

Domencich, T. and McFadden, D. (1975) *Urban Travel Demand: A Behavioural Analysis*, NorthHolland, Amsterdam.

Dong, X., Ben-Akiva, M.E., Bowman, J.L. and Walker, J.L. (2006) Moving from trip-based to activity-based measures of accessibility. *Transportation Research* **40A**, 163-180.

Doubleday, C. (1976) Spatial mobility and trip generation. *Urban Transport Planning Conference*, Leeds University.

Doubleday, C. 1977. Some Studies of the temporal stability of person trip generation models. *Transportation Research* **11**, 255–263.

Douglas, A.A. and Lewis, R.J. (1970) Trip generation techniques: (1) Introduction; (2) Zonal least squares regression analysis. *Traffic Engineering and Control* **12**, 362-5, 428-31.

Douglas, A.A. and Lewis, R.J. (1971) Trip generation techniques: (3) Household least squares regression analysis; (2) Category analysis and summary of trip generation techniques. *Traffic Engineering and Control* **12**, 477-9, 532-5.

Downes, J.D. and Gyenes, L. (1976) Temporal stability and forecasting ability of trip generation models in Reading. *TRRL Report LR 726*, Transport and Road Research Laboratory, Crowthorne.

Edinburgh City Centre Management (ECCM) (2004) *Proposed Congestion Charge scheme for Edinburgh: An assessment of its Potential Impact on the Retail Sector,* [online], Available from: http://www.edinburghcc.com/ECC/Retail_Impact_study.htm. [Accessed 25th January 2005].

Espino R., Román C., and Ortúzar J. de D. (2006) Analysing demand for suburban trips: a mixed RP/SP model with latent variables and interaction effects. *Transportation* **33**, 241-261.

European Commission (2004) *Urban Transport Pricing in Europe*, [online], Available from: http://www.transport-pricing.net. [Accessed 25th January 2005].

Faghri, A. and Hua, J. (1992) Evaluation of artificial neural network applications in transportation engineering. *Transportation Research Record* **1358**, 71-80.

Farag, S., Dijst, M. and Lanzendorf, M. (2003) Exploring the use of E-shopping and its impact on personal travel behaviour in the Netherlands. *Transportation Research Record* **1858**, 47-54.

Farrell, S.R. (2005) *Road-User Charging: Technically Optimal or Publicly Acceptable*? PhD Thesis, Napier University, Edinburgh.

Foerster, J.F. (1981) Nonlinear and non-compensatory perceptual functions of evaluations and choice. In P.R. Stopher, A.H. Meyburg and W. Brög (eds), *New Horizons in Travel Behaviour Research*. D.C. Heath and Co., Lexington, Mass.

Forinash, C. and Koppelman, F. (1993) Application and interpretation of nested logit models of intercity mode choice. *Transportation Research Record* **1413**, 98-106.

Fotheringham, A.S. (1982) A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning* **15A**, 15-36.

Freedman, O. and Kern, C.R (1997) A model of workplace and residence choice in two-worker households. *Regional Science and Urban Economics* **27**, 241-260.

Freeman, B.R. (1976) An accessibility approach to trip generation. *Proceedings of 8th ARRB Conference, Session 32*, 1-6, Perth.

Garrison, W.L. (1960) Connectivity of the interstate highway system. *Papers and Proceedings of the Regional Science Association* **6**, 121-137.

Garson, G. D. (2002) PA 765 *Statnotes: An Online Textbook.* [online], Available from: http://www2.chass.ncsu.edu/garson/pa765/statnote.htm. [Accessed 25th July 2002].

Garson, G. D. (2006) Multiple regression, from *Statnotes: Topics in Multivariate Analysis*. Available from: http://www2.chass.ncsu.edu/garson/pa765/statnote.htm [Accessed 25th November 2002]

Gaudry, M.J.I. and Wills, M.I. (1978) Estimating the functional form of travel demand models. *Transportation Research* **12**, 257-289.

Geertman, S.C.M. and J.R. Ritsema Van Eck. (1995) GIS and Models of Accessibility Potential: An Application in Planning. *International Journal of Geographical Information Systems* **9**, 67-80.

Geurs, K.T., Ritsema van Eck, J.R. (2001*) Accessibility measures: review and applications. Evaluation of accessibility impacts of land-use-transport scenarios, and related social and economic impacts. RIVM Report 408505006.* National Institute of Public Health and the Environment, Bilthoven.

Glass, G.V. and Stanley, J.C. (1986) *Métodos Estadísticos Aplicados a las Ciencias Sociales.* Prentice-Hall, Englewood Cliffs, NJ.

Golob, T.F. (2000) A simultaneous model of household activity participation and trip chain generation. *Transport Research* **34B**, 355-376.

Goodman, P.R. (1973) Trip generation: a review of the category analysis and regression models. Working Paper 9, Institute of Transport Studies, Leeds University.

Goulias, K.G., Pendyala, R.M. and Kitamura, R. (1990) A practical method for the estimation of trip generation and trip chaining. *Transportation Research Record* **1285**, 47-56.

Gourieroux, C., Monfort, A. and Trognon, A. (1984) Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica* **52**, 701-720.

Greater London Council. (1966) *London Traffic Survey.* Vol. II. London, London County Council.

Guevara, C.A. and Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record* **1977**, 60–66.

Guevara, C.A. and Thomas, A. (2007) Multiple classification analysis in trip production models. *Transport Policy* **14**, 514-522.

Guy, C.M. (1983) The assessment of access to local shopping opportunities: a comparison of accessibility measures. *Environment and Planning* **10B**, 219-238.

Guy, C.M. (1987) Recent advances in spatial interaction modelling: an application to the forecasting of shopping travel. *Environment and Planning* **19**, 173-186.

Guy, C.M. and Wrigley, N. (1987) Walking trips to shops in British cities. *Transportation Policy and Research* **58**, 63-79.

Hägerstrand. T. (1970) What about people in regional science? *Papers of the Regional Science Association* **24**, 1-21.

Hald, A. (1949) Maximum likelihood estimation of parameters of a normal distribution which is truncated at a known point. *Skandinavisk Aktuarietidskrift* **32**, 119–134.

Halden, D., McGuigan, D., Nisbet, A. and McKinnon, A. (2000) *Accessibility: Review of Measuring Techniques and Their application.* Scottish Executive Central Research Unit, Edinburgh.

Hall, M.D., Daly, A.J. Davies, R.F. and Russell, C.H. (1987) Modelling for an expanding city. *Proceedings 15th PTRC Summer Annual Meeting*, University of Warwick, England.

Handy, S.L. and Niemeier, D.A. (1997) Measuring accessibility: an exploration of issues and alternatives. *Environment and Planning* **29A**, 1175-94.

Hansen, W.G. (1959) How accessibility shapes land-use. *Journal of the American Institute of Planners* **25**, 73-76.

Hanson, S. and Schwab, M. (1986) Describing disaggregate flows: individual and household activity patterns. In S. Hanson (Ed.), *The Geography of Urban Transportation*, Guildford Press, New York.

Hanson, S. and Schwab, M. (1987) Accessibility and interurban travel. *Environment and Planning* **19A**, 735-748.

Hass-Klau, C. (1993) Impact of pedestrianization and traffic calming on retailing. *Transport Policy* **1**, 21-31.

HCG and TOI (1990). *A Model System to Predict Fuel Use and Emissions from Private Travel in Norway from 1985 to 2025*. Hague Consulting Group, The Netherlands.

Hedeker, D. and Gibbons, R.D. (1994) A random-effects ordinal regression-model for multilevel analysis. *Biometrics* **50**, 933-944.

Henderson, D., Koenig, B. and Mokhtarian, P. (1996) Using travel diary data to estimate the emissions impacts of transportation strategies: the Puget Sound Telecommuting Demonstration Project. *Journal of the Air Waste Management Association* **46**, 47-57.

Henderson, D.K. and Mokhtarian P.L. (1996) Impacts of center-based telecommuting on travel and emissions: analysis of the Puget Sound Demonstration projects. *Transportation Research* **1D**, 29-45.

Hensher, D.A. (1986) Sequential and full information maximum likelihood estimation of a nested logit model. *The Review of Economics and Statistics* **68**, 657-667.

Hensher, D.A. and Bradley M. (1993). Using stated response choice data to enrich revealed preference discrete choice data. *Marketing Letters* **4**, 139-151.

Hensher, D.A. and Greene, W. (2003) The mixed logit model: the state of practice. *Transportation* **30**, 133-176.

Hensher, D.A. and Greene, W. (2002) Specification and estimation of the nested logit model: alternative normalizations. *Transportation Research* **36B**, 1-17.

Hensher, D.H. and Johnson, L.W. (1981) *Applied Discrete Choice Modelling*. Croom Helm, London.

Hensher, D.A., Rose, J. and Greene, W.H. (2008) Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects. *Transport Geography* **16**, 126-133.

Hilbers, H.D. and Verroen, E.J. (1993) *Het beoordelen van de bereikbaarheid van lokaties. Definiëring, maatstaven, toepassing en beleidsimplicaties.* INRO-VVG 1993-09, TNO Inro, Delft. (Cited by Geurs and Ritsema, 2001).

Hobbs, F.D. (1979) *Traffic Planning and Engineering*. Pergamon Press, London.

Hope, S. (2002) *Scotland's People: Results from the 2001 Scottish Household Survey Volume 6: Technical Report.* A Scottish Executive National Statistics PublicationHope, S., Martin, C. and Dudleston, A. (2003) *Scotland's people: results from the 2001/2002 Scottish Household Survey. Volume 7: Annual Report.* Edinburgh: Scottish Executive. [online]. Available from: http://www.scotland.gov.uk/ Resource/Doc/47133/0024905.pdf. [Accessed 1st September 2003].

Hosmer, D.W. and Lemeshow, S. (2000) *Applied Logistic Regression.* Wiley, New York.

Hu, S. and Saleh, W. (2004) Incorporating impacts of accessibility on trip generation using logistic analysis. *Proceedings of the Fourth International Conference on Traffic and Transportation Studies (ICTTS)*, 296-307.

Hu, S. and Saleh, W. (2005) Impacts of congestion charging on shopping trips in Edinburgh. *Transport Policy* **12**, 443-450.

Hu, S. and Saleh, W. (2006) Modelling transport accessibility and its impacts on shopping trips. *Proceedings of the Fourth International Conference on Traffic and Transportation Studies (ICTTS)*, 196-206.

Huzayyin, A.S. (1978) *Importance of Accessibility in Trip Generation Modelling*. PhD Thesis, Department of Civil Engineering, Leeds University.

Iglesias, P., Godoy, F.J., Ivelic, A.M., and Ortúzar, J. de D. (2008) Un Modelo De Generación-distribución Y Partición Modal Conjunta Para Viajes Interurbanos. XV Panamerican Congress of Traffic and Transportation Engineering, Cartagena de Indias, September 2008, Colombia (In Spanish).

Ingram, D.R. (1971) The concept of accessibility: a search for an operational form. *Regional Studies* **5**, 101-107.

Jang T.Y. (2005) Count data models for trip generation. *Journal of Transportation Engineering* **131**, 444-450.

Jara-Díaz, S.R. and Farah, M. (1987) Transport demand and user's benefits with fixed income: the goods/leisure trade-off revisited. *Transportation Research* **21B**, 165-170.

Johnson, N.L. and Leone, F.C. (1964) *Statistics and Experimental Design: Volume II*. Wiley, New York.

Jones, P.M., Dix, M.C., Clarke, M.I. and Heggie, I.G. (1983) *Understanding Travel Behaviour*, Gower, Aldershot.

Kitamura, R. (1991) The effect of added transportation capacity on travel: a review of theoretical and empirical results. Paper for the *FHWA Conference on The Effects of Added Transportation Capacity*, Bethesda, MD, December 1991.

Karasmaa, N. and Pursula, M. (1997) Empirical studies of transferability of Helsinki metropolitan area travel forecasting models. *Transportation Research Record* **1607**, 38–44.

Kass, G.V. (1980) An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* **29**, 119-127

Kassoff, H. and Deutschman, H. (1969) Trip generation: a critical appraisal. *Highway Research Record* **297**, 15-30.

Keeble, D., Owens, P.L. and Thompson, C. (1982) Regional accessibility and economic potential in the European Community. *Regional Studies* **16**, 419-432.

Kelly, J.A., and Clinch, P.J. (2006) Influence of Varied Parking Tariffs on Parking Occupancy Levels by Trip Purpose. *Transport Policy* **13**, 487-495.

Kershaw, A., Barton, J., Noble, B., Williams, D. and Hird. D. (2001) *National Travel Survey Technical Report 2000*. Office for National Statistics, London.

Kitamura, R. (1991) The effect of added transportation capacity on travel: a review of theoretical and empirical results. Paper for the *FHWA Conference on The Effects of Added Transportation Capacity*, Bethesda, MD, December 1991.

Koenig, B., Henderson, D. and Mokhtarian P. (1996) The travel and emission impacts of telecommuting for the State of California Telecommuting pilot project. *Transportation Research* **4C**, 13-32.

König, D. (1936) *Theorie der Endlichen und Unendlichen Graphen.* Akademische Verlaggesellschaft M.B.H., Leipzig, Germany.

Koppelman, F. S. and Pas, E. I. (1984). Estimation of disaggregate regression models of person trip generation with multi-day data. In J. Volmuller and R. Hamerslag (eds), *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. VNU Science Press, Utrecht.

Koppelman, F.S. and Wen, C.H. (1998) Alternative nested logit models: structure, properties and estimation. *Transportation Research* **32A**, 289-298.

Koppelman, F.S. and Sethi, V. (2000) Closed-form discrete-choice models. In D.A. Hensher and K.J. Button (eds.), *Handbook of Transport Modelling*. Pergamon, Oxford.

Kouwenhoven, M., Rohr, C., Miller, S. and Daly, A.J. (2006) Evaluating a replacement ferry for the Isles of Scilly using a discrete choice model framework. *European Transport Conference*, Strasbourg.

Kraft, G. (1968) *Demand for intercity passenger travel in the Washington-Boston corridor, North-East Corridor*. Project Report, Systems Analysis and Research Corporation, Boston, Mass.

Kwan, M.P. (1998) Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis* **30**, 191–216.

Leake, G.R. and Huzayyin, A.S. (1979) Accessibility measures and their suitability for use in trip generation models. *Traffic Engineering and Control* **20**, 566-572.

Lenz, B. (2003) Will electronic commerce help to reduce traffic in agglomeration areas? *Transportation Research Record* **1858**, 39-46.

Lerman, S. (1979) The use of disaggregate choice models in semi-Markov process models of trip chaining behaviour. *Transportation Science* **13**, 273-289.

Lerman, S.R. and Louviere, J.J. (1978) The use of functional measurement to identify the form of utility functions in travel demand models. *Transportation Research Record* **673**, 78-85.

Levinson, D.M. (1998) Accessibility and the journey to work. *Journal of Transport Geography* **6**, 11–21.

Levinson, H.S. (1976) Urban travel characteristics. In J.E. Baerwald (ed.), *Transportation and Traffic Engineering Handbook*. Prentice-Hall, New Jersey.

Linneker, B.J. and Spence, N.A. (1991) An accessibility analysis of the impact of the M25 London orbital motorway on Britain. *Regional Studies* **26**, 31–47.

Litman, T. (2004) *Road Pricing – Congestion Pricing, Value Pricing, Toll Roads and HOT Lanes*, [online], Available from: http://www.vtpi.org/tdm/tdm35.htm. [Accessed 25th October 2004].

Local Government OR Unit. (1975) *A Preliminary Study of Royal Transport in West Yorkshire*. Royal Institute of Public Administration.

London Planning Advisory Committee (LPAC) (1994). *Advice on Strategic Planning Guidance for London.* LPAC, London.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables.* Sage Publications, Thousand Oaks, CA.

Louviere, J.J., Hensher, D.A., and Swait, J.D. (2000). *Stated Choice Methods: Analysis and Application.* Cambridge University Press, Cambridge.

Luce, R.D. and Suppes, P. (1965) Preference, utility and subjective probability. In R.D. Luce, R.R. Bush and E. Galanter (eds), *Handbook of Mathematical Psychology.* John Wiley & Sons, New York.

Manfield, N. (1969) Recreational Trip Generation in the Lake District. *J. Transport Economics and Policy* **3**, 152-164.

Manheim, C.F. (1973) Practical implications of some fundamental properties of travel demand models. *Highway Research Record* **244**, 21-38.

Manski, C.F. (1977) The structure of random utility models. *Theory and Decision* **8**, 229-254.

Marschak, J. (1960) Binary choice constraints on random utility indications, in K. Arrow (ed.), *Stanford symposium on mathematical methods in the social sciences*, Stanford University Press, Stanford, CA, 312-329.

Martin, C. and Hope, S. (2003) *Scotland's People: Results from the 2001/2002 Scottish Household Survey. Volume 8: Technical Report.* Edinburgh: Scottish Executive National Statistics. [online], Available from: http://www.scotland.gov.uk/Resource/Doc/47133/0024906.pdf. [Accessed 1st September 2003].

Martin, K.M. and Dalvi, M.Q. (1976) The comparison of accessibility by public and private transport. *Traffic Engineering and Control* **17**, 509-513.

McDermott N. (1995) *A comparative evaluation of selected statistical software for computing multinomial models. CDE working paper No. 95-1.* Centre for Demography and Ecology, University of Wisconsin-Madison.

McDonald, J.F. and Moffitt, R.A. (1980) The uses of Tobit analysis. *The Review of Economics and Statistics* **62**, 318–321.

McDonald, K.G. and Stopher, P.R. (1983) Some contrary indications for the use of household structure in trip generation analysis. *Transportation Research Record* **944**, 92-100.

McFadden, D. (1975) The measurement of urban travel demand. *Journal of Public Economics* **3**, 303–328.

McFadden, D. (1978) Modelling the choice model of residential location. *Transportation Research Record* **672**, 72-77.

McFadden, D. (1978) Modelling the choice of residential location. In A. Karlquist, L. Lundquist, F. Snickars and J.W. Weibull (ends), *Spatial interaction theory and residential location*. North Holland, Amsterdam.

McFadden, D. (1981) Econometric models of probabilistic choice. In C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data*. MIT Press, Cambridge, Massachusetts.

Meurs, H. (1990) Dynamic analysis of trip generation. *Transportation Research* **24A**, 427-442.

Misra, R., Bhat, C.R., Srinivasan, S., 2003. Continuous time representation and modeling framework for the analysis of nonworker activity-travel patterns: tour and episode attributes. *Transportation Research Record* **1813**, 11–20.

Mohammadian, A., and E. J. Miller (2003) Empirical Investigation of Household Vehicle Type Choice Decisions. *Transportation Research Record* **1854**, 99-106.

Mokhtarian, P., Handy, S. and Salomon, I. (1995) Methodological issues in the estimation of travel, energy, and air quality impacts of telecommuting. *Transportation Research* **29A**, 283-302.

Morikawa, T. (1989).*Incorporating Stated Preference Data in Travel Demand Analysis*. PhD Dissertation, Department of Civil Engineering, M.I.T.

Morris, J.M, Dumble, P.L. and Wigan M.R. (1979) Accessibility indicators for transport planning. *Transportation Research* **13A**, 91-109.

Muraco, W.A. (1972) Intra-urban accessibility. Econ. Geog. **48**, 388-405.

MVA Consultancy (2004) *Use Of CEC / Tie LUTI Model: Compendium of Results – ECCS/ITI Package*, in association with David Simmonds Consultancy, Prepared for the City of Edinburgh Council.

Nakkash, T.Z. (1969) *Activity-accessibility models of trip generation.* PhD Thesis, Purdue University, U.S.A

Nakkash, T.Z. and Grecco, W.L. (1972). Activity-accessibility models of trip generation. *Highway Research Record* **472**, 98-110.

Nelder, J.A. and Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series A* **135**, 370-384.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996) *Applied linear Statistical models*, Fourth Edition, Irwin, Chicago.

Niemeier, D.A. (1997) Accessibility: an evaluation using consumer welfare. *Transportation* **24**, 377-396.

Ortúzar, J. de D. (1982) Fundamentals of discrete multinomial choice modelling. *Transport Review* **2**, 47-78

Ortúzar, J. de D. (1983) Nested logit models for mixed-mode travel in urban corridors. *Transportation Research* **17A**, 283-299.

Ortúzar J. de D. (2001) On the development of the nested logit model. *Transportation Research* **35B**, 213-216.

Ortúzar, J. de D., Martínez, F.J. and Varela, F.J. (2000) Stated preferences in modelling accessibility. *International Planning Studies* **5**, 65-85.

Ortúzar, J. de D. and Willumsen, L.G. (2001) *Modelling Transport*, 3rd ed. John Wiley & Sons, Chichester.

Páez, A., Scott ,D.M., Potoglou, D., Kanaroglou, P.S. and Newbold, K.B. (2006) *Elderly mobility: demographic and spatial analysis of trip making in the Hamilton CMA, CSpA Working paper, CSpA_WP_014*, McMaster University, [online]. Available from: http://www.science.mcmaster.ca/cspa/papers/CSpA%20WP%20014.pdf [Accessed 1st February 2006]

Puget Sound Governmental Conference (PSGC) (1964) *Puget Sound Regional Transportation Study (PSRTS) Staff Report No. 16.*

Rich, D.C. (1978) *Potential Models in Human Geography. Concepts and Techniques in Modern Geography* 26. University of East Anglia, Norwich, England.

Rickard, J. M. (1988) Factors influencing long distance rail passenger trip rates in Great Britain. *Journal of Transport Economics and Policy* **22**, 209-233.

Rickard, J.M. (1989) Application of a generalized linear modelling approaches to category analysis. *Transportation Planning and Technology* **13**, 121-129.

Rizzi L.I. and Ortúzar J. de D. (2006) Road safety valuation under a stated choice framework. *Journal of Transport Economics and Policy* **40**, 69-94.

Rose, G. and Koppelman, F.S. (1984) Transferability of disaggregate trip generation models. *Proceedings of 9th International Symposium on Transportation and Traffic Theory*. VNU Science Press, Utrecht, Netherlands, 47-56.

Ryan, T.P. (1997) *Modern Regression Methods*. John Wiley & Sons, Inc., New York.

Said, G.M. and Young, K.G. (1990) A generalized linear model framework for estimating work trip rates for households in Kuwait. *Transportation Research* **24A**, 187-200.

Said, G.M., Young, D.H. and Ibrahim, H.K. (1991) Trip generation procedure for areas with structurally different socioeconomic groups. *Transportation Research Record* **1328**, 1-9.

Said, G.M., Young, D.H. and Ibrahim, H.K. (1994) Statistical models for work trip rates of Kuwaiti households. *Journal of University of Kuwaiti Science* **21**, 135-150.

Saleh, W. and Farrell, S. (2005) Implications of congestion charging for departure time choice: work and non-work schedule flexibility. *Transportation Research* 39A, 773-791.

Saunders, J. (2004) *The City of Edinburgh Council Transport Initiatives Edinburgh Proposed Congestion Charge Order: John Saunder's Precognition for Public Inquiry*, [online], Available from: http://iti.tiedinburgh.co.uk/ [Accessed September 4th 2004]

Schmöcker, J.D., Fonzone, A., Quddus, M.A. and Bell, M.G.H., (2006) Changes in the frequency of shopping trips in response to a congestion charge. *Transport policy* **13**, 217-228.

Schmöcker, J.D., Quddus, M.A., Noland, R.B. and Bell, M.G.H., (2005) Estimating trip generation of elderly and disabled people: an analysis of London data. *84th Transportation Research Board Annual Meeting*, Washington, D.C.

Schuldiner, P.W. (1962) *Trip generation and the home*. Highway Research Board, Bulletin No. 347.

SECTRA (1998) Metodología de Análisis de Sistemas de Transporte Urbano. (http://www.sectra.cl/contenido/metodologia/transporte_urbano/Analisis_sistema_transporte_urbano.htm). Developed by Fernández y De Cea Consultants for SECTRA, MIDEPLAN, Chile.

Sermons, M.W. and Koppelman, F.S. (2001) Representing the differences between female and male commute behaviour in residential location choice models. *Journal of Transport Geography* **9**, 101-110.

Sharpe, G.B., Hansen, W. G. and Hamner, L.B. (1958) Factors affecting trip generation of residential land-use areas. *Highway Research Board Bulletin* **203**, 20-36.

Sheffi, Y. (1979) Estimating choice probabilities among nested alternatives. *Transportation Research* **13B**, 189-205.

Sheppard, E. (1986) Modelling and predicting aggregate flows. In S. Hanson, (ed), *The Geography of Urban Transportation.* The Guilford Press, New York, 91-118.

Sherman, L, Barber, B. and Kondo, W. (1974) Method of evaluating metropolitan accessibility. *Transportation Research Record* **499**, 70-82.

Shiftan, Y. (1999) A practical approach to model trip chaining. *Transportation Research Record* **1645**, 17-23.

Shimbel, A. (1953) Structural parameters of communications networks. *Bulletin of Mathematical Biophysics* **15**, 501-507.

Singer, E.H.E. (1973) Correcting a basic inconsistency in trip generation, *Traffic Engineering & Control* December.

Slovic, P., Fischoff, B. and Lichtenstein, S. (1977) Behavioural decision theory. *Annual Review of Psychology* **28**, 1-39.

Small, K.A. (1982) The scheduling of consumer activities: work trips. *American Economic Review* **72**, 467-479.

Small, K.A. and Rosen, H.S. (1981) Applied welfare economics with discrete choice models. *Econometrica*, **49**, 105-130.

Smith, R.L. and Cleveland, D.E. (1976) Time stability analysis of trip generation and predistribution modal choice models. *Transportation Research Record* **569**, 76-86.

Song, S. (1996) Some tests of alternative accessibility measures: a population density approach. *Land Economics* **72**, 474-482.

Still, B. and Simmonds, D. (2000) Parking restraint policy and urban vitality. *Transport reviews* **20**, 291-316.

Stopher, P.R (1975) Goodness-of-fit measures for probabilistic travel demand models. *Transportation* **4**, 67-83.

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

Strambi, O. and Bilt, K. (1998) Trip generation modelling using CHAID, a criterion-based segmentation modelling tool. *Transportation Research Record* **1654**, 24-31.

Supernak, J. (1979) A behavioural approach to trip generation modelling. *Proceedings of 7th PTRC Summer Annual Meeting*, University of Warwick, England.

Supernak, J. (1981) Transferability of the person category trip generation model. *Proceedings of 9th PTRC Summer Annual Meeting*, University of Warwick, England.

Supernak, J. (1987) A Method for Estimating Long-Term Changes in Time-of-Day Travel Demand. *Transportation Research Record* **1138**, 18-26.

Supernak, J., Talvitie, A.P. and DeJohn, A. (1983) Person category trip generation modelling. *Transportation Research Record* **944**, 74-83.

Svenson, O. (1979) Process descriptions of decision making. *Organizational Behaviour and Human performance* **23**, 86-112.

Swait, J. and Adamowicz, W. (1996) *The effect of choice environment and task demands on consumer behaviour: Discriminating between contribution and confusion.* Department of Rural Economy, University of Alberta, Working Paper.

Swait, J. and Louviere, J. (1993) The role of the scale parameter in the estimation and use of multinomial logit models. *Journal of Marketing Research* **30**, 305-314.

Takyi, I.K. (1979) *Household Trip Generation in Kumasi.* B.Sc. thesis. University of Science and Technology, Kumasi, Ghana.

Takyi, I.K. (1990) Trip generation analysis in a developing country context. *Transportation Research Record* **1285**, 9-21.

Taylor, M.A. (1968) *Studies of travel in Gloucester, Northampton and Reading. Report No. LR 141*. Road Research Laboratory, Crowthorne, England.

Thurstone, L. (1927) A law of comparative judgment. *Psychological Review* **34**, 273-286.

Tillema, F., Zuilekom, K.M. and Maarseveen, van M.F.A.M. (2004) Trip generation and trip distribution: Comparison of neural networks and traditional methods. *10th World Conference on Transport Research (WCTR)*, Istanbul, Turkey.

Tobin, J. (1958) Estimation for relationships with limited dependent variables. *Econometrica* **26** (2), 24–36.

Train, K. (2003) *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

Train, K., Ben-Akiva, M. and Atherton, T. (1989) Consumption patterns and self-selecting tariffs. *Review of Economics and Statistics* **71**, 62-73.

Train, K., McFadden, D. and Ben-Akiva, M. (1987) The demand for local telephone service: A fully discrete choice model of residential calling patterns and service choice. *Rand Journal of Economics* **18**, 109-123.

Train, K.E. and McFadden, D. (1978) The goods/leisure trade-off and disaggregate work trip mode choice models. *Transportation Research* **12**, 349-353.

U.S. Department of Transportation Federal Highway Administration Urban Planning Division. (1975) *Trip Generation Analysis - August 1975*.

Vickerman, R.W. (1974) Accessibility, attraction, and potential: a review of some concepts and their use in determining mobility. *Environment and Planning* **6A**, 675-691.

Vickerman, R.W. and Barmby, T.A. (1984) The structure of shopping travel: some developments of the trip generation models. *Journal of Transport Economics and Policy* **18**, 109-121.

Vickerman, R.W. and Barmby, T.A. (1985) Household trip generation choice: Alternative empirical approaches. *Transportation Research* **19B**, 471-479.

Vlek, C. and Steg, L. (1996) Societal reasons, conditions and policy strategies for reducing the use of motor vehicles. *Towards sustainable transportation*, Vancouver, Canada, 24-27.

Voges, E.M. and Naude, A.H. (1983) *Accessibility in urban areas: an overview of different indicators*. National Institute for Transport and Road Research, Pretoria.

Wachs, M. and Kumagai, T.G. (1973) Physical accessibility as a social indicator. *Socio-Economic Planning Science* **7**, 437-456.

Wallace, B., Barnes, J., Rutherford, G.S., 2000. Evaluating the effects of traveler an trip characteristics on trip chaining, with implications for transportation demand management strategies. *Transportation Research Record* **1718**, 97–106.

Wang, D. (1996) *Accessibility: activity-based measures of accessibility for transportation policy analysis*. 24th European Transport Conference (PTRC), Proceedings of Seminar D&E, part 2.

Wang, R.M. (1997) *An activity-based trip generation model*. PhD. dissertation. Department of Civil and Environmental Engineering, University of California, Irvine. Irvine, CA

Washington, S. (2000) Iteratively specified tree-based regression: theory and trip generation example. *Journal of Transportation Engineering* **126**, 482-491.

Washington, S. and Wolf, J. (1997) Hierarchical tree-based versus ordinary least squares linear regression models: theory and example applied to trip generation. Paper presented at *the 76th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2003) *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman and Hall/CRC: Boca Raton.

Wen, D. and Koppelman, F. (2001) The generalized nested logit model. *Transportation* Research **35B**, 627-641.

Wickstrom, G.B. (1971) Defining balanced transportation: a question of opportunity. *Traffic Quarterly* **25**, 337-350.

Wiggin, P. (1993) Streets, Traffic and Trade. *Research Report 6*. Environment City Trust, Leicester.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* **9A**, 285-344.

Wilmot, C.G. (1995) Evidence of transferability of trip generation models. *Journal of Transportation Engineering* **9**, 405–410.

Wilson, A.G. (1967) A statistical theory of spatial distribution models. *Transportation Research* **1**, 253-269.

Wilson, A.G. (1971) A family of spatial interaction models, and associated developments. *Environment and Planning* **3A**, 1-32.

Wilson, A.G. (1974) *Urban and Regional Models in Geography and Planning*. Wiley, Chichester.

Wilson, A.G., Hawkins, A.F., Hill, G.J. and Wagon, D.J. (1969) Calibration and testing of the SELNEC transport model. *Regional Studies* **3**, 337-350.

Wootton, H.J. and Pick, G.W. (1967) A model for trips generated by households. *Journal of Transport Economics and Policy* **1**, 137-153.

Yai T., Iwakura, S. and Morichi, S. (1997) Multinomial probit with structured covariance for route choice behaviour. *Transportation Research* **31B**, 195-207.

PUBLISHED RESEARCH

HU, S. and SALEH, W. (2006) Modelling transport accessibility and its impacts on shopping trips. *Proceedings of the Fourth International Conference on Traffic and Transportation Studies (ICTTS)*, pp196-206. ISBN: 7030174445.

HU, S. and SALEH, W. (2005) Impacts of congestion charging on shopping trips in Edinburgh. *Transport Policy* 12 (5), pp 443-450.

HU, S. and SALEH, W. (2004) Incorporating impacts of accessibility on trip generation using logistic analysis. *Proceedings of the Fourth International Conference on Traffic and Transportation Studies (ICTTS)*, pp 296-307. ISBN: 7030137809.

APPENDIX 1 THE QUESTIONNAIRE FOR THE SHOPPING TRIPS IN EDINBURGH

YOUR SHOPPING TRIPS

I am a research student at School of the Built Environment, Napier University. As part of my PhD study, I am carrying out a survey to investigate the characteristics of shopping trips made by individuals and the potential impacts of transport policies on individual's travel behaviour. Therefore, I need your help to collect information about your travel patterns and your attitudes towards current transportation provision and possible transport policies.

This questionnaire includes four sections. In section one you will be asked to fill out all the shopping trips you made over the course of the past week. In section two you will be asked about your opinions about transport systems in Edinburgh and possible transport policies. The potential impacts of such transport policies on your shopping trips will be addressed in section three. Finally, the last section will require you to provide some information about yourself and your family.

Section One: Your Shopping Trips

In this section, I'd like to ask some questions about your shopping trips. Here, your journey to shopping is treated as one trip, and your journey from shopping to home is treated as another trip.

Q1. Please indicate the number of shopping trips you usually make <u>every week</u> and the mode of travel you use for these trips?

	Weekdays	Weekends
To the city centre		
Somewhere else		

☐ Walking ☐ Car/van ☐ Bus/taxi ☐ Cycling ☐Other ________

Q2. If you make some of these trips by bus, how long does it (usually) take you to travel to the city centre? And
to travel to somewhere else?

	To the city centre		To somewhere else	
	Weekdays	Weekends	Weekdays	Weekends
Walking time to bus stop				
Waiting time at bus stop				
Bus travel time				

Q3. If you make some of these trips by car, how much do you pay for parking fees for such shopping trips in the city centre? (If you pay different charges depending on where you park, please supply information on the different type of parking facility/type, e.g. multi-story, on street etc.)

	Weekdays	Weekends
Surface		
Multi-story		
On-street		

Q4. If you make some of these trips by car, how long does it typically take you to search or wait for a parking space when you go shopping? ____________ Minutes

Q5. If you usually travel to shopping destinations other than the city centre, please give the reasons for not going to the city centre? (*You can choose more than one*).

☐ The buses are not convenient

☐ It is difficult to find a parking place

☐ I do not want to pay for parking my vehicle

☐ City centre is far from home

☐ Shopping is more expensive in the city centre

☐ There are too many people in the city centre

☐ Choice of goods is limited in the city centre

☐ Others, please specify ____________________

Section Two: Your Attitudes About Transport and Transport Policies

In this section, I would like to ask your opinions about the transport system/network in Edinburgh.

Q6. Please indicate how strongly you agree/disagree with each of the following statements (Please tick one box on each line).

	Strongly agree	Agree	Slightly agree	Neither agree/disagree	Slightly disagree	Disagree	Strongly disagree	Not applicable	Don't know
1. There is a serious congestion problem in Edinburgh.									
2. The air pollution caused by transport in Edinburgh is a problem.									
3. I think noise caused by vehicles is a problem.									
4. I feel safe to cycle in the city centre of Edinburgh.									
5. I am satisfied with the bus services in Edinburgh.									
6. It is easy to find a parking place in the city centre.									
7. Parking costs are reasonable in the centre area of Edinburgh									
8. Congestion charging would be very effective in tackling congestion.									
9. Increased parking fees would be very effective in tackling congestion.									
10. Reducing parking places in Edinburgh would help reduce congestion in the city.									

Section Three: The Influence of Transport Policies on Your Shopping Trips

In this section, you are questioned on how different transport policies may influence you making shopping trips.

Q7 If the following transport policies would be considered to improve traffic and reduce congestion in Edinburgh, could you rate each of them for their effectiveness (with 10 being the most effective and 1 the least effective)?

Transport policies		Rating
1. To apply congestion charging in the city centre	£2 per day	
	£3 per day	
	£5 per day	
2. To increase parking fees in the city centre (car parking fees <u>increase</u> per hour)	£0.50	
	£1.00	
	£1.50	
3. To reduce parking places in the city centre (New searching/waiting time for a parking place)	10 minutes	
	20 minutes	
	30 minutes	

A. CONGESTION CHARGING SCENARIOS

<u>SEENARIO ONE</u> In this scenario it is assumed that the charging area refers to the inner cordon, and that there is one charging rate which is applied all day during weekdays, but not applied during weekends.

Q8. How many shopping trips would you make per week if such a congestion charging were introduced?

Congestion charge per day	Number of shopping trips per month	
	Weekdays	Weekends
£2		
£3		
£5		

Q9. If these congestion-charging programs were introduced, would you change from car to other modes when travelling for shopping?

Congestion charge per day	Mode change from car				
	No change	Walk	Bus	Cycling	Other
£2					
£3					
£5					

SCENARIO TWO In this scenario it is assumed that the charging area refers to the inner cordon, and that there is one charging rate which is applied throughout the week (seven days a week).

Q10. How many shopping trips would you make if such a congestion charging were introduced?

Congestion charge per day	Number of shopping trips per week	
	Weekdays	Weekends
£2		
£3		
£5		

B. CAR PARKING COST SCENARIOS

This section is to assist in an investigation into the effect of parking charging policies on travel behaviour.

Q11. If the cost of car-parking in the city centre is increased by the following VALUES, how many shopping trips would you make per week?

Car parking fees increase per hour	Number of shopping trips per week	
	Weekdays	Weekends
£0.50		
£1.00		
£1.50		

C. PARKING MANAGEMENT SCENARIOS

The objective of limiting the number of parking spaces in the city centre is to discourage car users from driving to the city centre. This reduction in parking spaces would result in an increase in the time spent searching for a parking space or waiting for a parking space.

Q12. If it took longer for you to search/wait for a parking place, how many shopping trips would you make per month?

New searching/ waiting time for a parking place	Number of shopping trips per week	
	Weekdays	Weekends
10 minutes		
20 minutes		
30 minutes		

D. PUBLIC TRANSPORT SCENARIOS

Q13. In order to attract more people to use public transport, the government needs to improve public transport services. Which of the following measures do you think would be most effective (please rate in order with 1 being the most effective)?

Public transport targets	(Please rate)
To reduce walking time to bus stop (more bus stops)	
To reduce waiting time at bus stops (more frequent buses)	
To reduce the bus fare (cheaper buses)	
To reduce bus travel time by bus priority measure (quicker buses)	
Trams	
More train stations with frequent train services	
Other ______	

Q14. If the following public transport improvement were obtained, how many shopping trips would you make per week?

Time reduction of total travel time to the City Centre by bus	Number of shopping trips per week	
	Weekdays	Weekends
15%		
30%		
50%		

Section Four: Yourself and Your Family

This section seeks some information about yourself and your family.

Q15. Are you male or female? ☐ Male ☐ Female

Q16. Which of the following age groups are you in?

☐ Under 16 ☐ 16-24 ☐ 25-35 ☐36-45

☐46-55 ☐ 56-65 ☐ 66+

Q17. Do you hold a full driving licence? ☐ Yes ☐ No

Q18. How many dependent children normally live in your household?

☐ None Under 5 ____ 5-12 ____ Over 12 ____

Q19. Which of the following best describes your current situation?

☐ Full time employed/Self-employed ☐ Part time employed

☐ Look after home/family ☐ Permanently retired

☐ Unemployed and seeking work ☐ Higher/further education

☐ Permanently sick or disabled

Q20. What is the first half of your home postcode (e.g. EH10) E H ___ ___

Q21. How many people normally live in your household?

☐ 1 ☐ 2 ☐ 3 ☐4 ☐5+

Q22. How many people in your household are in employment (full-time or part-time)?

☐ None ☐ 1 ☐ 2 ☐ 3+

Q23. How many cars are available for your household?

☐ None ☐ 1 ☐ 2 ☐ 3+

Q24. How many licensed drivers are there in your household?

☐ None ☐ 1 ☐ 2 ☐ 3+

Q25. What is your gross personal and household income (Please tick one for each)?

Per Week	Per Month	Per year	Personal	Household
Under £99	Under £419	Under £5,199		
£100-£199	£420-£859	£5,200-£10,399		
£200-£299	£860-£1,299	£10,400-£15,599		
£300-£399	£1,300-£1,733	£15,600-£20,799		
£400-£599	£1,734-£2,602	£20,800-£31,199		
£600-£769	£2,603-£3,332	£31,200-£39,999		
£770-£961	£3,333-£4,166	£40,000-£49,999		
£962-£1,153	£4,167-£4,999	£50,000-£59,999		
£1,154 or more	£5,000 or more	£60,000 or more		

Thank you for taking the time to complete this questionnaire. Please return it to me in the freepost envelope provided (no need for a stamp).

If you have any comments about the issues raised in this questionnaire, please provide them in the space below.