Research into Changing Trip Rates over Time and Implications for the National Trip End Model: Final Report

Department for Transport

May 2009

QM

Issue/revision	Issue 1	Revision 1	Revision 2	Revision 3
Remarks	Draft 1	Draft 2		
Date	13 May 2009	18 August 2009		
Prepared by	Kaveh Jahanshahi, Ian Williams, Xu Hao	Kaveh Jahanshahi, Ian Williams		
Signature				
Checked by	John Bates			
Signature				
Authorised by	Ian Williams	Ian Williams		
Signature				
Project number	11642024	11642024		
File reference	DraftFinalReport.doc	DraftFinalReport.doc		

WSP Development and Transportation
66-68 Hills Road
Cambridge
CB2 1LA

Tel: +44 (0)1223 558050
Fax: +44 (0)1223 558051
http://www.wspgroup.com

Reg. No: 2382309

Contents

List of Figures

List of Tables

1 Introduction

1.1 AIMS

1.1.1 The aims of this study are:

- to evaluate the evidence available to understand and explain the apparent overall decrease in trip generation rates per capita that has been observed within NTS data over the last decade ;
- to examine whether the current variables in use within the UK Department for Transport's National Trip End Model (NTEM) are the best determinants of trip rates and to provide advice on the form and inclusion of any new variables through which NTEM could be improved.

1.1.2 This understanding will be used to assess the most appropriate adjustments to the methods used to estimate base trip production rates within the National Trip End Model (NTEM) and to provide guidance on any revisions that should be made in the approach used to forecast these trip production rates through future years for use as inputs to TEMPRO.

1.1.3 The WSP Policy and Research team from Cambridge that has carried out this study is the same group which in 1999 (under their then name ME&P) developed the original multi-modal trip rates now being used in NTEM. These rates were estimated through analysis of travel behaviour using the NTS database for the years up to 1996. This previous work is a key input to this study because it identified many of the main influences on trip production rates. It also provided a comprehensive understanding of how the NTS database had evolved over time up to then and of the potential pitfalls that can emerge in its use, unless care is taken to understand its data definitions and how these may have changed over time.

1.1.4 The current study builds on this past work but differs in two main respects:

- The NTS database that is used to estimate the trip rates now includes newer travel data up to the year 2006;
- The trip production rates are being estimated using more sophisticated statistical estimation techniques (e.g. negative binomial regression) than those of the 1999 study - this methodological improvement in the analysis techniques substantially improves both the precision of the estimates of the trip rates themselves and the precision with which the main factors that influence these trip rates can be determined.

1.2 STRUCTURE OF THE REPORT

1.2.1 Chapter 2 outlines the overall objectives of this study, while Chapter 3 provides an overview of the methodology that is being adopted for the study. Chapter 4 presents a literature review on trip rates which was carried out by Rand Europe. Chapter 5 presents the mathematical specification of the approach which is employed in analysing trip rates. Chapter 6 reviews the NTS database; the changes through time in the coverage and design of this survey; together with the impact of these changes on the inferences that can be drawn from this database about changes in trip rates. Chapter 7 explains how the current modelling approach relates to that previously used for the original NTEM, while Chapter 8 presents the findings from examining changes in trip rates based on analysing 1995 to 2006 NTS data. Chapters 9 to 13 provide results from analysing various variables and travel purposes, followed by Chapter 14 which suggests some interesting topics and areas for future research and makes recommendations for changes to TEMPRO. Finally, Chapter 15 summarises the main research findings.

2 Objectives of the Study

2.1 OVERVIEW

2.1.1 This research study attempts to achieve four linked objectives:

- Validate whether in reality trip rates are tending to fall;
- Analyse the strength of the link between trip rates and generalised cost and/or accessibility;
- Investigate how to improve trip end growth estimates by rail;
- Identify changes required within TEMPRO.

These objectives are discussed in more detail in turn below.

2.2 VALIDATE WHETHER IN REALITY TRIP RATES ARE TENDING TO FALL

2.2.1 The key issue is to ascertain firstly, whether this apparent trend could be an aberration due primarily to data issues that arise from differences over time in the methods, response rates, expansion factors or other features of the NTS survey. Then if there is shown to be unambiguous evidence of a genuine behavioural trend through time of reductions in trip rates, the requirement is to measure the speed of this reduction, the types of trips and of persons to which it most pertains and the likelihood that it would accelerate or reduce through into future years. The potential for travel time budget trends to explain some of the changes in trip making behaviour will also be considered.

2.2.2 There is a certain irony about this first task, in that the DfT commissioned Validation study of WSP's original version of the National Transport Model (NTM) Pass1 Demand model had criticised it for failing to take due account of an apparent tendency for trip rates to *increase* over the years, whereas this current study is to examine whether these rates are *falling* through time! This apparent contradiction in the direction of trip rate trends is an important indicator of the need to examine with great care the internal consistency over time and the sampling errors and other potential sources of error within the NTS dataset that is being used to examine trends in travel behaviour.

2.2.3 It is certainly not the case that we think that the NTS is a substandard survey. On the contrary our experience suggests that it is of high quality and that it is likely to be more consistent through time than other periodic surveys. It is simply that even the best surveys need to be scrutinised carefully when using them to examine the existence of subtle changes in behaviour.

2.3 ANALYSE THE LINK BETWEEN TRIP RATES AND GENERALISED COST / ACCESSIBILITY

2.3.1 The requirement was to determine whether transport cost changes had a significant empirical impact on trip rates. This potential impact could be examined in two dimensions: space and time.

2.3.2 The spatial dimension is studied in some detail, both through examining the impacts of the various accessibility variables available within the NTS and through examining how trip rates vary across area types (e.g. conurbation, urban, rural) that have inherently different levels of accessibility.

2.3.3 The temporal cost dimension has dropped out, because as will be seen later, no significant trend in trip rates over time has been demonstrated so that the impacts on trip rates of cost changes through time are likely to be unimportant.

2.4 INVESTIGATE HOW TO IMPROVE TRIP END GROWTH ESTIMATES BY RAIL

2.4.1 Particular interest focuses on the rail mode, because over the last decade rail trip rates have proved to have been difficult to forecast realistically. The rapid growth rate in trips by rail over the last decade has not been captured in the NTEM procedures that underlie the forecasts in TEMPRO.

2.4.2 The results from the analysis of trip rates on rail are discussed in Chapter 12, which explains why the NTEM has not picked up the rapid growth in rail trips and then presents the substantial extensions to the model that would be required in order to really improve rail trip end forecasts.

2.5 IDENTIFY CHANGES REQUIRED WITHIN TEMPRO

2.5.1 Any trends in trip rates and any proven linkage to spatial or temporal cost changes would need to be considered for inclusion within the TEMPRO system. Following on from the findings on the degree of temporal stability in the trip rates a decision would be made on whether a simple one-off rebasing of the trip rates to 2001 or a later year is sufficient; or whether a more sophisticated trend mechanism or cost sensitivity would need to be built into the TEMPRO system to enable these effects to be projected forward to future years.

2.5.2 It is now clear that this later sophisticated set of methodological developments is not required, since the main influences on trip rates per person appear to be relatively stable over time, and that the geographical differences can be represented well through using area types.

2.5.3 Other behavioural findings would also need to be considered for inclusion within TEMPRO. These include the introduction of segmentation of trip rates by income or socio-economic group (SEG) in order to aid in the development of models that provide the degree of segmentation that is required by DfT for the assessment of road pricing (TIF) and of other forms of cost based policy measures.

3 Overview of the Approach for this Study

3.1.1 The work in this study was carried out as a series of tasks each of which is summarised below in turn, explaining the approach that has been adopted. Later Chapters then present in greater detail the outcomes from these tasks.

3.2 REVIEW EXISTING LITERATURE ON TRIP RATE CHANGES

3.2.1 Prior to embarking on the statistical analysis a literature review was carried out by Rand Europe (Chapter 4) which focused on two main topics.

3.2.2 The first aim was to see whether there are other relevant studies that have found sound empirical evidence of changes in trip rates through time. This focused particularly on UK studies but it did not find any major recent empirical studies that shed new light on this topic. This absence appears mainly to be caused by the need for a large survey sample that has been collected in a fully consistent fashion over a reasonably long time horizon. Few existing UK databases can meet these onerous requirements, other than the NTS and perhaps the periodic travel surveys in metropolitan areas such as London and Manchester.

3.2.3 The other aim of the literature review was to examine recent methodological developments in the modelling and estimation of trip generation rates. This part of the review found a number of interesting developments (e.g. Larsen, 2003; Daly & Miller, 2006) that are relevant to and were used in the current study, while others could be of potential use in follow-on work subsequent to this study. In order to take appropriate account of their non-negative integer characteristics when estimating models of counts of trips/tours, the review focused particularly on the need to use appropriate estimation procedures, based on Poisson, negative binomial regression or logit choice models, rather than on traditional liner regression. As will be described in Chapter 5, the negative binomial regression is the estimation approach adopted here to uncover and analyse the main influences on trip production rates through time, space and by person type.

3.3 ASSEMBLE THE NTS DATABASE

3.3.1 The ideal source material for this study would be to acquire the NTS data for a long time series, perhaps from 1988 to the most recent year available 2006. The larger variations in the survey procedures for years prior to 1988 make results prior to 1988 less useful for purposes of comparisons over time. As explained in Section 6.2, WSP have obtained the NTS database in a consistent form, suitable for the full analysis of various factors including income, socio-economic groups and accessibility, but only for the years 2002 to 2006. The conduct of the survey was changed somewhat in 2002 when a different survey organisation was introduced. At a later stage in the study we were provided by the NTS team with a further dataset that covers the period from 1995 to 2001 but this additional dataset was not in a sufficiently consistent form to be used rigorously for detailed analysis of the complete set of socio-economic factors, accessibility and income variables of interest. In the phase of the study that examined the scale of changes in trip rates through time, we therefore decided to use only those variables that are consistent in definitions over the whole 1995-2006 dataset in order to evaluate the scale of changes in trip rates for this longer period. The results of this analysis of temporal effects are provided in Chapter 8. In parallel using only the 2002 to 2006 dataset, more detailed analyses were carried out of the effects on trip rates of income, socioeconomic factors and accessibility variables.

3.4 EXAMINE SOURCES OF POTENTIAL BIAS IN NTS RESULTS THROUGH TIME

3.4.1 We have examined the substantial volume of methodological documentation available on the NTS survey. This includes the past work by the NTS team on potential sources of problems in the survey and on the benefits from applying weights to try to offset some of the effects of the reduction over time in the survey response rate that has been achieved.

3.4.2 The NTS methodological documentation explains that the national response rate has declined significantly over the years from a healthy 79% in 1991 to a much lower 61% in 2006. In areas such as Inner London the response rate of 49% in 2006 is much lower again. The implications of these lower response rates have needed careful consideration and analysis, to measure the extent to which they are a source of bias among those who have completed the survey towards persons who spend more time at home and less time travelling. This topic is discussed in greater detail in Chapter 6.

3.5 ANALYSE INFLUENCES ON TRIP RATES

3.5.1 This is the main task of the study and the results are presented in Chapters 9 to 13. Because of the large number of segments that need to be investigated, the empirical analysis and estimation of trip production rates was carried out using Negative Binomial regression techniques, which is a more general form of Poisson regression. For each travel purpose this identified its trend through time as well as the other potential influences that can lead to variations in trip rates through space or between types of individuals.

3.5.2 Negative Binomial (NB) regression is the equivalent approach for count data (trip rates), to the use of regression methods for continuous variables. It avoids the problems with small samples that were encountered in the original 1999 study when using cross-categorisation based methods to identify the appropriate segmentation of person types for trip rates. NB regression has the added advantage of providing informative measures of the statistical significance of these various influences on trip rates, provided that these variables are not heavily collinear.

3.5.3 This statistical estimation of the trip rate model was carried out using the R statistical package. This contains flexible routines for NB regression as well as for other related techniques for count estimation, such as Poisson regression. R provides robust procedures with clear diagnostics that can be used, for example:

- to test for significant interactions between categorical variables;
- and to guide in determining the best supported parsimonious model structure for each travel purpose.

3.5.4 We also test for overdispersion in the variance, which would imply that the negative binomial model should be used in place of the Poisson model, through implementing in R a test due to Cameron and Trivedi (1990).

3.5.5 It is comparatively straightforward to estimate the changes through time in trip production rates, provided that the NTS dataset is consistent throughout the period. However, there are a number of reasons why the estimation of the direct effects of cost and other transport supply changes on trip rates is more challenging. The data on travel supply costs in the NTS has not been validated, so that its reliability is not known. The NTS by its nature only provides supply information for the mode chosen for a trip, not on the modes not chosen. These two shortcomings are particularly important when trying to investigate the impacts of important supply characteristics such as car parking cost and the availability of parking supply.

3.5.6 To circumvent this problem we analysed the impacts of supply characteristics indirectly through examining the extent to which trip rates differ between area types with different inherent levels of accessibility. There are also a number of locally specific accessibility variables in the NTS database, relating to population density, access to bus services and to rail terminals, which have been included within this analysis.

3.6 PROPOSALS TO IMPROVE METHODS FOR TRIP END MODELLING

3.6.1 This task uses the understanding gained in the previous steps to specify improvements to the methods to be used to estimate trip rates for zonal trip productions, segmented by person type and trip purpose. It includes recommendations on changes in segmentation details and on the area types for which differential trip ends are calculated. These recommendations cover the estimation of revised trip productions within NTEM for input to TEMPRO.

3.6.2 Guidance is provided on key issues including: the question of how they might be changed over time; the reliability that can be placed on the modal subdivision of trip rates; how to improve the modelling of rail; the impact of accessibility changes on trip rates; and on other findings that have emerged from the analysis.

4 Literature Review

4.1 INTRODUCTION

4.1.1 This Chapter presents the results of a survey, which was undertaken for this study by Rand Europe, of the literature and the current practice in modelling trip generation. The primary objectives of this literature review were: (a) to better understand the various methods of modelling and forecasting trip rates, with specific focus on trip generation; and (b) to gather any available evidence on the temporal stability of trip rates. The review examined foreign practice and research as well as UK and European methods.

4.1.2 The main objective of the trip generation step in any transport planning process is to produce reliable estimates of the number of trips generated (or originating) in a zone and the number of trips attracted to a zone. Models for forecasting the number of trips that originate in a zone (also referred to as trip productions) can be estimated at the zonal level as a function of zonal population and other attributes, as in traditional aggregate models, or at the individual level in terms of the frequency of trips made by an individual or a household. Models for forecasting trip attractions are typically estimated at the zonal level, as a function of zonal land use attributes, in aggregate models. In disaggregate travel demand models, forecasting of trip attractions is achieved through the use of choice models that model an individual's choice of trip destination. It is important to note, however, that in applying these trip production and attraction models for forecasting, aggregate models can only be applied using zonal aggregate characteristics. Disaggregate models, on the other hand, may be used to forecast disaggregate, individual-level trip rates using micro-simulation methods. Alternatively, as is more often the case in practice, disaggregate models are applied using aggregate zonal characteristics. The relative merits of aggregate and disaggregate models of trip generation are discussed in greater detail in section 4.3. This literature review was more focused on trip production models since the objective of this study is to better understand trip rates by individuals and households; trip attraction modelling therefore receives less attention.

4.1.3 Throughout, the emphasis is much more on the estimation of these models than on their application. In a literature review this is natural, because that is where the emphasis of the literature has been. However, application procedures for these models are generally straightforward. The issue is much more to provide input data than to operate the models themselves and the issue of forecasting the population and its characteristics is outside the scope of the study we have undertaken.

4.1.4 The rest of this Chapter is structured as follows. Sections 4.2 discusses the various issues associated with modelling trip generation rates. This then leads in Section 4.3 to a discussion of the pros and cons of modelling techniques, both in practice and academic research, ranging from aggregate methods such as linear regression and cross-classification to disaggregate approaches such as Poisson distribution models and models of random utility maximisation. Section 4.4 discusses the connection in trip generation between different travel purposes, including the importance of capturing correlations across travel purposes in modelling trip generation rates. Section 4.5 reports on the very small body of literature that is relevant to understanding temporal variations in trip rates. Section 4.6 concludes this Chapter with a brief overall summary.

4.2 MODELLING ISSUES

4.2.1 In this Section we describe in detail the various issues related to trip generation modelling as found in the literature. The first aspect of modelling trip generation is to define the unit of analysis. Next we touch briefly upon the differences between modelling trip productions and trip attractions. The rest of the Chapter then focuses on modelling trip production since the primary objective of this study is to understand how trip rates vary between individuals. This includes listing the key explanatory variables that were found to be significant in modelling trip rates and specifically examining the role of accessibility as an explanatory variable,

UNITS OF ANALYSIS

4.2.2 Aggregate models of trip generation are based on the trip as a unit of analysis. A trip is described as the travel between a specific origin and destination with a well-defined activity being performed at either end. Trip-based models typically develop separate model specifications for home-based and non-home-based trips. Since most travel surveys and traffic counts observe trips[1], they form a convenient unit of analysis. The choice of trips as a unit of analysis, however, largely ignores the fact that the trips undertaken by an individual are linked temporally and spatially although modelling on a production-attraction basis (rather than an origin-destination basis) alleviates some of these issues.

4.2.3 This shortcoming is further alleviated by the use of tours as a unit of analysis. The use of tours is typically accompanied by the use of disaggregate, individual level, modelling methods. In practical models, a tour is defined as the journey made by an individual from home (or work) to the primary destination and back. Tours therefore retain some of the links between trips – the fact that the trips in a tour follow one and another and typically use the same mode, the fact that the trips that make up a tour influence the choice of mode for the tour, and the geographical links between the trips that make up a tour. Tour frequency models typically form the state-of-the-practice in modelling trip generation, though there are also many trip-based models in use world-wide.

4.2.4 However, many tour-based models split travel into home-based tours and non-home-based trips and use separate models for each, for the sake of computational simplicity. Tour modelling also does not address the underlying factors that motivate travel in general i.e. the need to participate in activities and the fact that activities are correlated. In fact, most tour frequency models do not consider the chaining of trips for multiple purposes. Chaining of activities in a tour can produce positive correlations of the numbers of trips for different purposes. On the other hand, the numbers of trips undertaken for specific purposes (such as leisure) could suppress the number of trips undertaken for other purposes (such as social visits) due to time constraints or the satisfaction of the needs that motivate travel, thus exhibiting a negative correlation in trip generation for these purposes. These issues, which have become increasingly relevant with the advent of in-home activities such as social networking and e-shopping on the internet, are addressed by the modelling of activities.

[1] Many travel surveys also observe the stages that make up a trip (a stage being defined by the use of a distinct mode). However, for the purpose of trip generation models trips are a more convenient unit of analysis since the sequence of stages that make up a trip are not directly a function of the trip purpose but rather a function of the transport network.

4.2.5 Activity-based modelling techniques are based on the activity as a unit of analysis – with the activity being work or education or shopping etc. Activity-based models aim to model the activities that a person undertakes in order to satisfy his/her needs, in conjunction with the needs of the household (See, for instance, Scott and Kanaroglou, 2001, for an ordered probit model of activity episode generation. Larsen, 2003, models 'visits' or 'sojourns' in his paper, which are also derivable from an activity perspective). This is then translated to travel through the modelling of the choice of location for activity participation (home, work, school etc.), the choice of mode and time of day for travelling to the location if the location is out-of-home etc. Although the modelling of activities is best handled by a paradigm change in travel demand modelling, trip generation models could adopt activities as the unit of analysis to be converted to trip rates at a later stage using spatial-temporal disaggregation procedures.

TRIP PRODUCTION VS. ATTRACTION

4.2.6 There are two kinds of trip generation models used in traditional 4-step transport planning processes: production models and attraction models. Trip production models estimate the number of home-based trips to and from zones where trip makers reside. Trip attraction models estimate the number of home-based trips to and from each zone at the non-home end of the trip. Different production and attraction models are used for each trip purpose.

4.2.7 Most land-use development projects in the UK (especially in the south of England) use a trip attraction modelling package known as TRICS. TRICS is a computerised database and trip rates analysis package that relates at-site traffic counts to aggregate land-use characteristics. Effectively, TRICS provides trip attraction rates by over a hundred different land-use categories.

4.2.8 The equivalent of TRICS in the U.S. is the ITE (Institute of Transportation Engineers) Trip Generation Manual, although this provides both trip production and attraction rates for numerous land use and building types. However, a common critique of the ITE trip generation rates is that they are nearly twice as large as equivalent data from survey-based sources such as the National Personal Transportation Study and the various regional planning agencies. A number of potential reasons for the difference have been cited including biased sample demographics, underreporting, and the prevalence of non-home-based trips (or in other words, the effect of trip chaining).

4.2.9 Tools such as TRICS and the ITE Manual are typically based on cross-classification or regression methods using national data sources. These methods are discussed in greater detail in Section 4.3. For a detailed review specifically of trip attraction models, including methods to estimate pass-by and diverted link trips to a specific land-use, please see the relevant TRL report (Dasgupta et al., 1996).

4.2.10 The trip attraction model can be dealt with more rigorously in disaggregate travel demand models. These models typically begin with a tour frequency step that predicts the number of tours undertaken by an individual by purpose. This is followed by a mode-destination choice model that simultaneously predicts the choice of primary destination for each trip undertaken by an individual and the choice of mode for the specific trip. The 'attraction' estimates are therefore produced indirectly by the mode-destination choice model in a more realistic manner which avoids bias (see Daly, 1982). Although a disaggregate estimation of trip attraction models leads to model parameters that are statistically more reliable, the application of this model for forecasting is often undertaken in the same way as a more aggregate model.

4.2.11 A further important issue here is that of 'constraint', i.e. adjusting the numbers of trips or tours attracted to each zone so as to match exactly (or exactly proportionally) the predictions of an attraction model. For purposes such as commuting and education, the justification for this procedure is that the numbers of arrivals in a zone is determined by processes outside the transport sector. However, the critique is that this process, implemented by the use of 'balancing factors', equivalent to what economists would call shadow prices, gives too much credence to both the forecast data and the forecasting model, both of which are subject to error. No satisfactory approach has been found in the literature to answer these conflicting arguments.

4.2.12 The focus of this report (and the literature review) is predominantly on trip production. First, the primary objective of this study is to better understand the changes in trip rates by individuals and households, which is essentially forecasted by trip production models. Second, with the increasing use of tour-based models which model trip destination choice, traditional trip attraction models are becoming less important. Instead the focus is shifting towards capturing the spatial availability and accessibility of opportunities (or 'attractions') and their influence on trip production rates as well as the choice of destination and mode.

EXPLANATORY VARIABLES

4.2.13 There are two key factors in ensuring that trip generation is modelled in a behaviourally realistic manner. One is the choice of model structure (discussed in Section 4.3) and the other is the choice of covariates included in the trip generation model to capture the observed sources of variability in individual trip frequencies. Maximising the observed part of the model specification can help to ensure that the unobserved error in model specification is kept to a minimum.

4.2.14 The following is a list of variables that seem to be identified by several trip generation studies as being important sources of observed heterogeneity.

Individual characteristics

- Age
- Gender
- Ethnicity
- Marital Status
- Occupation and employment status
- Student status
- Licence holding

Household characteristics

- Household size
- Number of workers in household
- Number of children in household (< 5 years, 5-16 years)
- Household and personal income
- Car ownership
- Number of licensed drivers in household
- Housing type and home ownership status

Locational/Spatial characteristics

- Length of residence at current location
- Distance of residence from the different land uses
- Land-use mix around residence
- Accessibility of attractions/opportunities from residence and workplace

4.2.15 The use of locational and spatial characteristics, such as the ones listed above, in modelling trip generation rates appears to be more common in the U.S. as is evidenced by travel demand models developed by the North Carolina Department of Transport (NCDOT), the Bay Area Metropolitan Transportation Commission (MTC) and the Portland Metro in the state of Oregon to name a few.

4.2.16 In addition to the socio-demographic, land-use and accessibility related covariates listed in this section, there have been several studies on the effects of trip chaining on trip/tour frequencies. See, for instance, Goulias and Kitamura (1989), Golob (1999) and Noland and Thomas (2007).

4.2.17 Some of the other interesting studies include the impacts of telecommuting on trip frequency (Pendyala et al, 1991); the impacts of work-trip duration on trip frequency (Purvis et al, 1996); and the spatial-temporal travel patterns of the baby-boomer cohort (Miranda-Moreno and Lee-Gosselin, 2008).

ROLE OF ACCESSIBILITY

4.2.18 The accessibility of a zone is another key covariate in modelling trip/tour frequency. Accessibility captures not only the land use mix around the zone but also the ease of mobility from the zone to potential sources of 'attraction'. The logsum (from a mode-destination choice model, for instance) is often used as measure of accessibility. Daly (1997), Daly & Miller (2006) and Larsen (2003) discuss the importance, and the difficulty, of incorporating accessibility within a trip generation model. Several studies by RAND Europe also indicate significant accessibility effects on trip (or tour) generation and travel demand. See, for instance, Fox et al. (2003) and Kouwenhoven et al. (2006).

4.2.19 There is a fairly large body of literature looking at the value of including accessibility measures (or indices) in travel demand models, and specifically in trip production models although there is no definite conclusion from these. For instance, the paper by Leake and Huzayyin (1980) is specifically focused on the importance of accessibility measures in trip production models. They test a variety of accessibility measure formulations within a regression-based model of trip production and demonstrate the improvements in model fit and overall forecasts. On the other hand, others such as Ewing et al (1996) have found that, after controlling for socio-demographic variables, residential density, mixed use, and accessibility do not have significant, independent effects on household trip rates. Whereas, Koenig (1980) found, based on a study of five French cities, that accessibility (including walk trips) was a very good determinant of trip generation for non-working people. He argues that accessibility may be ineffective in explaining trip generation when calculated by mode and when walking trips are excluded. Unfortunately most of these studies are limited to demonstrating significant statistical relationships between accessibility measures and trip rates, and we were unable to locate any studies that explicitly quantify the value of these effects.

4.2.20 Another interesting study from recent times is one by the Metropolitan Transportation Commission (MTC) in the San Francisco Bay Area that incorporates work trip accessibility in the form of home-based work trip duration in non-work trip generation models (see Purvis et al, 1996, 1998) and quantifies the effect of this accessibility measure. This effort was motivated by the evidence that the survey data and models exhibited an inverse relationship between work trip duration and home-based non-work trip frequency i.e. as work trip duration increases, non-work trip frequency decreases. Elasticity analyses of the final trip generation model demonstrated that a 10% decrease in the regional work trip duration yields a 1.2% increase in regional home-based shopping/other trips and a 0.9% increase in regional home-based social/ recreation trips. Purvis et al. and the studies referenced by Fox et al. (2003) include all modes (including walk and bicycle) in the estimation of the non-work trip frequency model and therefore provide valuable evidence in favour of including accessibility measures.

4.3 MODELLING TECHNIQUES

4.3.1 There is a variety of techniques available for modelling trip rates, ranging from cross classification and regression analysis to Poisson and nested logit models. While some of these are estimated at the zonal, and therefore aggregate, level others are estimated at a disaggregate individual (or household) level. There also exists a range of models that are aggregate or disaggregate to different degrees. In application for forecasting, however, most models in practice use relatively aggregate data. While aggregate zonal models require zonal level attributes of population, household size etc for forecasting, other models that are either estimated using less aggregate (for instance, use a large number of categories of person-types) or entirely disaggregate data are also applied in forecasting using zonal level attributes thus rendering them equivalent to the aggregate models.

4.3.2 The primary advantages of the disaggregate models are really twofold. First, from a statistical standpoint, the estimation of trip rates models using disaggregate data results in more reliable and unbiased model parameters. Second, disaggregate trip rates models can also be applied in a disaggregate manner, if desired, using micro-simulation methods, though this approach is not yet widely adopted.

PRACTICE VS. RESEARCH

Aggregate Models

4.3.3 At the one end of practice are zonal level trip generation models, such as the ITE Trip Rates in the US, which are typically statistical models that describe the observed flows between zones as a function of the aggregate demographic, land use and economic attributes of a zone. These models usually use methods such as multiple regression analysis, cross-classification (sometimes called category analysis) and experience-based analysis (discussed further in the following sections). These methods are described in the standard textbook (Ortúzar and Willumsen, 2001).

4.3.4 The literature in aggregate trip generation modelling also extends to tobit regression models and Poisson regression models. Tobit models account for the fact that the number of trips originating in a zone is bounded by 0 on the lower end, while continuing to maintain a continuous formulation for the number of trips. Poisson regression models on the other hand acknowledge the fact that trips are countable and treat the number of trips originating in a zone as a non-negative integer.

Disaggregate Models

4.3.5 The basic emphasis of disaggregate models of trip generation is to capture the factors, both observed and unobserved, that drive travel behaviour and accordingly disaggregate estimation methods yield statistically better model parameters. Such disaggregate trip generation models use a variety of modelling techniques such as Poisson distribution models, negative binomial distribution models, structural equations models, ordered probit models and binary or nested logit models as described later in this section. Most of these models have a basis in the theory of utility maximisation.

Practice

4.3.6 The typical practice of trip generation in the UK is to use TEMPRO forecasts of trip ends, which are based on the National Trip End Model (NTEM). TEMPRO effectively provides trip rates by travel purpose, household size, car availability and few other zonal socio-demographic variables and is based on category analysis of the NTS data. However, other than by their indirect impact through car ownership, TEMPRO does not account for the effects of accessibility or household income on trip generation, which have been identified as being independently important as described in the previous section. While TEMPRO is commonly used for large scale planning purposes, most land-use development projects in the UK (especially in the south of England) use TRICS.

4.3.7 In the US, the typical practice of trip generation in smaller planning areas is to use the ITE Manual. Tools such as TEMPRO and the ITE Manual are typically based on cross-classification or regression methods using national data sources. The inherent assumption is that, in the absence of local data, the national trip rates can be safely applied for local planning purposes. This approach can be improved by the use of methods such as that suggested by Dey and Fricker (1994) who posit a Bayesian updating of national trip generation data using relatively small samples of local data.

4.3.8 The state-of-the-practice in the U.S., as evidenced in larger planning areas, is typically person and household level models of trip generation by trip purpose. The person-level model used by Portland METRO is one such example, although this model does not include accessibility and land-use characteristics among its explanatory variables. The trip generation models developed by the Metropolitan Transportation Commission (MTC) of the San Francisco Bay Area, on the other hand, capture accessibility effects quite well (see secrion 2.4) but are hampered by the use of a more aggregate, linear regression modelling approach.

4.3.9 At the other extreme of the state-of-the-practice are disaggregate models of trip and tour frequency, such as those applied by RAND Europe in the strategic model for the West Midlands in the UK (PRISM), the Dutch National Model, the RATP model for Paris, and the transport models for Copenhagen, Stockholm and Sydney. These are mostly logit models of tour frequency based on national travel survey data sources that model the number of tours made by an individual, for a specific travel purpose, as a function of the socioeconomic characteristics of the individual, the land-use and accessibility variables. See Fox et al. (2003) for a further description of these models. As discussed earlier, these models contain statistically reliable parameters estimated through disaggregate techniques, but for the sake of convenience and computational tractability they are applied for forecasting using aggregate zonal data.

4.3.10 There are also a handful of activity-based travel demand models used in practice, such as in Columbus (Ohio), New York, Sacramento (California), Portland (Oregon) and Tel Aviv (Israel). Since the basic paradigm of these modelling systems involves activity patterns, they do not explicitly model trip rates. Rather they model activity generation and scheduling and the trip rates are derived from the implementation of the entire modelling system.

Research

4.3.11 There are several other academic approaches that extend the envelope of these state-of-the-practice methods, such as the activity generation models and models of tours that explicitly consider trip chaining effects (see, for instance, Goulias and Kitamura, 1989, and Golob, 1999). The state-of-the-art focuses conceptually on capturing the inter-relationships between trips and the effects of these interactions on trip rates; and more fundamentally on understanding the underlying motivations for travel. Two broad areas of research include the activity-based (Bhat and Koppelman, 1999) and need-based (Arentze and Timmermans, 2007) paradigms. From a theoretical perspective, the state-of-the-art focuses on the incorporation of suitable explanatory variables, such as accessibility (e.g. Daly and Miller, 2006), and the choice of appropriate modelling structure, such as Poisson (e.g. Berkhout and Plug, 2004), Poisson-logit (e.g. Larsen, 2003), negative binomial (e.g. Greene, 2007), ordered probit (e.g. Noland and Thomas, 2007) and so on.

4.3.12 In the following sections we examine the various modelling techniques in a little more detail, starting with experience-based analysis to structural equations models.

EXPERIENCE-BASED ANALYSIS

4.3.13 One of the oldest, most commonly used, techniques of modelling trip generation is through 'Experience-Based Analysis'. The ITE Manual of Trip Generation, and TRICS are two of the best examples of generalised trip generation rates produced by experience-based analysis. For instance, the ITE manual is a compilation of data from all over North America on many different types of land uses. Within the manual, productions and attractions for each type of land use are related to some measurable variable. For example, a shopping centre might produce a certain number of trips for each employee. Simply asking for the employment roster would allow a transportation engineer to estimate the total number of trips that are generated by the shopping centre employees. To establish local credibility, a survey of similar land uses in the area may also need to be conducted. Early travel forecasting used extrapolation of past trends to estimate future travel. Such an approach is still used occasionally for estimating future traffic on a single facility, in a relatively isolated area, where only moderate and uniform growth or change in development pattern is anticipated.

4.3.14 Most transport planning procedures today have moved away from trend analysis towards the behavioural modelling of travel. Consequently, most trip (or activity) generation methods in use today attempt to capture the observed and unobserved sources of variability in the generation of trips (or tour or activity) by individuals.

CROSS-CLASSIFICATION

4.3.15 'Cross-Classification' procedures measure the changes in one variable (trips) when other variables (land use etc.) are accounted for. Cross-Classification resembles multiple regression techniques but is based on tabulation rather than mathematical modelling. Cross-classification models group individual households together according to common socioeconomic characteristics (car ownership level, income, household size, etc.) to create relatively homogeneous groups. Average trip production rates are then computed for each group from observed data. Cross-classification analysis can be similarly performed for trip attraction calculations. Most trip production models are two- or three-way cross-classification tables with the dependent variable being trips per household or trips per person. The independent variables are most often income, car ownership, and household size. Virtually all of the trip attraction models use employment and an identifier of location as independent variables.

4.3.16 The primary advantages of cross-classification techniques are that they are simple to apply, capture correlations among the independent variables well, and imposes no a priori assumptions about functional relationships among the variables. The drawbacks of this technique are:

- large numbers of categories or dimensions lead quickly to empty or sparsely populated cells;
- because of this sample size problem, it is typically necessary to minimize the number of cells either by limiting the number of variables or by aggregating the values for each variable into a few ranges;
- also because of the sample size problem, confidence intervals on cell mean values may exhibit wide variation among cells;
- the method is sensitive to the grouping applied in defining ranges for each variable;
- it is not possible to investigate hypotheses that the impact of one variable is constant across different values of other variables, as all cells are assumed to be independent of each other;
- when the dependent variable is a zonal average, the cross-classification method is sensitive to the zone system used (though the use of average per capita variables alleviates this issue to some extent); and

- 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.

4.3.17 Nevertheless, cross-classification continues to remain one of the most common methods in current practice, and is a reliable method when a small number of variables is thought to be sufficient for a good trip generation model.

LINEAR REGRESSION ANALYSIS

4.3.18 Linear Regression Analysis is a statistically rigorous form of cross-classification that is based on modelling trip generation as a function of one or more independent variables. First, data regarding the actual number of productions and attractions is coupled with data about the area that is thought to impact the production and attraction of trips. For instance, the total population is believed to impact the number of trips produced. If we know the number of trips produced and the population for the present and a few time periods in the past, it is possible to develop a relationship between these parameters using statistical regression. Once we are satisfied with the relationship that has been developed, we can extrapolate into the future by plugging the future population into our relationship and solving for the number of productions. Early trip generation models were commonly developed by regression analysis because of its power and simplicity. The independent variables in such models were usually zonal averages of the various factors of influence. Trip generation equations developed by regression are still used by some planning agencies, more commonly for attraction models than for production models. This is because only zonal averages of trip attracting characteristics are usually available since most travel surveys do not survey at trip destinations. Obtaining more detailed data for individual attraction zones requires a survey of trip attractors, such as a workplace survey.

4.3.19 The imposition of linearity introduces a number of problems in regression models of trip generation. For example, most surveys have shown that trip-making is not linearly related to car ownership, but increases dramatically with the first car and to a declining extent as the number of cars increases. The use of a linear form in such circumstances represents a basic misspecification. Transforms of variables (e.g., log-linear forms, exponential forms, Box-Cox transforms, Box-Tukey transforms) provide a way of overcoming some of these difficulties while retaining the use of linear regression estimation software. Nonlinear regression techniques allow more modelling flexibility but are less frequently available in basic statistical software packages and hence are less commonly applied. Nevertheless, these techniques are finding their way into use in some regional planning models, primarily because nonlinear models allow both a high degree of flexibility in functional form (much like cross-classification) and a large number of explanatory variables.

4.3.20 As trip generation models became more disaggregate and move towards the generation of trips by an individual, it became increasingly important to pay close attention to the model structure. For instance, linear regression models entirely ignore the fact that the number of trips generated by an individual are ‘countable’ (or discrete) and non-negative. This motivated several academic studies to develop model structures that are better suited to modelling the frequency of trips. For instance, Cotrus et al (2005) compare tobit models against linear regression models of trip generation in Israel. The basic motivation for using a tobit model is that in a tobit model the dependent variable (number of trips) is bounded by 0, which improves the behavioural realism of the model structure. Several other researchers have since proved the value of using a Poisson distribution model in modelling count data such as trip frequency.

POISSON MODEL

4.3.21 Barmby and Doornik (1989) were one of the first to specify and estimate Poisson distribution models of trip frequency. The basic concept of a Poisson distribution model is that the number of trips (Y) generated by an individual follows a Poisson distribution. Therefore, the probability that the number of trips made by an individual is equal to a specific value y (y=0,1,2,3…) is given as

$$P(Y = y) = f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad \lambda > 0 \qquad \text{(Eq. 1)}$$

where λ is the parameter of the distribution and represents both the mean and the variance of the Poisson distribution. λ may in turn vary with covariates *x* expressing observed heterogeneity among individuals, as shown in equation 2 (β in equation 2 is a vector of parameters that measure the effects of covariates x, such as age, gender etc., on λ). The Poisson model is typically estimated using maximum likelihood techniques.

$$\lambda = e^{x'\beta} \qquad \text{(Eq. 2)}$$

4.3.22 The primary merits of the Poisson distribution model are two-fold. First, it represents trips realistically as non-negative and integer random variables. Second, the Poisson maximum likelihood estimator is robust to distributional misspecification, and maintains certain efficiency properties even when the distribution is not Poisson. However, Poisson models suffer from the limitation that they have only one parameter (which in fact is equal to both the mean and the variance) defining all moments and make the underlying assumption that successive events are independent. Several researchers suggest that individuals exhibit a high level of heterogeneity in their trip frequencies which is not adequately captured by observed characteristics thereby leading to a high variance (or dispersion) in the Poisson distribution (see, Larsen, 2003, and Berkhout and Plug, 2007 for a detailed discussion). A Poisson distribution model typically underestimates this variance (known as misdispersion in the literature) either because the assumption of a Poisson distribution is incorrect or because there is a high level of unobserved heterogeneity in the trip frequency. This can be overcome to some extent by introducing an error term into the Poisson distribution model as shown in equation 3.

$$\lambda = e^{x'\beta+\varepsilon} \qquad \text{(Eq. 3)}$$

NEGATIVE BINOMIAL MODEL

4.3.23 Another possible extension suggested to the Poisson model to overcome this limitation is the negative binomial model. The negative binomial distribution is a generalization of the Poisson distribution with separate parameters for the mean and variance thus overcoming the issue of misdispersion. It is equivalent to the generalisation of the Poisson distribution with an error term, as in equation 3, providing the error has a gamma distribution. Maximum likelihood estimation techniques can be used to estimate the parameters of a negative binomial model. Greene (2007) presents a detailed description of the negative binomial model. Larsen (2003) demonstrates empirically that although there is not much dispersion in work trip frequency (variance practically identical to the mean), the trip rates for other purposes such as business, shopping, and private visits exhibit overdispersion suggesting that a negative binomial model structure would be more suitable.

RUM MODELS

4.3.24 Another class of models that is often used to model trip and tour frequencies is the random utility maximization (RUM) models. The most commonly used of these are the multinomial logit models, ordered probit models and nested logit models, although Daly and Miller (2006) recently demonstrated that the Poisson distribution models also have some basis in random utility theory. The more traditional RUM models (logit and probit models) have the advantage that they treat the number of trips (or tours) undertaken by an individual as discrete, integer, numbers. Moreover the random utility theory on which these models are based have a strong behavioural basis, and account quite rigorously for various sources of observed and unobserved variability (and correlations therein) in individual behaviour. This gives these models a strong link to the appraisal process.

4.3.25 Ordered probit models are a special case of probit models that are applicable specifically to ordinal multinomial dependent variables such as the number of trips undertaken by an individual on a given day. The central idea is that there is a latent continuous metric (that can be looked upon as a latent 'desire' to make trips) underlying the ordinal responses observed by the analyst. Thresholds partition the real line into a series of regions corresponding to the various ordinal categories (such as the number of trips in a model of trip frequency). As described above these models can also be loosely based on the RUM principle. See Boarnet et al. (2003), Schmöcker et al. (2006) and Sugie et al. (2003) for sample applications. An empirical comparison of negative binomial and ordered probit models by Noland and Thomas (2007) indicates that for the most part both models produce similar results with a few notable exceptions in the coefficient estimates. The more important limitation of the multinomial logit and ordered probit models is the fact that they treat a frequency of 0 trips in the same way that they treat 1 or 2 or 3 trips. Intuitively, however, the decision to make a trip or not guides the observations with zero trip frequency, whereas the choice between 1, 2, 3 or more trips may be guided by a combination of needs, time and monetary budgets etc.

4.3.26 This difference in motivation between the decision to make a trip and the number of trips made given the decision to make a trip are best addressed by model structures such as nested logit and Poisson-logit. Both these model structures are similar in that they have a higher level choice between 0 and 1 or more trips, followed by a choice between 1,2,3,4 etc. trips conditional on the probability of the individual choosing to make one or more trips. While the nested logit model assumes a Gumbel distribution for the higher level choice, the Poisson-logit model assumes a Poisson distribution for the same. Larsen (2003) describes the structure of the Poisson-logit model in detail and compares it to a simple Poisson model both theoretically and empirically. We could find hardly any empirical evidence comparing the performance of the Poisson-logit and the nested logit models and the choice between these two model structures must be driven by beliefs in the distributional form of the binary random variable denoting the decision to undertaken 1 or more trips.

4.3.27 The extension of the logit and Poisson models to nested logit and logit-Poisson is analogous to the extension of the Poisson regression to negative binomial, as both extensions try to capture the same effect of 'over-dispersion'.

STRUCTURAL EQUATIONS MODELS

4.3.28 Another class of models that has often been used to model trip frequencies is the structural equations models (SEM). Structural equation modelling is a statistical technique for testing and estimating causal relationships using a combination of statistical data and qualitative causal assumptions. SEM encourages confirmatory rather than exploratory modelling and is better suited to theory testing than theory development. It usually starts with a hypothesis, represents it as a model (typically a system of equations), operationalises the constructs of interest with a measurement instrument, and tests the model. For an application of SEM to trip generation, see Golob (1987). While there have been several applications of SEM to test hypotheses of trip frequencies there are hardly any applications within a travel demand forecasting system.

4.4 CORRELATION ACROSS TRAVEL PURPOSES

4.4.1 Another key methodological and conceptual issue discussed in the trip generation literature is the correlation between the numbers of trips undertaken for different purposes (work, education, shopping etc.). Most trip/tour frequency models treat travel undertaken for different purposes as independent observations. However, people often combine trips both with the same purpose and for different purposes within a tour. As discussed in Larsen (2003),

4.4.2 "…there are at least two reasons why we may expect some interdependence between the numbers of visits with different purposes. <u>One</u>, the combination of two or more visits in a tour occurs because the opportunity cost and/or utility of one visit is influenced by the presence of other visits in a tour. The most obvious reason for this is that travel time and travel cost for a tour with two visits may be less than the travel time and cost for two tours with one visit each. This type of interdependence should emerge as a positive correlation. <u>Two</u>, already having made one tour with one or more visits for some purpose(s) may increase the opportunity cost of making another tour to carry out a visit with still another purpose. This should emerge as a negative correlation".

4.4.3 In order to get a reliable forecast of trip/tour frequencies, it is therefore important to accurately capture the interdependencies between trip/tour frequencies for different travel purposes. There are several methods that can be used to detect interdependence. One method is to examine the covariance matrix of the estimated residuals from the independently estimated models for the different trip/tour purposes. Another method is to calculate the sum of the variances for each trip/tour purpose with the variance calculated for the total number of trips/tours.

4.4.4 Larsen suggests a conditional Poisson model structure or a joint Poisson-logit and multinomial model structure to account for the interdependencies between travel purposes. A Poisson-logit model, as described by Larsen, is essentially a Poisson model that relaxes the assumption of equidispersion and has a behavioural interpretation consistent with the random utility framework. The Poisson-logit model treats the parameter for zero occurrence of an event as different from the parameter for one more occurrences in a Poisson model (a more general version of the Poisson-logit model is the negative binomial model). In order to capture the interdependencies between travel purposes, Larsen takes the Poisson-logit model further to propose a simultaneous Poisson-logit model for the number of trips and a multinomial model for the distribution of the total number of trips between purposes conditional on the total number of trips. However, he stops short of deriving a working model.

4.4.5 Berkhout and Plug (2007) propose a conditional Poisson model to capture interdependencies between travel purposes and develop it into a working model with an empirical application for visits to cultural events and tourist attractions. The basic theory underlying this model is one of conditional probability.

4.5 TEMPORAL STABILITY

4.5.1 As part of this literature review we also looked for studies documenting any temporal trends in trip generation rates. Interestingly, there are hardly any studies that examine broad temporal trends in trip rates although there are several studies examining temporal variations in trip rates for an individual (see, for example, Kitamura and Van Der Hoorn, 1987).

4.5.2 The only relevant literature in this area appears to be a series of studies based in Reading that suggest that trip rates are generally temporally stable. However, this is a rather old study and is based on travel data over the period 1962-1971. The other relevant study by Hupkes (1988) suggests the "law of constant travel time and trip rates". However, this study appears to be suggesting that the combination of trip rates and travel times tends to be a constant, which is tied in with the extensive body of literature focused on 'Travel Budgets' (see, Zahavi, 1974, and Shafer, 2000).

4.6 CONCLUSION

4.6.1 In conclusion, applied trip generation models range from aggregate, experience-based, trip rates models to disaggregate, individual-level, tour frequency models. The state-of-the-practice is typically disaggregate tour-based models that model home-based tour frequency and non-home-based trip frequency by purpose. The primary advantage of a disaggregate formulation is the ability to estimate reliable and statistically rigorous model parameters. In application for forecasting, although disaggregate models may be applied at the individual-level using micro-simulation methods they are more commonly applied using aggregate data.

4.6.2 Several academic papers propose conceptual and theoretical enhancements to trip generation model structures and the most practical of these approaches are presented in this report. Although the studies referenced in this report are not exhaustive by any means they are fairly representative of the wide range of topics in the area of trip generation modelling. For more general details on trip generation modelling methods the reader is referred to the book by Ortúzar and Willumsen (2001).

5 Trip Rates: Statistical Analysis Techniques

5.1 INTRODUCTION

5.1.1 A brief description of the use of Poisson and Negative Binomial regression to estimate trip rates has been provided in the Literature Review in Section 4.3. As discussed, the Poisson regression methodology was developed to model non-negative integer (count) data such as trip frequency. This methodology was initially chosen for analysing trip rates for this research study. However, the outcome from conducting the Cameron and Trivedi (1990) statistical test on overdispersion has revealed that the distribution of individual's trip rates for the majority of travel purposes is not in line with the key Poisson regression assumption– that the variance is equal to the mean. Therefore, negative binomial regression (NBR), which is the general form of Poisson regression, has instead been used for this study.

5.1.2 This Chapter provides a more detailed explanation and mathematical specification of Poisson regression which is followed by a discussion of how the NBR overcomes the limitation of the Poisson regression . It then provides the specification and results of applying Cameron and Trivedi statistical tests on our dataset which shows that the simpler Poisson would underestimate the error terms associated with each independent variable because it suffers from overdispersion.

5.2 THEORETICAL BACKGROUND

5.2.1 As noted earlier, the Poisson regression methodology is a form of regression analysis used to model count data such as the number of trips made by an individual.. Poisson regression is based on the assumption that the dependent variable Y_i has a Poisson distribution, which in turn is a limiting case of the binomial distribution.

5.2.2 The binomial distribution relates to the discrete random variable which is the number of successes in a sample of *n* observations during a time interval of given length. Given that λ is the expected number of successful occurrences in the interval, the probability of success, which should be equal over all observations, is $p = \lambda / n$. The probability of exactly y successes in the given time interval can then be estimated by multiplying the number of all possible combinations of *y* successes in *n* observations by the probability of y successes in each of the combinations as follows:

$$P(y \mid n, \lambda/n) = \binom{n}{y}(\lambda/n)^y(1-\lambda/n)^{n-y} = \frac{n!}{y!(n-y)!}(\lambda/n)^y(1-\lambda/n)^{n-y}$$

Equation 5-1

5.2.3 In practice, evaluating the terms in the binomial probability formula is tedious, and with the extra assumption that *n* is very large and *p* is relatively small, it can be shown that the Poisson distribution with parameter λ is a close approximation[2]. With this assumption, Equation 5-1 could then be replaced by the form shown in Equation 5-2. For more information on the mathematical details, please refer to Appendix A.

Equation 5-2

$$P(y,\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$$

[2] NB There is a rule of thumb stating that the Poisson distribution is a good approximation of the binomial distribution if n is at least 20 and p is smaller than or equal to 0.05. According to this rule the approximation is excellent if n ≥ 100 and np ≤ 10

Hence taking the log gives

$$\ln P(y,\lambda) = -\lambda + y\ln\lambda - \ln(y!)$$ Equation 5-3

5.2.4 Poisson regression models are estimated by specifying the Poisson parameter (λ_i) as a function of explanatory variables *X*. The most common relationship between explanatory variables and the Poisson parameters is the log-linear model,

$$E[y_i] = \lambda_i = EXP(\beta X_i) \text{ or } LN(\lambda_i) = \beta X_i$$ Equation 5-4

5.2.5 Here i is an observation, y_i is the observed number of events, X_i is a vector of explanatory variables and λ_i denotes the expected/predicted number of events per period

5.2.6 Note a fundamental difference between a classical linear regression model and the specification for the conditional mean in the Poisson regression model: the latter does not contain a random error term (in its "pure" form). Consequently, unlike the approach taken for the linear regression, the log-likelihood is not derived from the joint density of the random errors, but from the distribution of the dependent variable itself.

5.2.7 By substituting Equation 5-4 into Equation 5-3, the log likelihood function results, as shown in Equation 5-5. The value of β is then estimated by maximizing this log likelihood function.

Equation 5-5

$$LL(\beta) = \sum_{i=1}^{n}[-EXP(\beta X_i) + y_i\beta X_i - LN(y_i!)]$$

5.2.8 In order to maximise the likelihood function the first order condition ∂L/∂β=0, yields a system of K equations (one for each β) of the form:

Equation 5-6

$$\sum_{i=1}^{n}[-EXP(\beta X_i) + y_i]X_i = 0$$

As β enters this equation in a non linear form, there is no analytical solution for solving Equation 5-6. It is typically solved using the Newton-Raphson method.

5.3 POISSON REGRESSION GOODNESS OF FIT

5.3.1 The goodness of fit indicator R^2, which is widely used in OLS regression, is not directly available for the Poisson regression or similar models that are estimated by Maximum Likelihood Estimation (MLE). There are several alternative approaches to estimate the goodness of fit for the Poisson regression. Two of those which are close to the approach being used by the statistical package R are described below

5.3.2 As an analogue to the F-test in the classical regression model, the likelihood ratio test is a common test used to assess two competing models in a Poisson regression. A full model and a restricted one are normally compared against each other. The restricted model is considered to have all its coefficients (i.e. β) set to zero and only its constant value to be non zero (NB this is actually the "null" model: the point is that any more fully specified model can be compared with a restricted model, based on a set of linear constraints on the estimated coefficients). The Full model, on the other hand, is the estimation based on testing all independent variables included in the model. Based on the Wald test, the difference between the log likelihood functions of these two models is assumed to have the χ^2 distribution. Therefore, the χ^2 test is used to assess if they are significantly different and to measure whether the extra variables included are adding significant value to the model. Equation 5-7 shows the mathematical expression of the test. The degrees of freedom (df) would be equal to the difference between the number of estimated coefficients in the restricted and full models,

$$\chi^2{}_{df} = 2[LL(\beta_R) - LL(\beta_{Full})]$$ Equation 5-7

5.3.3 An additional indicator which is provided by R and is useful in comparing several models estimated on the same data is Akaike's Information Criterion (AIC). AIC is normally used when the maximum likelihood approach is used to estimate the variables. It is defined as:

$$AIC = -2LL(\beta_{Full}) + 2K$$ Equation 5-8

where $LL(\beta_{Full})$ is the log likelihood value of the full model and K is number of estimated parameters to control for the sample size. Therefore the smaller the AIC, the better the model is estimated.

5.3.4 This AIC measure approach is used in this study as the basis for comparing various estimated models. The benefit of using AIC is that it takes account of the number of parameters used in the model when estimating the overall goodness of fit, which helps to focus on simpler parsimonious models with fewer parameters.

5.4 TESTING THE SIGNIFICANCE OF VARIABLES

5.4.1 T test or Z scores are used to test the significance of each of the specific variables in its ability to explain the variations in the dependent variable - the number of trips. The package R will estimate for each coefficient its Z score and the probability that the Null hypothesis is true (i.e. that the true value of the coefficient is zero so that the inclusion of this variable does not add any value to the model). Therefore, the closer the probability of the Null hypothesis being true is to zero, the more significant is that dependent variable.

5.5 VIOLATION OF ASSUMPTIONS - TESTING THE VALIDITY OF THE POISSON REGRESSION

5.5.1 The following set of assumptions is made when using the Poisson regression. To guarantee we have a valid model, it is essential to be aware of and to control for these assumptions.

5.5.2 Failure to recognize that the data might be truncated is one possible source of error in analysis. Estimating the Poisson regression without accounting for this error might result in a biased estimate of the parameter vector β. In terms of trip rate analysis, this issue is less likely to happen as people are not limited to doing a specific number of trips in a week for any specific purpose.

5.5.3 As noted, the Poisson distribution has the assumption that all its moments are defined by a single parameter. Among other things, it is implied that the dependent variable has the same value for its mean and variance. If the variance has been proved to be much larger than the mean, the data would be called "over dispersed". The negative binomial regression (NBR) has been proven to do better in this case. The NBR is derived by rewriting Equation 5-4 as $E[y_i] = \lambda_i = EXP(\beta X_i + \varepsilon_i)$ where $EXP(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α^2. The addition of this error term allows the variance to differ from the mean. α is called the over-dispersion parameter (note: in the R statistical package used for this study and most other statistical packages it is the inverse $1/\alpha$ which is considered as the overdispersion parameter).

5.5.4 One useful test for investigating overdispersion is provided by Cameron and Trivedi (1990). Appendix B gives more details on its mathematical specification. After coding the test in R, we checked for all travel purposes the performance of our models built using the Poisson regression approach. The results of this test along with the estimated dispersion parameter from running the negative binomial model are shown in Table 5.1 for the main purposes. These results suggest that Poisson regression is not a suitable approach to be used for modelling trip rates for almost any of the purposes tested, although it can be argued that HBPB and HBEB (for FT and PT workers only) suffer to a lesser extent from overdispersion than the other purposes. This finding is close to Larsen's (2003) results that suggest that there is not much dispersion in work trip frequency but that the trip rates for other purposes such as business, shopping, and private visits exhibit overdispersion. Our findings suggest that various factors including the dataset, number of parameters used in the model and the units and format of dependent variables are all important in the severity of overdispersion.

Table 5.1 Comparison of the Cameron and Trivedi test results by trip purpose[3]

Purpose (b)	T value	Pr (> |t|)	Sig code	Dispersion parameter
HBW (FT and PT)	-7.107	1.21e-12	***	1.700
HBEB (FT and PT)	-2.718	0.00657	**	0.115
HBShop	-4.849	1.24e-06	***	1.095
HBHol	4.231	2.33e-05	***	0.351
HBPB	-2.298	0.0215	*	0.492

5.5.5 We therefore decided to use the negative binomial regression approach as the main method for estimating trip rates in this study. A more detailed description and mathematical specification of this approach is provided in Appendix C.

5.5.6 The dependent variable in the regression is the weekly number of trips per person for the trip category and person type of interest. The explanatory variables are the set of continuous and categorical variables that are expected to influence these trip rates. Note that the basic model form (Eq 5-4) implies a multiplicative relationship: this is because the estimated coefficients β need to be exponentiated to produce the required estimate of the number of trips

[3] All models which compared against each other are the simplest tested form with the same parameters as those had used in the original NTEM. So the they can be compared against each other in terms of the extent by which they were hit by overdispersion The test were carries out by assuming both cases of g(E[y_i] = E[y_i] and g(E[y_i]) = $E[y_i]^2$. However, the latter is reported in this table.

6 National Travel Survey: Trends and Consistency

6.1 INTRODUCTION

6.1.1 The primary aim of this study is to ascertain whether trip rates have been changing systematically through time or whether it is reasonable to assume them to be approximately constant over time. One relevant topic for investigation was to check whether the apparent reduction in trip rates that has been observed in the NTS may simply be a result of differences over time in the household response rates that have been achieved by this survey. This Chapter analyses the evolution of the NTS methodology and survey operation to improve understanding of its potential impact on measured trip rates.

6.1.2 The hypothesis that is tested here is:

> *Trip rates in reality are constant through time but in periods where the NTS can only achieve lower response rates in the survey, the trip rates that it measures are biased downwards due to differences in the characteristics of respondents compared with non-respondents.*

6.1.3 The assumption underlying this hypothesis is that those most likely to respond to home-based surveys, such as the NTS, are those people who spend the most time at home. This would include the unemployed, the elderly and infirm and other individuals who are known to make fewer than average trips. In contrast it appears likely that those who would be most difficult to "persuade" to complete a survey are those who spend a lot of time away from their home travelling on various types of trips. Here we investigate the evidence that would support or contradict the validity of the above hypothesis.

6.1.4 This apparent contradiction in the past in the direction of trip rate trends is an important indicator of the need to examine with great care the internal consistency over time and the sampling errors and other potential sources of error within this NTS dataset that is being used to examine trends in travel behaviour.

6.1.5 The key issue here is to ascertain whether the apparent recent trend to reductions in trip rates could be an aberration due primarily to data issues. Such issues could arise from differences in the methods, response rates, expansion factors or other features of the NTS survey over time.

6.1.6 This Chapter reviews various general sources of empirical evidence about trends in trip rates through time. Then detailed empirical analysis of the existence and strength of any temporal trends is carried out in Chapter 8 using fully specified behavioural models. The intermediate Chapter 7 introduces the data elements from the NTS that are used in constructing these models.

6.2 CHECKING THE PROCESSING OF THE NTS DATABASE

6.2.1 DfT initially provided the NTS data to this study in SPSS format for the years 2002-2006 only. The SPSS data set contained both the data and the definitions (attributes) of the variables. Eight sets of data and attributes tables were provided as shown in Table 6.1.

Table 6.1 Data tables provided from NTS for 2002-2006

Data table	Description / contents
Day	Travel survey day (1 to 7)
Household	Household related variables – eg numbers of resident adults, income etc
Individual	Individual related variables – eg gender, age etc
Journey	Variables specific to each journey made – eg purpose from, time started
Ldj	The long distance survey: entries cover a longer interval than the survey week
Psu	Variables specific to the post code sector unit in which household is located.
Stage	Journey stage
Vehicle	Information on vehicles available to households surveyed.

6.2.2 A number of checks summarised below have been carried out on this dataset to make sure it has all the categories required to examine the temporal stability of trip rates and that when processed it matches exactly to the published data on travel patterns.

6.2.3 Tables of households, individuals, journeys and area type (ie psu) were imported into a local database both for checking and for creating the data required to be analysed in the R statistical package. The first check was to make sure that all imported tables are consistent when they have been linked to each other. The total number of recorded trips in the Journey table was 1,538,673 records. The processing to link the trips to the other attributes for individuals and households retained all this information.

6.2.4 The other check was to apply the set of weights to make sure that the same numbers of trips were obtained as are published in the National Travel Survey: 2006 (DfT, 2007) in its Table 1.1. This table is reproduced below as Table 6.2 and provides data on the sample numbers obtained in each survey year of the NTS from 1995 to 2006.

6.2.5 The totals calculated from the NTS data received did match the published data. It should be noted that the published total trips reproduced in Table 6.2 exclude the expansion factor for short walk trips which are only collected for one day (day 7) of the weekly survey. Thus the numbers shown are not the numbers that were subsequently used to calculate trip rates. Instead, the trip rate calculations also included the day to week expansion factor for the short walk trips.

6.2.6 For purposes of calculating trip rates, we focused on the **diary sample** as trips are only provided for those people/households who have filled their travel diary. This is done by dividing the total weighted trips by the total weighted number of individuals.

Table 6.2 Sample numbers on which analyses are based, years 1995 - 2006

Number/thousands

		1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Households	Unweighted diary sample	3,211	3,210	3,139	2,935	3,020	3,435	3,469	7,437	8,258	8,122	8,430	8,297
	Unweighted interview sample	3,491	3,505	3,465	3,300	3,376	3,783	3,760	8,849	9,196	8,991	9,453	9,261
	Weighted diary sample	3,209	3,211	3,139	2,938	3,018	3,431	3,472	7,437	8,258	8,122	8,430	8,297
	Weighted interview sample	3,490	3,506	3,463	3,303	3,374	3,782	3,761	8,849	9,196	8,991	9,453	9,261
Individuals	Unweighted diary sample	7,723	7,665	7,473	6,842	6,970	8,056	7,978	16,886	19,467	19,199	19,904	19,490
	Unweighted interview sample	8,521	8,504	8,452	7,945	8,000	9,054	8,833	20,827	21,990	21,588	22,702	22,141
	Weighted diary sample	7,675	7,704	7,487	6,986	7,109	8,114	8,132	17,494	19,578	19,302	20,103	19,794
	Weighted interview sample	8,348	8,405	8,258	7,857	7,945	8,942	8,811	20,789	21,795	21,369	22,539	22,098
Children (<16 yrs)	Unweighted diary sample	1,744	1,666	1,650	1,433	1,466	1,731	1,659	3,413	4,178	4,129	4,150	3,938
	Unweighted interview sample	1,945	1,869	1,917	1,708	1,703	1,971	1,860	4,398	4,702	4,690	4,776	4,497
	Weighted diary sample	1,602	1,634	1,547	1,447	1,467	1,666	1,624	3,523	3,900	3,815	3,963	3,848
	Weighted interview sample	1,741	1,782	1,704	1,626	1,631	1,835	1,764	4,181	4,336	4,222	4,443	4,296
Adults (16+)	Unweighted diary sample	5,979	5,999	5,823	5,409	5,504	6,325	6,319	13,473	15,289	15,070	15,754	15,552
	Unweighted interview sample	6,576	6,635	6,535	6,237	6,297	7,083	6,973	16,429	17,288	16,898	17,926	17,644
	Weighted diary sample	6,074	6,069	5,940	5,539	5,640	6,447	6,508	13,971	15,678	15,487	16,142	15,945
	Weighted interview sample	6,606	6,623	6,553	6,231	6,315	7,107	7,047	16,608	17,459	17,147	18,097	17,801
Motor vehicles	Unweighted diary sample	3,296	3,301	3,238	3,121	3,217	3,772	3,707	8,195	9,264	9,065	9,847	9,758
	Unweighted interview sample	3,642	3,691	3,653	3,608	3,681	4,240	4,081	9,954	10,452	10,190	11,228	11,118
	Weighted diary sample	3,339	3,374	3,313	3,201	3,270	3,843	3,780	8,391	9,408	9,261	10,059	9,875
	Weighted interview sample	3,629	3,694	3,663	3,601	3,664	4,237	4,090	9,959	10,465	10,270	11,264	11,028
Trips	Unweighted diary sample	126,088	124,748	122,397	112,867	114,501	130,179	129,998	278,916	314,845	310,065	322,500	312,347
	Unweighted interview sample	-	-	-	-	-	-	-	-	-	-	-	-
	Weighted diary sample	129,356	133,896	127,242	120,996	123,182	137,689	139,240	302,796	333,833	326,869	345,997	336,802
	Weighted interview sample	-	-	-	-	-	-	-	-	-	-	-	-
Stages	Unweighted diary sample	131,548	129,690	127,273	117,269	119,072	136,324	134,036	289,048	327,230	322,602	335,940	326,076
	Unweighted interview sample	-	-	-	-	-	-	-	-	-	-	-	-
	Weighted diary sample	135,017	139,423	132,494	125,838	128,346	144,406	143,953	314,728	348,024	341,321	361,449	352,392
	Weighted interview sample	-	-	-	-	-	-	-	-	-	-	-	-
Great Britain demographic data for survey periods:													
Population ('000s)		56,279	56,381	56,496	56,627	56,802	56,960	57,149	57,625	57,851	58,125	58,485	58,846
Grossing up factors		7,009	7,356	7,560	8,276	8,150	7,071	7,163	3,413	2,972	3,027	2,938	3,019

Source: National Travel Survey: 2006 (DfT, 2007) Table 1.1

6.2.7 As the diary sample was used to extract the trip rates, all other variables were also extracted based on the diary sample and not the interview sample. Therefore, the NTS weight W2 rather than W3 was used in all cases.

6.3 NTS EVIDENCE ON THE SURVEY RESPONSE RATES ACHIEVED

6.3.1 It may or may not be coincidence that:

- In both the 1970s and in the current decade when the NTS response rates were in the low 60% range the trip rates measured in the survey were considerably lower than in the early 1990s, a period when the NTS response rates were around the 80% mark;
- The lowest trip rates in the NTS have been observed for residents of Inner London who have generally had the lowest survey response rates.

6.3.2 We seek here to determine how much of the change in the trip rates, as measured in the NTS, is a genuine behavioural phenomenon describing the travel behaviour of the population; and how much is a side effect of methodological issues related to the collection and processing of the NTS data.

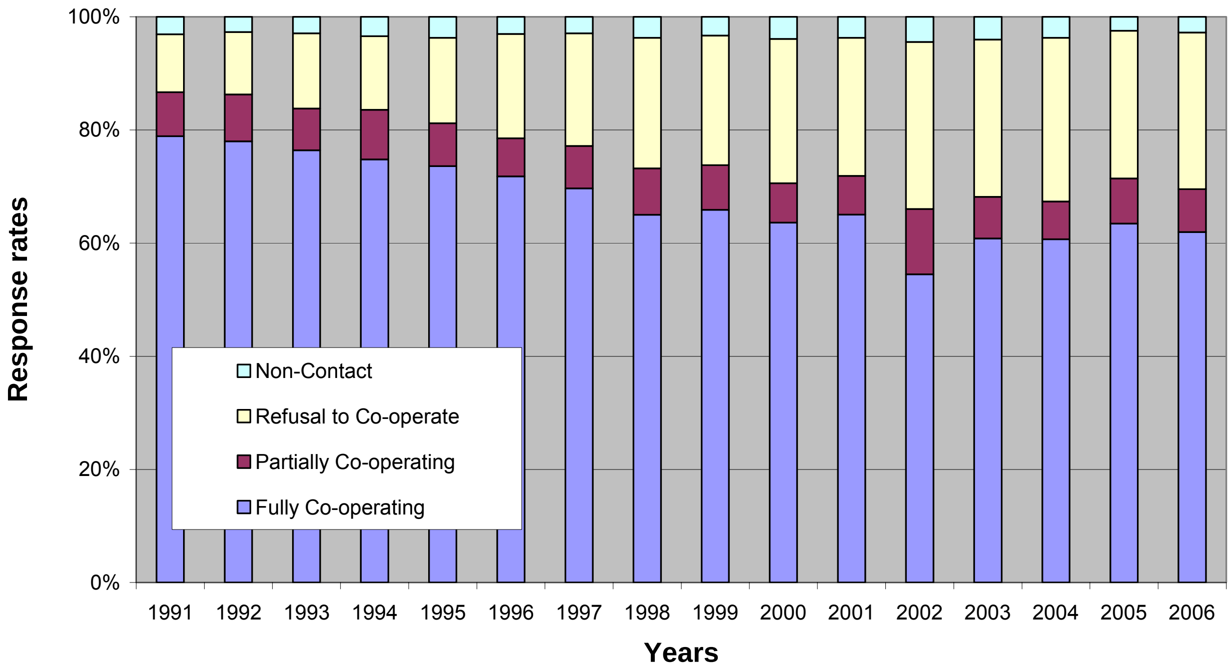

Figure 6.1 Response rates for NTS by year

6.3.3 The NTS methodological documentation records that the response rate for Great Britain for fully co-operating households has declined significantly over recent years from 79% in 1991 to a much lower 61% in 2006 as illustrated in Figure 6.1. The lowest response rate obtained was the 54% in 2002, the year in which the organisation carrying out the survey changed. In areas such as Inner London the current response rate of 49% is much lower again than the average for Great Britain as a whole. A much lower than average response rate in London has occurred consistently over the years.

6.3.4 These lower response rates need careful consideration, in case they are a source of bias among those who have completed the survey towards persons who spend more time at home and less time travelling. The NTS team have already carried out research into the nature of the lower response rates and have introduced a weighting of surveyed individuals in order to off-set some of the bias that can potentially be caused by these lower response rates.

6.3.5 To understand the extent to which the lower response rates might be one of the causes of the reduction in trip rates, we have carried out regressions of annual trips per person against the percentage in each year of fully co-operating respondents in the sample, as illustrated in Figure 6.2. Here two regressions are shown: the dark blue diamonds and red line represent values from the shorter series from 1988/90 onwards in which the data collection method is more consistent over time than in earlier years; the green squares are from NTS surveys prior to 1988 and are included together with the later years in the dotted line regression.

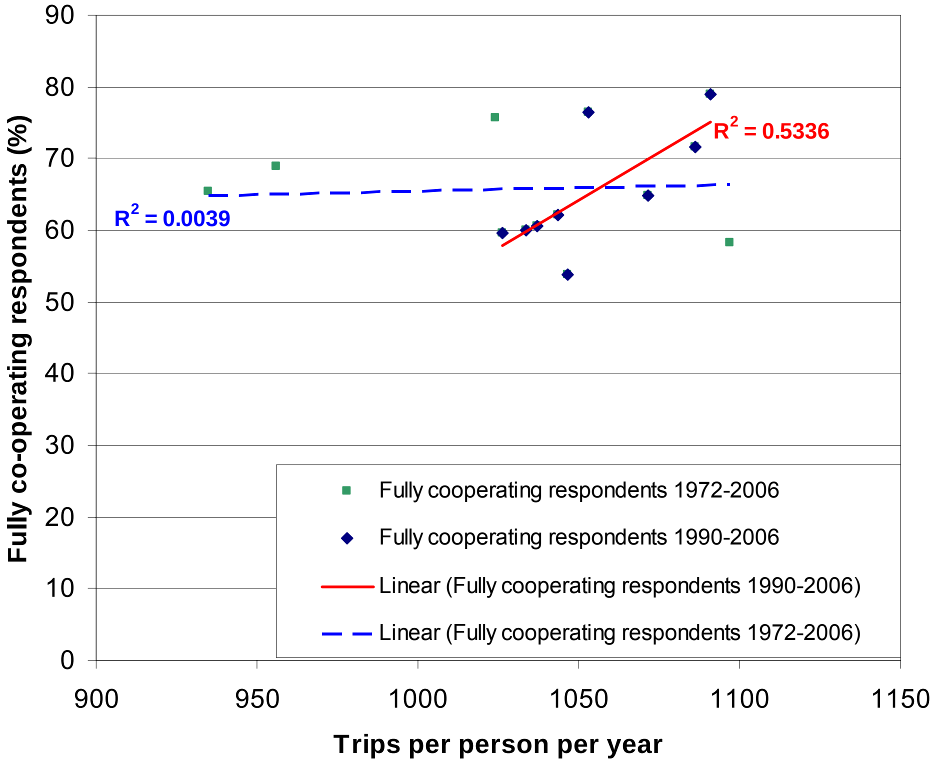

Figure 6.2 Regression of percentage of fully co-operating respondents versus trip rates

6.3.6 In this regression the R^2 value of 53% appears to provide clear statistical evidence of a relationship since 1990 between the annual trip rate per person and the percentage of fully co-operating households, though this does not of course guarantee that this is actually a causal relationship. Two other categories of Figure 6.1: partially co-operating and not contactable respondents were quite small throughout the years and showed no strong trends either through time or of a correlation with trip rates.

6.3.7 The results of this regression are consistent with the hypothesis that:

> *in periods where the NTS can only achieve lower response rates in the survey, the trip rates that it measures are biased downwards.*

6.4 OTHER RESEARCH ON SURVEY NON-RESPONSE AND TRIP RATE STABILITY

6.4.1 Strong support for the previous hypothesis is provided by evidence obtained in a research study into non-response in the 2001 LATS survey that was carried out by Polak (2002) for Transport for London. Following on from the initial LATS Pilot Household Survey, a survey of its non-responding households was initiated in order to ascertain whether and how these households might differ in behaviour from these who had responded fully to the initial pilot survey. The research findings on non-response were summarised as:

> "The nature of the non-response mechanism is in general such as to reduce the representation of high mobility households in the final sample. Simple tests using the sample data collected in the pilot survey suggest that the biases in the estimation of mobility that are introduced in this way can be highly significant. …
>
> This significance is twofold. First, since highly mobile households are less likely to participate in the main survey than less mobile ones, estimates of mobility based on data from the main survey will be biased downwards. Second, ex-post treatments such as re-weighting that are based solely on demographics are unlikely to adequately account for these effects."

6.4.2 Although these findings were based on the LATS household survey, they are very likely to carry across directly to the NTS because both surveys are of comparable complexity in terms of the scale of effort required to complete them and both cover similar subject matter in terms of requesting detailed information on travel patterns for all of the members of the surveyed household.

6.4.3 Recent analysis of trip rates in London (TfL, 2008) comparing data from the LATS 2001 and the current LTDS household surveys has shown that the average number of trips of 2.8 per weekday per Londoner was almost unchanged between 2001 and 2006/07. Unpublished household surveys in the city of Dublin have also indicated no change in trip rates in recent years. Accordingly, there is support from other recent independent surveys for the hypothesis of no trend through time in trip rates.

6.5 LONG TERM TRENDS IN NTS TRAVEL PATTERNS

6.5.1 With a view to understanding better the relationship between trends in different aspects of travel we have also examined the overall trends through time. Figure 6.3 illustrates that the long term trend since the 1970s of ever increasing annual travel distance per person had stabilised around 1999, with no consistent increase since then. This stability coincides with the start of the apparent recent minor reduction in trip rates.

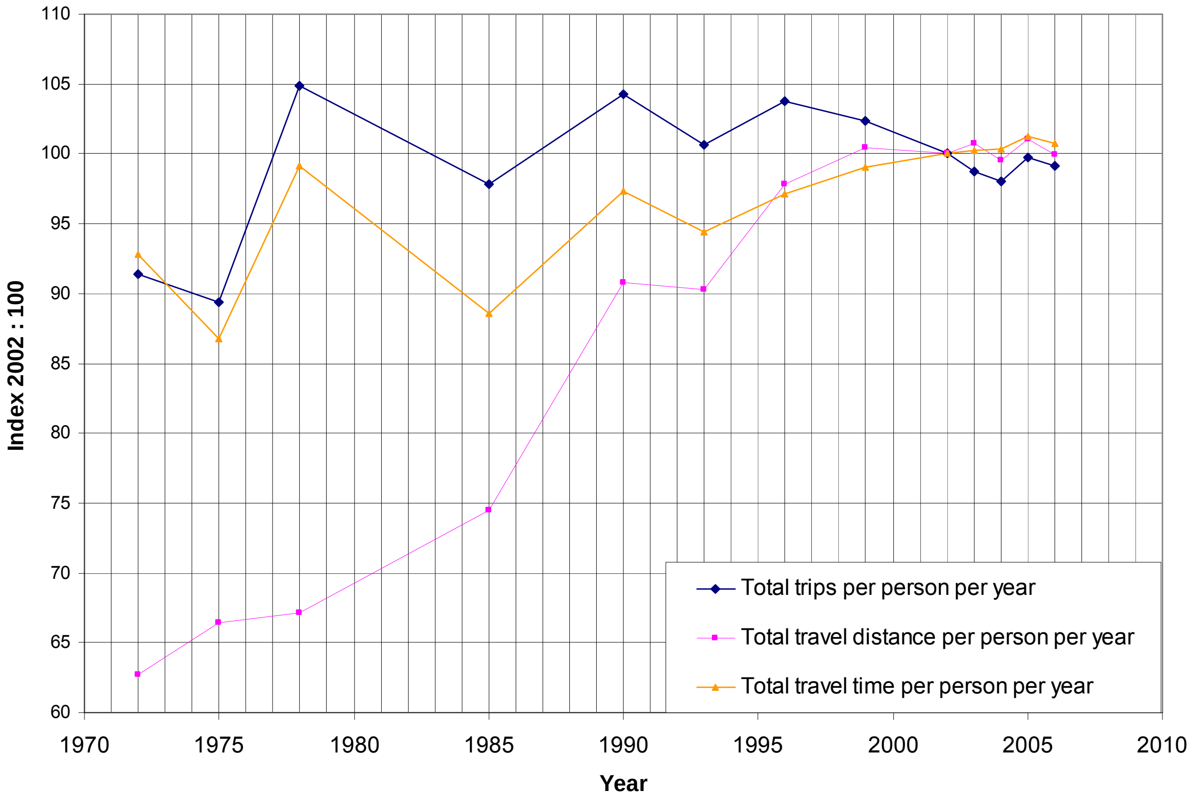

Figure 6.3 Annual Index of trips, distance travelled, travel time per person (1972-2006)

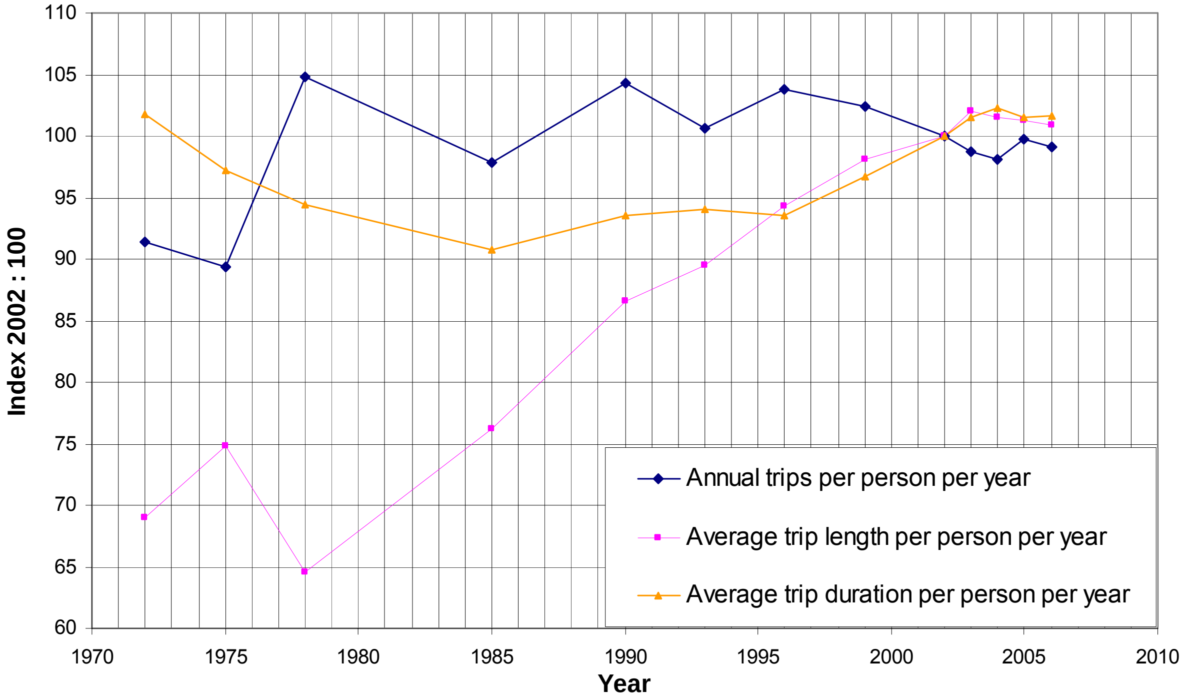

Figure 6.4 Annual index of trips per person, average trip length and average travel time per trip (1972-2006)

6.5.2 The same basic travel data is presented in a different form in Figure 6.4, which illustrates that despite the major increases in average trip lengths, the average number of trips per annum and the average travel time per trip were both fairly constant from the late 1970s through to the mid 1990s, allowing for sampling errors and possible differences in data collection methodology. It also appears to suggest that from the mid-1990s onwards, increases in average trip lengths started to imply increases in average travel duration. Perhaps it is this termination of the previous continuing trend to faster travel speeds that has encouraged the end of the trend to increasing trip length and the start of a trend to fewer trips? Alternatively, it may be a reflection that the reduction over time in response rates was leading to a reduction in the proportion of the most mobile individuals who were captured in the survey, which in turn has biased downwards all measures of travel intensity.

6.6 RECENT TEMPORAL TRENDS IN TRIP RATES BY TRAVEL PURPOSE

6.6.1 Later in Chapter 8, we will investigate the hypothesis of temporal stability in trip rates in more detail for specific trip purposes. Here we carry out some simple summary analysis.

6.6.2 Statistical estimations were carried out to examine temporal trends from 2002 to 2006 in overall trip rates for individual home-based trip purposes. Seven trip categories were examined, work, employer's business, education, shopping, recreation, holiday and visiting friends. For each travel purpose test, two approaches were adopted to analyse the influence of different survey years (NTS variable "H96") on the observed trip rate per capita. One was to test the significance of the model estimate based on inputting survey year as a continuous variable; the alternative was to test the significance of introducing it as a set of year specific dummy variables. In the latter case, year 2002 was set as the base (zero value) of the dummy variables.

6.6.3 The negative binomial model has been used to estimate the influence of the survey year variable on its own on the weekly trips per capita as presented in the seven sub-tables of Table 6.3 below for each home-based travel purpose. The third column contains the parameter estimate, while the remaining columns contain respectively: the standard error of this estimate; its normalised score; the probability that this parameter estimate is not significantly different from zero; and the classification of its statistical significance.

Table 6.3 Tests of the temporal trends from 2002 to 2006 in trip rates per person by travel purpose (using weighted NB regression)

Purpose	**Work**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	-0.020778	0.0120982	-1.7174	0.08590	
	2004	-0.0206567	0.0122235	-1.6899	0.09104	
	2005	-0.0482788	0.0122112	-3.9537	7.697e-05	***
	2006	-0.0561299	0.0122058	-4.5986	4.253e-06	***
Continuous	H96	-0.0139572	0.0027118	-5.1468	2.649e-07	***

Significance levels: * 95%; ** 99%; *** 99.9%

Purpose	**Employer's Business**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	-0.021450	0.053280	-0.4026	0.687242	
	2004	0.140557	0.052640	2.6702	0.007581	**
	2005	0.165142	0.051535	3.2045	0.001353	**
	2006	0.167031	0.051177	3.2638	0.001099	**
Continuous	H96	0.052136	0.011212	4.6499	3.321e-06	***

Purpose	**Education**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	0.0676756	0.0298880	2.2643	0.02356	*
	2004	0.0185822	0.0301612	0.616	0.53783	
	2005	0.0086124	0.0300897	0.2862	0.77471	
	2006	-0.0333925	0.0303871	-1.0989	0.27181	
Continuous	H96	-0.013618	0.006536	-2.0835	0.03720	*

Purpose	**Shopping**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	-0.011724	0.0146410	-0.800	0.4232	
	2004	0.00070538	0.0147883	0.0477	0.9619	
	2005	-0.022255	0.0147089	-1.5130	0.1302	
	2006	0.0261675	0.0146949	1.7807	0.07496	
Continuous	H96	0.0044449	0.0032827	1.354	0.1757	

Purpose	**Recreation**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	-0.003349	0.0188908	-0.1773	0.85927	
	2004	0.014113	0.0189564	0.7445	0.45657	
	2005	0.029073	0.0187470	1.5508	0.12095	
	2006	0.053070	0.0190011	2.7930	0.00522	**
Continuous	H96	0.014080	0.0042216	3.3351	0.00085	***

Purpose	**Holiday**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	0.054546	0.032289	1.6893	0.09116	
	2004	0.061596	0.032488	1.8960	0.05796	

	2005	0.192870	0.031750	6.0747	1.242e-09	***
	2006	0.206788	0.033132	6.2413	4.339e-10	***
Continuous	H96	0.055599	0.007208	7.7135	1.224e-14	***

Purpose	**Visit Friends**	Estimate	Std. Error	Z value	Pr(>|z|)	Significance
Dummy Variables	2003	0.016835	0.019391	0.8682	0.3853	
	2004	0.000787	0.019730	0.0399	0.9682	
	2005	0.016423	0.019327	0.8497	0.3955	
	2006	-0.019959	0.019579	-1.0194	0.3080	
Continuous	H96	-0.004229	0.004288	-0.9861	0.3241	

Significance levels: * 95%; ** 99%; *** 99.9%

6.6.4 Analysing the results in these tables demonstrates that when survey year is introduced as a continuous variable:

- for the purposes work and education, trip rates have a mild declining trend that is only significant at the 95% level but not the 99% level;
- in contrast, for the purposes employer's business and holiday, trip rates have an increasing trend that is significant at the 99.9% level;
- the remaining three travel purposes exhibit no significant overall trend in trip rates per capita over the five years.

6.6.5 Further examination of the dummy variable estimates for each of the individual years does not suggest any pronounced consistent trend through time for travel purposes, except for the increasing trend in trip rates for employer's business and holidays.

6.6.6 *These initial simple regression results do not suggest that there is a pronounced trend in recent years to reductions in trip production rates per capita*.

6.6.7 Further more detailed analyses were carried out in this study to evaluate the robustness of this viewpoint, through:

- use of a longer time-series from 1995 through to 2006 which provided a greater number of data points at which to see whether or not there is a pronounced trend;
- introducing behavioural variables that explain much of the variation in trip rates between types of individuals. Such influences could potentially currently be off-setting genuine trends through time in trip rates, and so in summary analyses such as those reported in Table 6.3, such trends could be masked.

The findings from analysing this 1995-2006 NTS data are reported in Section 8.

7 NTEM and Current Data Definitions and Approach

7.1 REVIEW OF DEFINITIONS IN USE IN THE NTEM MODEL

7.1.1 Before discussing the current trip estimation approach and findings, it is essential to have an understanding of both the definitions and assumptions that were used in the original NTEM and of how these match to the approach now taken in our current analysis when extracting the data from the NTS database. In most instances, the same approach continues to be used now when defining travel purposes and specifying categories for variables in our analysis. This Chapter aims to provide details on the assumptions used in the original NTEM and to identify wherever a different approach has now been implemented in our current analysis.

7.1.2 The eight home-based purposes explicitly distinguished in the original NTEM model are shown in the second column of Table 7.1, together with the detailed NTS travel purpose categories that have been used to create them.

Table 7.1 Composition of home-based travel purposes of NTEM

No.	NTEM purpose	Code	Composition by NTS purpose
1	HB Work	(HBW)	Work, Escort to work
2	HB Employer's business	(HBEB)	In course of work, Escort in course of work
3	HB Education	(HBEd)	Education, Escort Education
4	HB Personal business	(HBPB)	PB medical, PB eat/drink (only available from 1995), PB other
5	HB Shopping	(HBShop)	Shopping, Escort PB/shopping
6	HB Recreation	(HBRec)	Eat/drink with friends Other social Entertainment Sport Day trip - just walk trips Other non-escort Other escort
7	HB Holiday / Day trip	(HBHol)	Holiday Day trip (excluding just walk trips)
8	HB visit friends	(HBVF)	Visit friends Escort home

7.1.3 The comparison between Table 7.1 and the set of definitions of main trip purposes that was sent to us by the NTS team shows (with the exception of other work - code 2 which is a tiny number of trips) that all trip purposes within the current NTS database are included in the original NTEM trip purposes presented in Table 7.1.

7.1.4 It had been decided in the original NTEM to incorporate the escort trips into the same purpose categories as those of the primary trips being escorted. The reason was because the location of the trip-end, the time of day and the mode selected would all be the same for both. This combined definition of primary purpose and its escort purpose was then used for the estimation of trip productions as well as of trip attractions. For our research, which is focused on home-based trip productions, however, the factors affecting the rate of generation of primary home-based trip productions might not necessarily be the same as those affecting the rate of generation of any related escort trips. Escort trip characteristics are strongly influenced by the characteristics of the person who is being escorted rather than / as well as by those of the person who is escorting.

7.1.5 Generating escort trips as a function of the number of primary trips and of those who are making the primary trips is one of the areas with potential for improvements/modifications to the NTEM model. This approach is examined in more detail in Chapter 13. In the other Chapters of this Report we focus on home-based primary trips (i.e. that exclude escort trips), but this temporary exclusion of the escort trips from our estimation must be taken into account when comparing the trip production rate estimates below with those in the original NTEM.

7.1.6 Within NTEM the non-home-based trips made by those originating from a previous trip for purpose visiting friends/relatives are included within the home-based trip category because the spatial pattern of origins will tend to coincide with residential areas. We continue to take the same approach for extracting home-based trips from the NTS dataset and for analysis.

7.1.7 When extracting data, trips by all vehicle types which have been reported in NTS were included. The vehicles which are included within the NTS survey are presented in Table 7.2.

Table 7.2 Transport modes in NTS dataset

Vehicle Types
Walk, less than 1 mile; and Walk, 1 mile or more
Bicycle
Private (hire) bus
Private car – driver and passenger
Motorcycle/scooter/moped – driver and passenger
Van/lorry – driver and passenger
Other private transport
Stage, express and tour buses
LT underground, surface and light rail
Air
Taxi
Minibus
Other public transport

WEIGHTING FACTORS WITHIN THE NTS

7.1.8 An important difference between the original NTEM analysis and the approach adopted in the current study lies in the manner in which the NTS now makes use of weighting factors to help to offset differential response rates to the survey.

7.1.9 A weighting strategy for the NTS was developed following a recommendation in the 2000 National Statistics Quality Review of the NTS. The NTS results for 2005 were based on weighted data for the first time. The weighting methodology was then applied to data back to 1995 so that all NTS figures for 1995 onwards which have recently been published or released are now based on weighted data. As well as adjusting for non-response through a household weight, the weighting strategy for the NTS also uses a trip weight to adjust:

- for the drop-off in the number of trips recorded by respondents during the course of the travel week;
- for uneven recording of short walks by day of the week; and
- for the short-fall in reporting of long distance trips.

7.1.10 The relevant two types of weights used in the analysis are:

- Household/Individual weights (i.e. W2: the diary sample household weight within the NTS dataset); and
- Trip weights (i.e. W5xhh which excludes the household/individual weight).

7.1.11 As recommended in NTS documentation[4], the preferred calculation of the weekly trips or trip rate is to use the NTS Diary sample as follows:

- the trips in each record should be multiplied by their trip weight W5xhh and by the individual weight W2; then summed over the week for all relevant individuals;
- these total trips are then divided by the weighted sum of individuals, where each is weighted by the household weight W2.

7.1.12 At the start of the study, all travel purposes were examined by using a simple NB regression. In this approach all trips were scaled by both the trip and household weights and were then used as dependent variables in our analysis. However, it was discovered later in the study that the R statistical package provides an option to run *weighted* NB regression. This enabled the analysis to switch to a more precise approach for estimating trip rates in which the trips are weighted only by trip weights and the household weights are used instead to provide input weights within the weighted regression model. For the rest of this report, the former approach will be called "simple" and the latter "weighted" NB regression.

7.1.13 It was then decided to re-estimate using weighted NB regression all the main tests of various simple and more complex forms of models for all of the home-based travel purposes. Comparing the results from the simple and weighted regressions showed that, with the exception of area type, no difference is found in the significance level or sign of the coefficients for the main variables and the coefficients themselves do not differ greatly between the two approaches.

7.1.14 The reason why area type differential effects appear to have more influence in the simple regression than in the weighted regression is because area type is one of the main characteristics that is used in creating the household weights. The response rates to the survey are much lower than the average in areas such as Inner and Outer London so that the household weights are scaled up for the successful survey responses from such areas in order to make the results overall be representative of the whole population of Great Britain.

7.1.15 Series of calls trips were not included when extracting trips as they are given the trip weight of 0 within NTS dataset and its publications. The short walk trips on the other hand (which are collected only for 1 of the 7 days of the survey) are given trip weights of 7 to represent all 7 days of the week. The NTS user guide provides the following explanation:

> 7.1.16 "…Because trips of less than one mile in distance are recorded only on the seventh day of the travel week, these trips must be weighted by a factor of seven when analysed. Also for consistency with earlier surveys 'series of calls' trips are excluded from analysis of stage and trip counts and time."

[4] See page 4 " National Travel Survey (NTS) data: User guidance" supplied by the NTS team

7.2 MAIN EXPLANATORY VARIABLES FOR TRIP RATES USED IN NTEM

7.2.1 We now review the main explanatory variables that had been used in the original NTEM to represent the variations in trip production rates for individual travel purposes. These entered as segmentation variables to determine the cross-categories for which separate rates were produced.

7.2.2 **Household car availability** mainly affects the estimation of the choice of mode. The inclusion of **socio-economic group** added little improvement to the trip rates model and the inclusion of **area type** mainly led to a differentiation between Inner London and rest of GB. Accordingly, in the original NTEM model the household type of a resident was segmented:

- by household size (1, 2 or 3+ adults);
- by car ownership (0, 1 or 2+ cars); and
- by area type into Inner London and the rest of Great Britain.

7.2.3 The definition of person type within a household type that was used in the original NTEM is presented in Table 7.3. Children are defined as those with age 0 to 15, which is in line with our current definitions to be consistent with the many Census tabulations used to extract population data for the 16+ age group. However, the gender of children was not distinguished in NTEM as it had found that gender had little effect on trip rates for the majority of travel purposes.

Table 7.3 Person type definitions in use in NTEM

Person Type	NTEM Definition
1	children (0 to 15)
2	males in full time employment (16 to 64)
3	males in part time employment (16 to 64) NB the sample size for these in NTS is very small
4	male students (16 to 64)
5	male not employed / students (16 to 64) - Unemployed plus other Inactive
6	male 65+
7	females in full time employment (16 to 64)
8	females in part time employment (16 to 64)
9	female students (16 to 64)
10	female not employed / students (16 to 64) - Unemployed plus other Inactive
11	female 65+

7.2.4 However, within our improved approach that employs Negative Binomial regression analysis rather than categorical analysis, we are able to model the variable "age", which includes children as one of its categories. This is a separate variable from "gender" so we can evaluate the independent effect of each on changes in trip rates. Findings for some purposes such as shopping suggest that male children make significantly fewer trips than female children so segmenting children by gender can make a noticeable improvement in predicting trip rates.

7.2.5 Both the variable "age" and "work status" in the NTS dataset include "children" as one of their categories. This makes it technically impossible (due to full inter-correlation) to model them as two separate variables in one model. We therefore combined "work status" and "age" to define one joint variable "Age_Workstatus" to be used in all the models that were estimated.

7.2.6 In the current study, “area type” is segmented into 6 categories so we can consider the relative importance of each individual type on trip rates. This is an improvement compared to the original NTEM model in which the distinction is just between Inner London and the rest of Great Britain. The definitions used for area types are provided in Table 7.4.

Table 7.4 Area type definitions in use in the research

Area Type	Definition
Inner London	London area within north-south circular
Outer London	London area outside north-south circular
Metropolitan Area	Metropolitan areas such as the city of Manchester, Liverpool etc
Urban Big	Urban area with the population of over 250k
Urban Medium	Urban area with the population of 25k to 250k
Urban Small	Urban area with the population of 10K to 25K
Rural	Rural area with the population of less than 10K

8 Analysis of Changes over Time in Trip Rates

8.1 INTRODUCTION

8.1.1 Apparent changes in trip frequency from 2002 to 2006 have been examined in detail both by modelling the variable "year" as the only variable in the model (see section 6.6) and in combination with other factors (i.e. income, accessibility etc). No strong evidence suggesting a behavioural change over time has been found from these analyses.

8.1.2 In order to have a fuller picture of changes in trip rates over time, it was decided to combine the 1995-2001 NTS dataset with the 2002-2006 one in order to analyse the most common three travel purposes: "shopping", "commuting" and "leisure" in general detail.

8.1.3 This would require a version of the 1995-2001 NTS dataset that is fully consistent with the 2002 to 2006 dataset. However, the 1995-2001 data that WSP received did not contain all of the variables that were in 2002-2006 dataset, while some of the other variables included were not available in a consistent form across the period as a whole. It has therefore been decided to include in the estimation presented in this Chapter only those variables that are fully consistent across the full period from 1995 to 2006. Table 8.1 lists the set of years that were included together with all the variables that were used for analysing changes in trip rates for each of the three purposes. These are the variables which were used in the original NTEM and which were available in a form that is consistent between the 1995-2001 and 2002-2006 NTS datasets, for estimating trip rates for the specified trip purpose.

Table 8.1 Variables used for 1995-2006 models

Purpose	Period tested	Variables used
Shopping	1995-2006	Year Household size Car availability Gender Age
Commuting (FT,PT workers)	1998-2006	Year Work status Household size Car availability Gender Age
Leisure	1995-2006	Year

8.1.4 "Work status" in the 1995-2001 NTS dataset was given under two headings as "1998 onward" and "1995-1997". The definitions of categories in "1998 onwards" are consistent with those in the 2002-2006 dataset. "Work status" was not considered for modelling "shopping" and "leisure" purposes so that we can model the longer period from 1995 to 2006. For commuting, however, it is essential to consider "work status" to be able to model the commuting trip rates of workers, rather than rates for the population as a whole.

8.1.5 The analysis of changes in trip rates over time has been carried out in two steps for each of the three travel purposes:

- Modelling the variable "year" as the only variable in the model both in "continuous" and in "categorical" fashion.
- If modelling the year alone in any of these two forms shows it has a statistically significant role, then "year" was modelled with all other relevant variables to see how important it is when all other factors are taken into account.

8.1.6 Year when included as the only variable did not show any significant role in explaining changes in "Leisure" trip rates.

8.1.7 When year was modelled as one continuous variable, it was statistically significant for both the commuting and shopping travel purposes. When modelling year as a categorical variable, we found that there were significant reductions in trip rates compared to the base year (1995), particularly from 2002 onwards. Therefore, it was decided to create fuller models of "commuting" and of "shopping" that included other relevant behavioural variables together with the parameter "Year" as a categorical variable.

8.2 VISUALISING AND INTERPRETING RESULTS FROM NBR

8.2.1 Before discussing these results further, it is helpful to digress in this Section to provide an overview of how to visualise and interpret the results from the NBR which are output by the R statistical package.

8.2.2 The dependent variable in the regression is the weekly number of trips per person for the trip category and person type of interest. The explanatory variables are the set mainly of categorical variables plus a few continuous variables that each are expected to influence these trip rates.

8.2.3 The R package outputs detailed information on the estimated parameters within a tabular presentation that is not particularly easy to assimilate or to compare between competing model formulations. Accordingly WSP have experimented in order to devise and implement an automated visualisation procedure that makes it easier to compare and interpret the relative importance of each behavioural factor in explaining observed differences in trip rates.

8.2.4 Figure 8.1 and Figure 8.2 demonstrate the resulting NB regression charts that present the estimated relative influence of the year and of various categorical variables on trip rates.

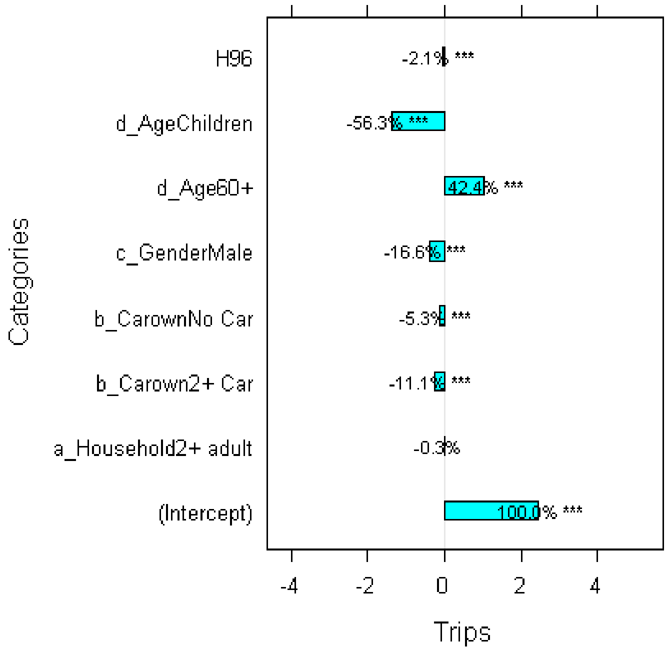

Figure 8.1 Simple model of shopping trips based on 95-06 NTS data – year (H96) modelled as a continuous variable (weighted NB Regression)

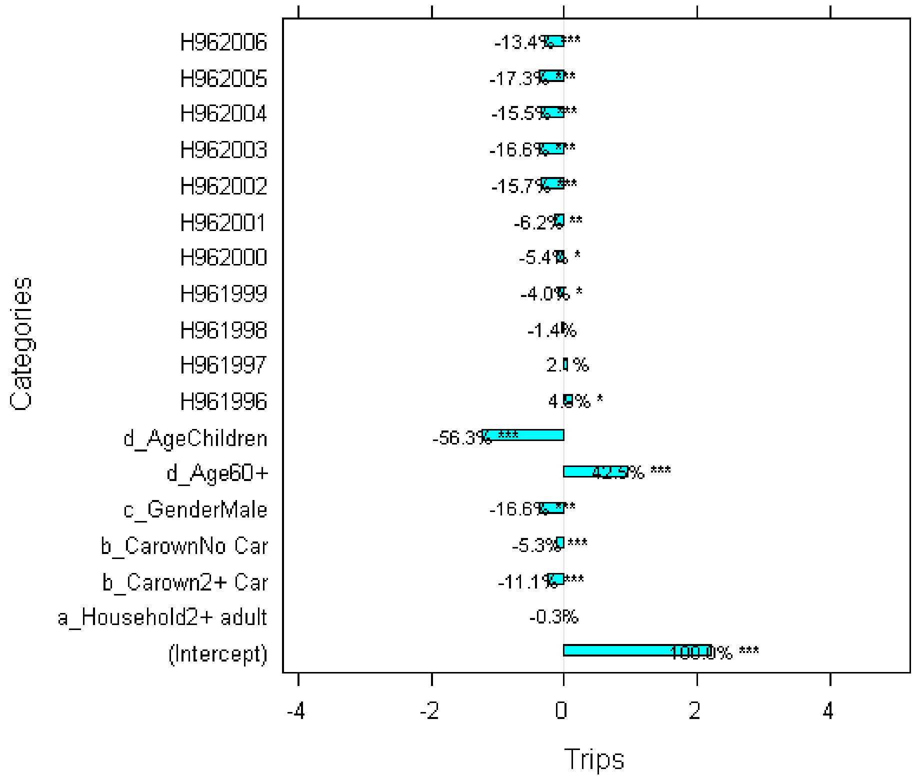

Figure 8.2 Simple model of shopping trips based on 95-06 NTS data - year (H96) modelled as a categorical variable (weighted NB regression)

8.2.5 These charts are automatically produced via a procedure that WSP have coded into R. Each chart presents two components: percentage change and significance code, for every continuous and categorical variable in the model.

8.2.6 The percentage change component measures the relative impact of the specified category on the trip rate, compared to the impact of the omitted reference category for that variable. Every categorical variable has just one omitted reference category for which the coefficient is automatically set to zero and so is not show in the chart. Normally this reference category is selected to represent a very commonly occurring category. Some examples based on Figure 8.2 will explain the concepts. The estimated average shopping trip rate for a male is 16.6% lower, *ceteris paribus*, than that for a female (the reference category). Likewise, the individuals from households owning either no car or 2 cars make 5.3% and 11.1% fewer shopping trips respectively than those from a household which has 1 car (the reference category).

8.2.7 When a continuous or categorical variable is modelled, the significance code shows within which confidence interval (as listed in Table 8.2) the null hypothesis can be rejected. This null hypothesis is that the variable has a coefficient equal to zero and so has no impact on the trip rate or in other words is statistically insignificant. Because of the very large size of this NTS dataset, a variable may be found to have a statistically significant impact on the trip rate at the 95% level even when this coefficient may have an effect in practice that is numerically too small to be noticed. Accordingly, in general here a significant variable is likely to have a noticeable impact only when it is at a 99.9% or a 99% significance level.

Table 8.2 Significance codes

Code	Confidence interval
*	95%
**	99%
***	99.9%

8.2.8 Further details on how to interpret these NB regression results charts are provided in Appendix D.

8.3 FINDINGS FROM ANALYSING 1995 TO 2006 NTS DATA

8.3.1 Figure 8.1 and Figure 8.2 for the purpose shopping contrast the results from modelling the survey year as a continuous and as a categorical variable. Figure 8.1 indicates that the introduction of a continuous variable for year is significant; it has a negative sign that denotes a trend through time of reduction in trip rates. However, analysing year as a categorical variable in Figure 8.2 presents a fuller picture. From 1995 to 2000, year is significant at no more than the 95% confidence level if at all, and its magnitude implies at most a 5% reduction from the rate for the base category (i.e. 1995). The trip rate then shows a sudden drop in 2002 and from then through to 2006 it stays in the range of a 13% to 17% decrease in trips compared to year 1995. No consistent growth trend over time can be observed from 2002 to 2006..

8.3.2 One hypothesis that explains this pattern is that the reason for the apparent drop in trip rates does not relate primarily to pure behavioural changes over time but is influenced more by changes in the effectiveness in the NTS and in its survey response rates over the years. These findings are in line with what was discussed in Chapter 6. The gradual minor decline in trip rates from 1995 to 2001 may well be an indirect result of the continuing reduction in NTS survey response rates from 79% in 1991 down to 54% in 2002. The sudden drop in trip rates from 2002 onwards coincides with the change from 2001 to 2002 in the surveying organisation and in the methodology for collecting NTS data. This change may well have had a direct impact on the observations collected. Based on the evidence that we have collected we believe that it is these methodological changes that have led to the apparent reductions in trip rates through time more than any genuine behavioural change.

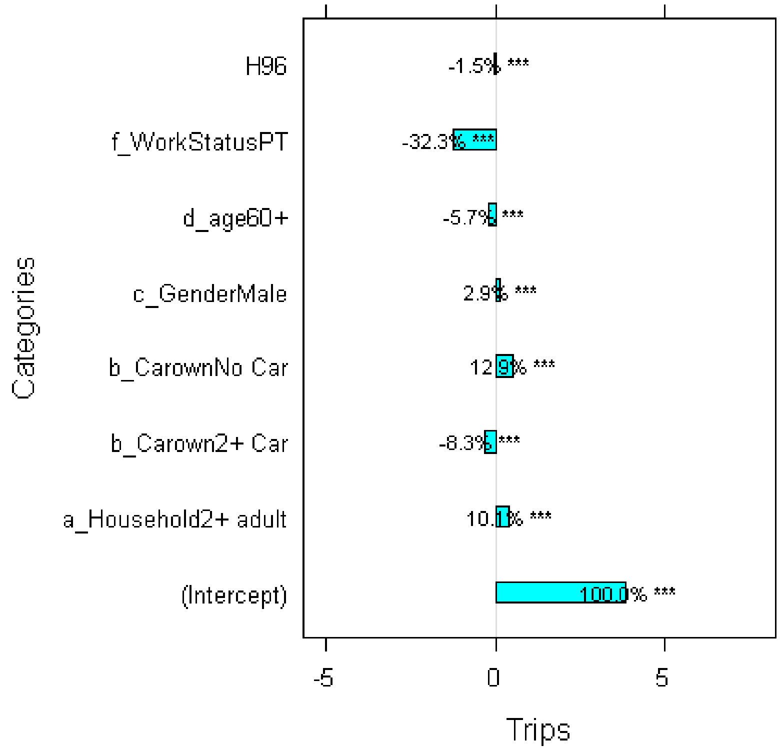

Figure 8.3 Simple model of commuting trips based on 95-06 NTS data - year (H96) modelled as continuous variable (weighted NB regression)

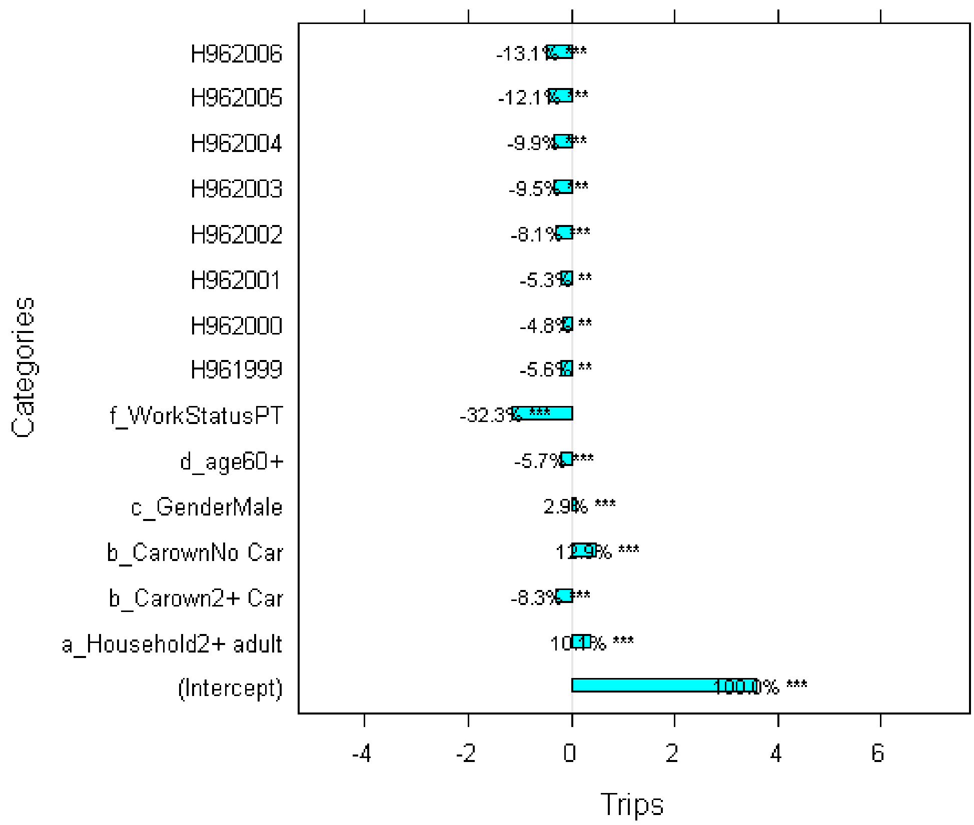

Figure 8.4 Simple model of commuting trips based on 95-06 NTS data - year (H96) modelled as categorical variable (weighted NB regression)

8.3.3 The assumption underlying this hypothesis is that those most likely to respond to home-based surveys, such as the NTS, are those people who spend the most time at home. Strong support for this hypothesis is provided by evidence obtained in a research study into non-response in the 2001 London Area Transport Survey (LATS) that was carried out by Polak (2002) for Transport for London. Following on from the initial LATS Pilot Household Survey, a survey of its non-responding households was initiated in order to ascertain whether and how these households might differ in behaviour from these who had responded fully to the initial pilot survey. The research findings on non-response were summarised as:

> "The nature of the non-response mechanism is in general such as to reduce the representation of high mobility households in the final sample. Simple tests using the sample data collected in the pilot survey suggest that the biases in the estimation of mobility that are introduced in this way can be highly significant. …
>
> This significance is twofold. First, since highly mobile households are less likely to participate in the main survey than less mobile ones, estimates of mobility based on data from the main survey will be biased downwards. Second, ex-post treatments such as re-weighting that are based solely on demographics are unlikely to adequately account for these effects."

8.3.4 Although these findings were based on the LATS household survey, they are very likely to carry across directly to the NTS because both surveys are of comparable complexity in terms of the scale of effort required to complete them and both cover similar subject matter in terms of requesting detailed information on travel patterns for all of the members of the surveyed household.

8.3.5 Recent analysis of trip rates in London (TfL, 2008) comparing data from the LATS 2001 and the current LTDS household surveys has shown that the average number of trips of 2.8 per weekday per Londoner was almost unchanged between 2001 and 2006/07. Unpublished household surveys in the city of Dublin have also indicated no change in trip rates in recent years. Accordingly, there is support from other recent independent surveys for the hypothesis of no trend through time in trip rates.

8.3.6 Figure 8.3 and Figure 8.4, analyse some of the influences on commuting trips, through use of a continuous and a categorical representation, respectively, of the year variable. They show broadly similar findings to those reported above for shopping trips. It can be observed that that although there is a 5.6% jump initially in 1999 from the implied reference case year 1998, there is subsequently little difference in the trip rates observed from 1999 through to 2001 but that the difference becomes pronounced after 2002 and exhibits a declining trend through from 2002 to 2006.

8.3.7 It is also of interest to note that the proportional strength and the significance of each of the other categorical variables remains almost unchanged for both shopping and commuting, irrespective of whether year is introduced as a continuous or as a categorical variable. This suggests that just testing the effect of one or other approach should be sufficient in later explorations of model form, for travel purposes where year is statistically significant.

8.3.8 There is a potential explanation for the particularly large reduction in shopping trips that occurred in 2002, the year in which the survey methodology changed. One important new feature introduced into the NTS survey from 2002 until 2007 was the requirement for respondents to record the postcode of the end-point of each trip. The NTS has always experienced 'drop-off', where respondents tend to report fewer trips towards the end of the travel week.

8.3.9 Accordingly, research was undertaken by the NTS team into the effect on response rates arising from the extra burden of this requirement to record postcodes. They found that the drop-off rates for day 7 (which is designed to fall uniformly on any day across the week) substantially increased from 2002 onwards, when post-coded data was first required on this day. This day 7 is the only day of the survey week for which the NTS requests that respondents record **all** short walk trips, whereas on the other six days of the survey only those walk trips longer than 500 metres are required. The recently introduced NTS weighting strategy makes an adjustment to account for drop-off. However, this adjustment for drop-off is not made for short walks, because they are only recorded on Day 7. Consequently, the requirement to record postcodes is likely to have a particular impact on walking data.

8.3.10 In practice, comparative studies of trends in the NTS data have shown that it is walk trips that have declined most since 2002 and it is known that short walk trips are particularly common for the trip purpose shopping.

CONCLUSIONS ON TRIP RATE STABILITY

8.3.11 The following conclusions can be drawn by combining the range of model based findings in this and subsequent Chapters, together with the methodology related findings from the earlier Chapter 6.

8.3.12 There is no convincing evidence of any substantial behavioural trend through time towards higher or lower trip rates per person for any travel purpose. Although summary analysis of the NTS trip rates appears to suggest a reducing trend in trip rates, more detailed analysis shows that this apparent trend is likely to be due to a mixture of other effects, including:

- Reductions in household response rates in the NTS, leading to a lower rate of representation of those making the most trips;
- Reductions in trips reported, especially short walk trips, in part due to the requirement to report the postcode of trip destinations;
- Changes in population profiles (e.g. a greater proportion of those in work being part-time workers).

8.3.13 The NTS data covering the years 1995-2006 has been reweighted by the NTS team to try to account for the influence of lower response rates and lower trip capture. However, the results from the research by Polak (2002) indicate that standard weighting using demographic characteristics is unlikely to be sufficient to off-set the downward bias in observed mobility arising from survey non-response.

8.3.14 The preferred approach when using the estimated models to forecast trip rates for a particular type of person is to set the year variable to zero (i.e. the base year 1995). This should then minimise the effects from any reductions in survey response rates and in trip drop-off. However, for a number of variables there was not a full consistent data series all the way back from 2006 to 1995, so that this approach is not always available as an option.

9 Analysis of Variables Used in the Current Version of NTEM

9.1 MAIN FINDINGS

9.1.1 This chapter describes the findings from estimating home-based trip rates by using the same variables as are used in the current NTEM model. Negative Binomial (NB) models were tested in two forms for each purpose:

- a weighted NBR estimation for a **simple model** which includes without interactions all variables used in the current NTEM;
- a weighted NBR analysis of a more **complex model** that takes into account interactions between some of the variables in addition to including each variable individually.

9.1.2 For each purpose, the inter-correlation between variables and the categories of each variable are carefully examined to ensure that the estimation of the model has not been biased. It is also used as a means to investigate the categories of each variable which can be combined to improve the goodness of fit of the model after ensuring that the differences are minor (not statistically significant) between the coefficients and standard errors of these categories. An example of a correlation matrix chart for the simple model for shopping is shown in Figure 9.1.

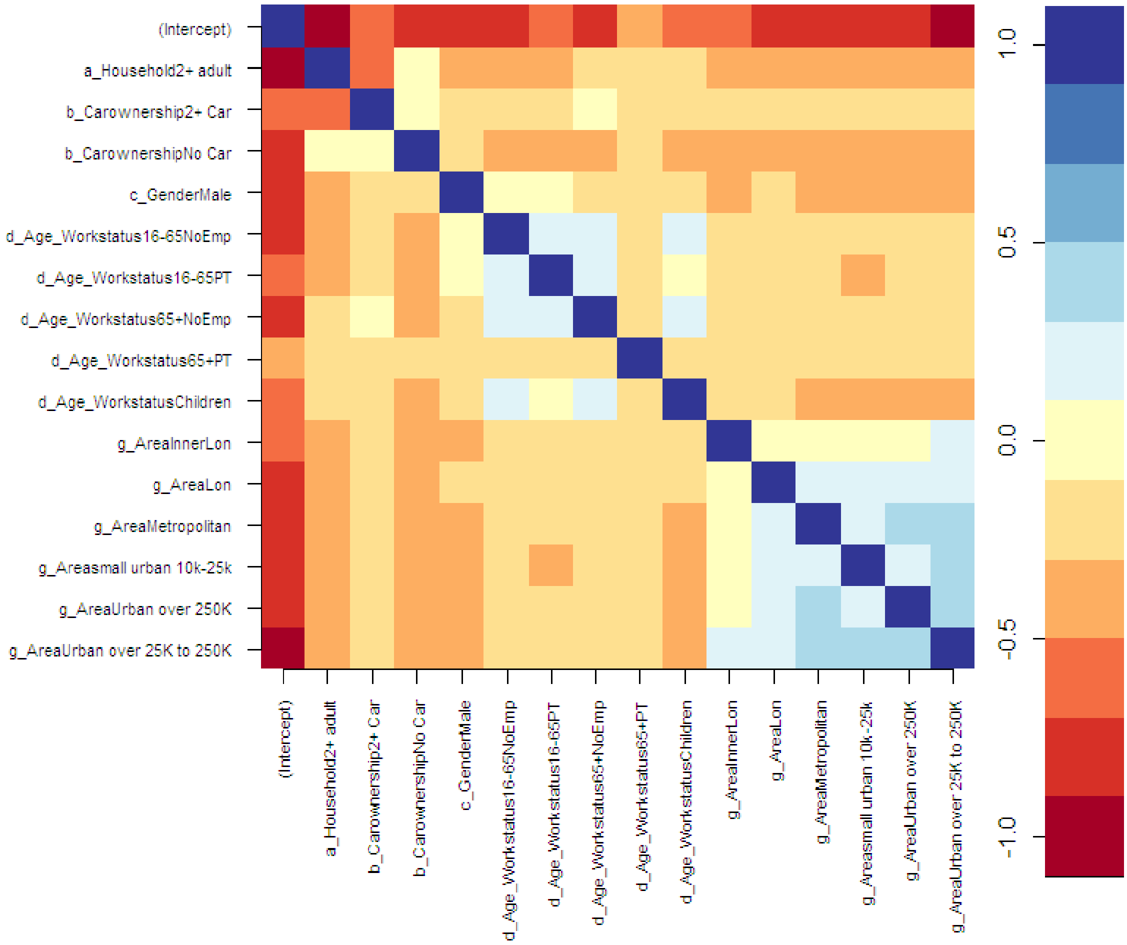

Figure 9.1 Example of correlation matrix chart: shopping - simple model

9.1.3 Commuting and business trips were modelled separately for workers (Full-time, FT and part-time, PT) and non-workers (student, pensioners etc). The reason is the latter group will include many people with no weekly working trips and so is very different in nature from the former.

9.1.4 The main results found from statistical analysis of the weekly number of home-based trips per person are provided in the next Section. Detailed purpose by purpose results are provided in Appendix E for the relative impact of each category of each variable tested and its statistical significance level.

FINDINGS FROM TESTING SIMPLE MODELS

9.1.5 The NTS survey year (between 2002 and 2006) was found *not* to be statistically significant at the 99.9% level in the estimation of the number of trips for each of the individual travel purposes examined, regardless of whether only the year variable alone was taken into account or whether other explanatory variables were also included.

9.1.6 All variables which are used in the current NTEM are significant and so are important in estimating trip rates for the majority of purposes. In some cases, however, to improve overall goodness of fit for a few variables we combined some of their categories if they had very similar coefficients and standard errors. Below we summarise some of the more interesting findings for various trip purposes from analysing the simple model.

- Segmentation by ***area type*** does not appear to have a significant effect for personal business and employer's business purposes. However, it appears to have an important role in explaining changes in trip rates for every other home-based purpose. This effect is not always shown to be similar for all purposes. For instance, people residing in Inner and then in Outer London make the least number of recreation and holiday trips, whereas the most are by those residing in rural areas. The number of trips for shopping, however, is equally small for London and for rural residents compared to all other area types. Overall, It is proved that trip rates vary noticeably across all 7 area types (i.e. Inner London, Rest of London, Small Urban, Medium Urban, Big Urban, Metropolitan and Rural) for the majority of purposes, rather than varying just between the two segments (i.e. Inner London and Rest of GB) which are currently in use within NTEM.
- More detailed segmentation by ***age groups*** (i.e. 5-11, 11-16, 16-18, 18-64 and 64+) based on individual's school levels was tested and proved to have a noticeable effect in improving the estimation of education trip rates.
- ***Gender*** proved not to be statistically significant in the models for Personal Business and Holiday purposes. but is an important factor to be considered for every other purpose

FINDINGS FROM TESTING THE MODELS AFTER INCLUDING INTERACTION TERMS

9.1.7 Important interaction terms were identified through examining several intermediate tests for each home-based trip purpose examined. These were developed to test full interactions between any two variables which were found to add some value to the model. At the next stage, the categories of these variables which were shown to have statistically significant interactions were kept and added as extra categorical variables within the simple model. The relative importance of each of the interactive variables was then examined and the AICs of the resulting models were compared to the equivalent simple ones to judge whether these interactive variables have a noticeable effect.

9.1.8 Interactions between some of the basic variables were proved to be statistically significant and that provides a means for improving the model for some of the travel purposes. However, the particular interaction terms which are significant and their relative importance vary across the purposes. For instance, some forms of interactions between age-work-status with car ownership, gender and area type appear to be significant for home-based shopping but no interaction variables add any value to the models of home-based work (HBW) or home-based employers business (HBEB) trips.

9.1.9 Table 9.1 gives the comparison between the AIC factors of the two forms of the model: simple one (i.e. with no interactions) and the final form of the more complex one (i.e. includes interactions). The purposes which are not shown in this table are those for which no statistically significant interaction terms were identified. The detailed purpose by purpose charts provided in Appendix E demonstrate the relative importance of each interactive variable in use.

Table 9.1 Comparing simple model versus model with interaction variables

Purpose	AIC - simple model	AIC - Model with interactions	AIC % difference	Important Interaction terms
Holiday	118366	118312	-0.045%	"0Car,16-65NotEmp" "0Car,65+NotEmp"
Recreation	294200	294056	-0.049%	"0Car,65+NotEmp" "2+Car,65+NotEmp" "2+Car,16-65NoEmp" "2+Car,Child" "0Car,Child" "0Car,Male"
Shopping	317432	316760	-0.21%	"65+NotEmp,NoCar" "Child,NoCar" "Student,NoCar" "16-65NotEmp,Male" "65+NotEmp,Male" "65+NoEmp,OutLon"

9.1.10 The models with interaction terms were proved to have a better goodness of fit for the three purposes of “Holiday”, “Recreation” and “Shopping”. However, apart from shopping, the two other purposes are just marginally improved by the inclusion of interaction variables.

9.1.11 In all cases, the most important interactions are proved to be between car ownership and work status, specifically for those not employed and students.

9.1.12 In order to have a better understanding of the influences that interactions between variables can have on the estimation of trip rates, the results for home-based shopping are examined as an example. Figure 9.2 shows the model of shopping trip rates where the interaction terms are included as the first six entries at the top of the chart. The main conclusions from analysing Figure 9.2 are the following.

- People over the age of 65 who are not employed (d_Age_Workstatus65+NoEmp) make 45.4% more shopping trips than the reference group (16-65 FT workers). People in households with no car (b_CarownNo Car) make 9.7% fewer trips than those in households with 1 car. However, in addition to these individual effects, people from the group: not employed, over 65 years age and with no car in their household (Int1_NoEmpONoCar), make 10.2% fewer trips than the average for all other groups combined. This additional effect can be captured either by including the interaction term (Int1) that identifies this specific group or alternatively by filtering out those not employed who reside in no car households to be examined in a separate model.

- Although children in general makes fewer shopping trips, children in households with no car makes more shopping trips than those in households with car.
- Although people living in Outer London in general make fewer shopping trips compared with other areas except Inner London , the unemployed people who are living in London show different behaviour and make more home-based shopping trips than the employed persons.

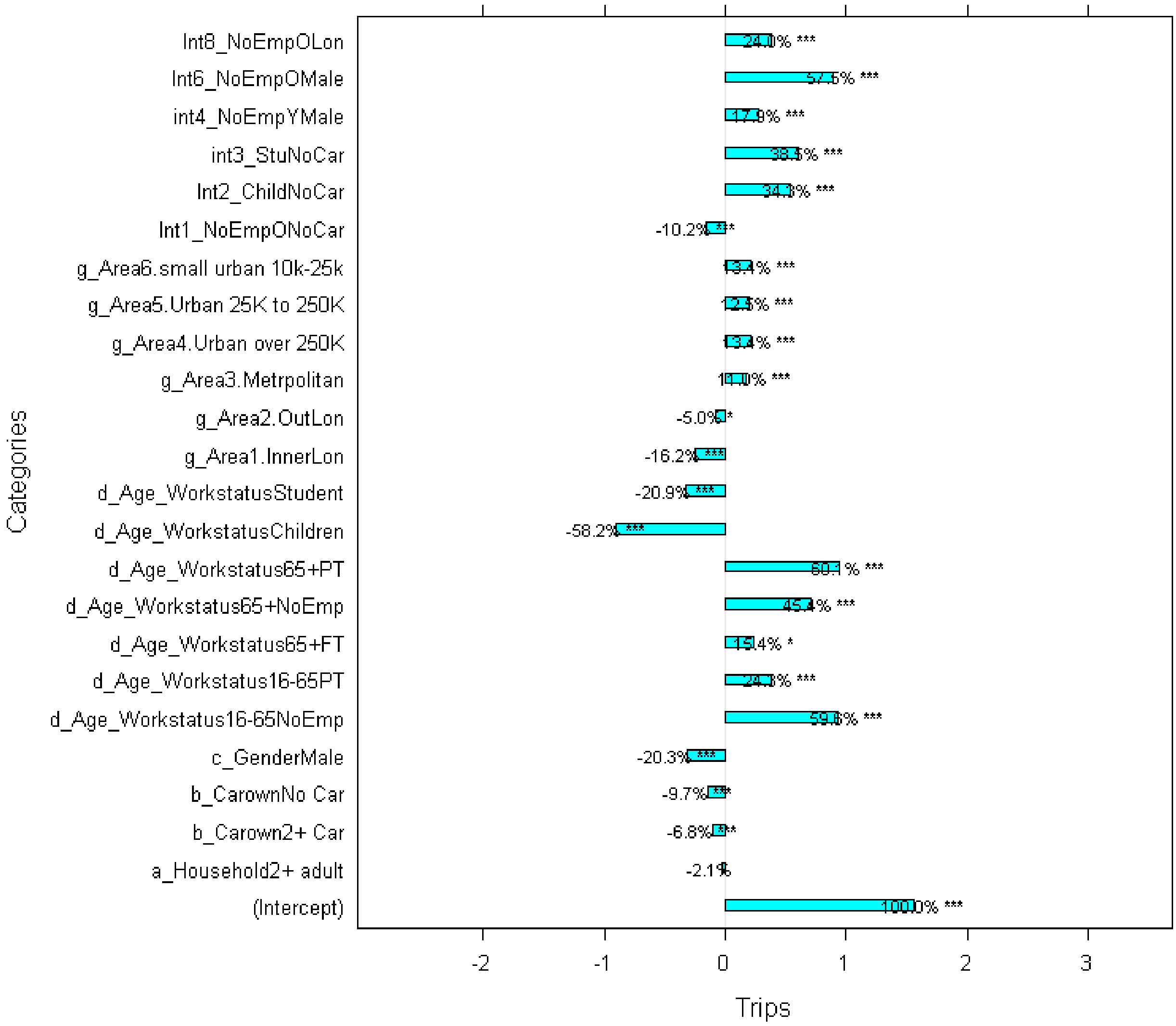

Figure 9.2 Model of shopping trip rates with interactive terms

9.1.13 As an example of the difference that the inclusion of interaction terms can make when estimating trip rates, Figure 9.3 demonstrates the effect on gender of including interactions when all other predictors are fixed at their typical value. The difference between males and females in their estimated number of trips is significantly bigger for each work status category after the inclusion of the interaction terms. To demonstrate the reason, the results of the simple shopping model provided in Figure 9.4 are now compared against those shown in Figure 9.2 which is an otherwise identical model that includes the six interaction terms.

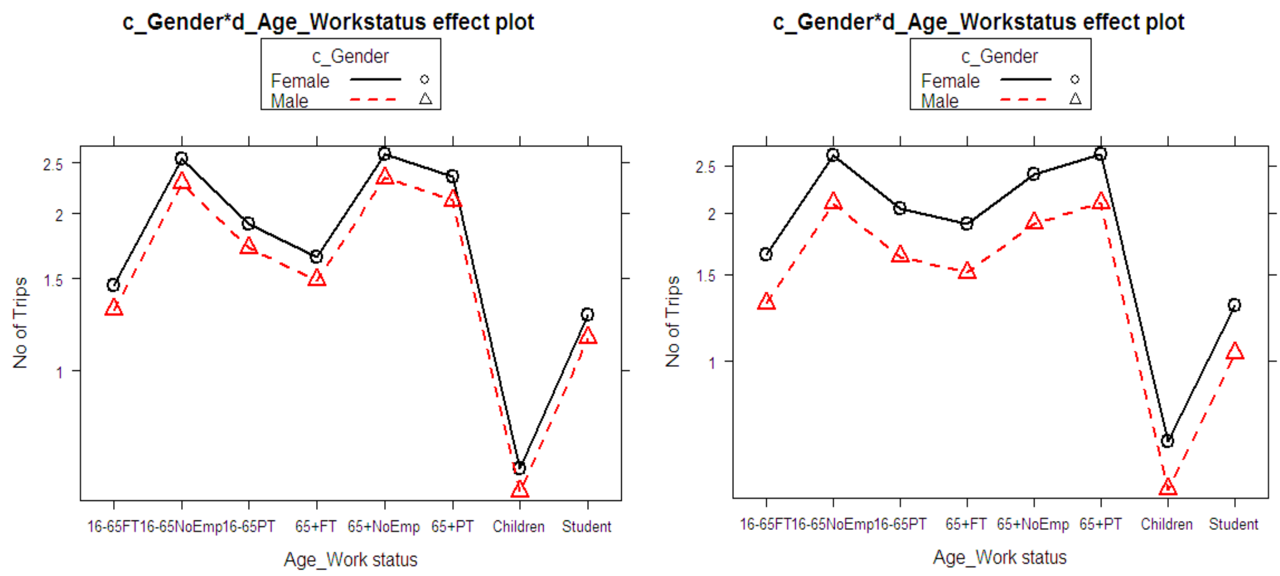

Figure 9.3 Comparison of the effect plot of gender before (left chart) and after (right chart) inclusion of interactive terms

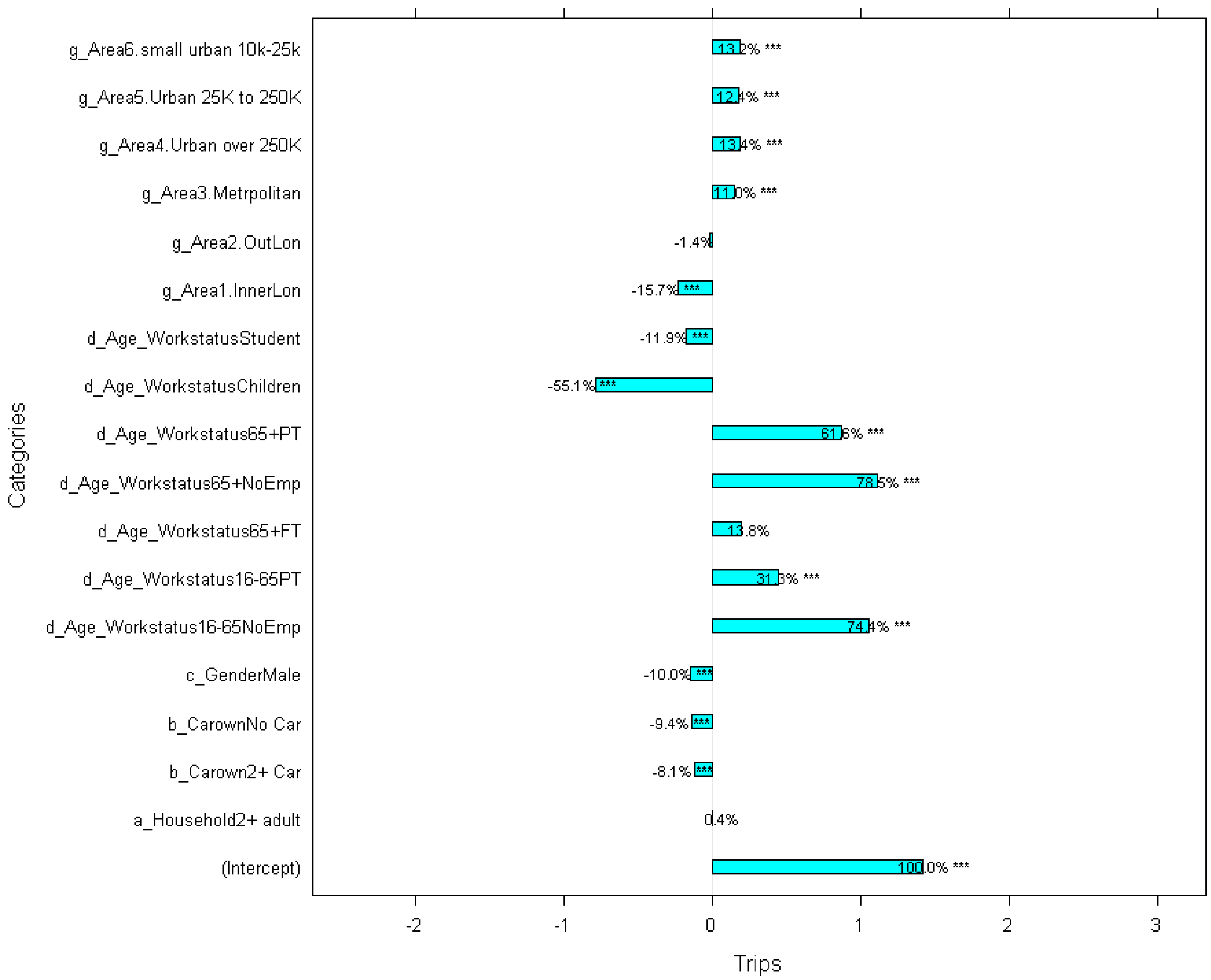

Figure 9.4 Simple model of shopping trip rates

9.1.14 The complex model of Figure 9.2 shows that the inclusion of the interaction variables: Int4 - “16-64 not Employed and male” and Int6 - “65+ not employed and male” leads to a reduction of 20% for male (c_GenderMale) shopping trips, compared to the 10% reduction estimated in the simple model of Figure 9.4; both reductions are relative to the female (reference category) rate. This comparison shows that the smaller reduction in male shopping trips in the simple model was due to not identifying the effect of males not in employment who tend to make between 18% and 58% more shopping trips than other males. When this effect is separated out in the interaction model, the rest of the males are now shown to make 20% fewer shopping trips than females. This increased discrimination from the inclusion of relevant interaction terms is what has been illustrated by Figure 9.3.

FURTHER FINDINGS

9.1.15 More detailed discussion on the effect of each explanatory variable could be provided here for each individual trip purpose. However, this approach has not been taken to prevent making this report unnecessarily long and repetitive. The interested readers should instead refer to Appendix E, where detailed charts showing the relative importance of the categories within each variable are provided for each trip purpose in turn. When one of the main variables is not shown on the chart or if it does not have a significance code associated with it, it can be considered to be statistically insignificant.

9.1.16 However, further discussion on the findings from home-based shopping is provided below as an example of the results that can be extracted. The following are the main findings from the simple model presented in Figure 9.4.

- The not employed and the part-time workers make more shopping trips than do full-time workers aged 16-64 (reference category), while children make many fewer shopping trips on their own. The influence of the 65+ FT workers is not significantly different from the reference category and so they could be combined within a broader new reference category.
- Persons in 2+ adult households make almost the same number of trips as those in 1 adult households. This distinction of household size could be dropped from this model.
- Those in both 2+Car and No Car households make fewer shopping trips than those in 1 Car households.
- Relative to the reference category of rural areas the number of shopping trips is 16% lower for residents of Inner London and 11% to 13% higher for those in Metropolitan, big, medium and small urban areas. Because this model is only for home-based shopping trips, this may partly be explained by the willingness of Londoners to combine their shopping trips with other trips as non-home-based trips: for example doing their shopping on their way home after work.

9.2 VALIDATING THE MODELLING APPROACH

9.2.1 This section explains the approach that was taken to validate our modelling techniques (i.e. usage of weighted NB regression) and to make sure that the results produced will be reliable.

9.2.2 It was decided to use as HBW (FT/PT) as the trip purpose for which results would be validated. Provided that this was successful the methodology would be assumed to be reliable for other trip purposes as the same approach is used for estimating trip rates for all trip types and purposes.

9.2.3 The first stage in validating results was to predict trip rates per person based on our estimated model. The following steps were followed.

- The most detailed test model of commuting trips is used for the validation test. This includes both the income and NS-SeC variables that are discussed further in Chapter 11.
- All possible combinations of levels of the explanatory categorical variables were defined and the estimated model was used to estimate trip rates for each combination. The set of variables and all of their associated categories used within this model are presented in Table 9.2.

Table 9.2 Variables and associated categories used in validating the model

Variables	a_Household	b_Carown	c_Gender	d3_Age_work_NSSEC	e_Income	g_Area
Level1	1 adult	No Car	Male	5_PO	1_below4k	1.InnerLon
Level2	2+ adult	1 Car	Female	6_PY_01_M&P	2_4k-9k	2.OutLon
Level3		2+ Car		6_PY_03_InterOcc	3_9k-20k	3.Metrpolitan
Level4				6_PY_04_SmallandOwn	4_20k-35k	4.Urban over 250K
Level5				6_PY_05_LowsupTech	5_35k-60k	5.Urban 25K to 250K
Level6				6_PY_06_SemiRoutine	6_over60k	6.small urban 10k-25k
Level7				6_PY_07_Routine		7.Rural
Level8				7_FO		
Level9				8_FY_01_M&P		
Level10				8_FY_03_InterOcc		
Level11				8_FY_04_SmallandOwn		
Level12				8_FY_05_LowsupTech		
Level13				8_FY_06_SemiRoutine		
Level14				8_FY_07_Routine		

9.2.4 The next stage calculated the number of individuals, weighted by the household weight (i.e. W2), in each category for which trip rates had been predicted. This was then multiplied by its relevant predicted trip rate to give the total number of predicted trips for each group. The overall sum of trips was then calculated by adding the total number of trips across all categories together.

9.2.5 The total number of trips of all categories was calculated alternatively by summing directly from the NTS dataset the number of trips, weighted by trip and by household weights.

9.2.6 These two predicted and actual total trip totals were then compared with each other and the difference was less than 0.1%. This confirms that the model is actually predicting the right number of trips overall so that no trips are missed by using this estimation and modelling approach. Figure 9.5 summarises the steps in this validation process.

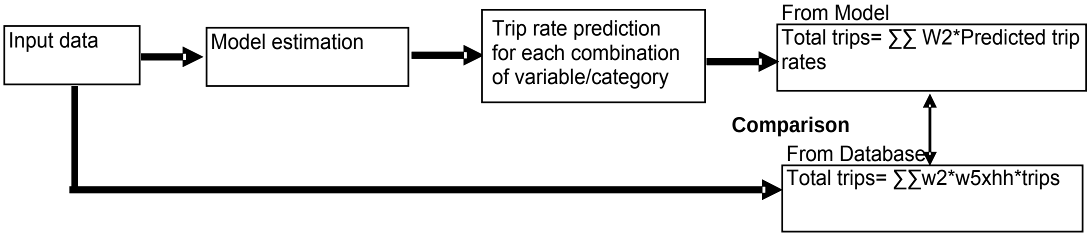

Figure 9.5 Model validation process

10 Analysis of Accessibility

10.1.1 NTS data gives some indicators of accessibility at the household's residence. These indicators could not be extended or refined except through conducting detailed GIS analysis on the respondents' residential location and through collecting further complimentary information which is outside the scope of this research. Instead, we decided to include the accessibility indicators available from NTS dataset into the model to find out more about their relative importance. In addition to area type, the main variables which were added to the basic model for testing accessibility are presented in Table 10.1.

Table 10.1 General accessibility variables tested

NTS codes for variables	Description
P6	Population density
H13	Walk time to bus stop
H14	Frequency of bus service
H15	Walk time to railway station
H16	Bus time to railway station
H17	Type of railway station
H100, H121, H122	Bus Satisfaction, reliability and frequency (perceived)

10.1.2 Furthermore, some accessibility variables that were purpose specific are provided in Table 10.2..

Table 10.2 Purpose specific accessibility variables

Purpose	Specific variables used
Home-based personal business - HBPB	**Distance variables**: H18 (walk dist to Doctor), H23 (Bus dist to nearest chemist), H28 (Walk distance to nearest General Hospital) and H29 (Bus distance to nearest General Hospital) **Time Variables**: H160 (JT to nearest GP), H161 (JT to nearest chemist), H162 (JT to nearest hospital), H164 (JT to nearest grocer) and H165 (JT to nearest post office)
Home-based shopping - HBS	**Distance variables**: H24 (Walk distance to nearest food store), H27 (Bus distance to nearest shopping centre), **Time variables**: H163 (JT to nearest shopping centre)
Home-based education - HBEd	**Time variables**: H166 (JT to nearest primary school), H167 (JT to nearest secondary school), H168 (JT to nearest college)

10.1.3 Testing all these variables together was not always possible due to the high or full correlation between some of them for some trip purposes. For instance bus time and walk time to rail stations were found to be highly correlated in most cases. Furthermore, in extracting data for tests, we were forced to ignore some records that had irrelevant answers to some of the accessibility questions (e.g. the code: do not know, for walk time to rail stations). However, these exclusions should not affect the significance levels as the remaining dataset still contains a very large number of records.

10.1.4 For those trip purposes that had area type as an important explanatory variable, it generally remains significant after the addition of extra accessibility variables. However, the marginal effect of some of the area type categories was reduced and in some cases became insignificant after the inclusion of other accessibility variables. For instance, the residents of Inner London make more shopping trips than those residing in rural areas in the model that includes other accessibility variables. This is the reverse of the pattern provided by the simple model without accessibility variables (see Figure 9.4). This suggests that the inclusion of accessibility variables, particularly when they are significant, is strongly correlated with area type.

10.1.5 In general, for every purpose it is the indicators: area type and population density that are the most important accessibility variables. The rest of the indicators listed previously have generally been shown to have limited explanatory influence.

10.1.6 It is not beneficial to include both area type and population density within the same model as they are highly correlated. Inclusion of both within the model would usually make population density insignificant. With the exception of education, which showed only a marginally better goodness of fit from using population density in place of area type, and of Personal Business, for which neither area type nor population density is significant, all other tests suggest that area type provides the better representation of the variation in the number of trips per person.

10.1.7 For those few travel purposes for which other accessibility variables happened to be significant, it is bus frequency and walk time to bus stop and rail stations that appear to be the most important accessibility factors. Reductions in the number of trips per capita were mainly observed when the bus frequencies are more than half an hour and the walk time to a bus stop and to a railway station are more than half and a quarter hour respectively, suggesting that accessibility to public transport stations is important in defining the number of trips, only when the distance or time is beyond a certain limit.

10.1.8 Table 10.3 shows the main findings from running accessibility tests for each travel purpose.

Table 10.3 Main findings from accessibility tests for each home-based trip purpose

Purpose	Main findings
HB Work	Examining the models with area type and other accessibility variables suggests that although some area type categories are significant, none of the additional accessibility variables are significant at the 99.9% level. Only walk time to bus stop and bus frequency are significant at the 99% level when the walk time and frequency are more than 27 minutes and 1 every hour, respectively.
HB Employers Business	None of the accessibility variables including area type and density are statistically significant.
HB Holiday	In addition to area type, walk time to rail station is significant at the 99.9% level, when the walk time is more than 44 minutes.
HBPB	Area type is but none of the other accessibility variables are significant.
HB Recreation	Area type and population density are but no other accessibility variables are significant.

Purpose	Main findings
HB Shopping	Travel time to shopping centre is statistically significant at 99% level when it is more than 40 minutes. Bus frequency is significant at the 99% level when it is more than half an hour Walk time to bus stop is highly significant (at 99.9%) in reducing the number of shopping trips when it is more than 6 minutes. It also proved that when local rail stations run less frequent services, the number of shopping trips will decline.
HB Visiting Friends	In addition to area type, bus frequency of more than half an hour at the 99% level is significant.
HB Education	In contrast to all other purposes, population density appears to be a more important variable than area type. None of the other accessibility variables has much importance, including walk time to schools

11 Analysis of Income and Socio-economic Class

11.1 INTRODUCTION

11.1.1 In this chapter the effect of adding individual income into the basic model of the number of trips is reported. The inclusion of socio-economic characteristics was also tested, both separately and in combination with income, in order to identify which of income or socio-economic status provides greater explanatory power in the trip rate model. These tests were done for each of the main home-based purposes and are based on weighted NBR. In this Chapter, the main findings from analysing all purposes are reported. Interested readers should refer to Appendix E for more detailed purpose by purpose findings.

11.1.2 Because the income and the National Statistics Socioeconomic Classification (NS-SEC) variables each included "children" as one of their categories, testing both together was impossible for all purposes except for commuting and HBEB, because for these only those in full-time (FT) and part-time (PT) employment groups were examined. Therefore, in all other tests, the "age_workstatus" variable is combined with income and NS-Sec to form a new variable called "Age_work_Income" and "Age_work_NS-SeC" to avoid the issue of multi-collinearity. Where both income and NS-SeC were tested in one model, the variable of "Age_work_Income_NSSec" was formed.

11.2 MAIN FINDINGS FROM ANALYSING INCOME AND SOCIO-ECONOMIC VARIABLES

11.2.1 For each purpose, three main tests were carried out by adding each one of "income", "NS-SeC" or both together into the best simple version of the model. These three resulting estimations were then compared with each other.

11.2.2 Interestingly, income and socio-economic class were not highly correlated. This provided the opportunity to include both of these variables into the model without generating serious problems of collinearity. This was ascertained through monitoring the correlation matrix and also by the fact that when either NS-SeC or income alone is significant in our estimation, adding in the other will not make the first then become not significant.

11.2.3 Table 11.1 compares the goodness of fit of the three estimated models for each purpose, where the best model is highlighted in grey. It shows that for most purposes, the AICs of the three models (i.e. income, NS-SeC and Income&NS-SeC) are of similar magnitude to each other except for the commuting and the business trip purposes.

Table 11.1 Comparison of the goodness of fit of three estimated models: Income; NS-SeC; Income & NS-SeC

Purpose	AIC of simple Model	AIC of simple model plus income	AIC of simple model plus NS-SeC	AIC of simple model plus income and NS-SeC
HBW (FT/PT)	205476	205180	204142	203890
HBEB (FT/PT)	60906	60816	60062	60021
HBHol	118366	118191	118175	118100
HBPB	226542	226432	226404	226399
HBRec	294200	293885	293861	293764
HBShop	317432	316645	316696	316667
HBVF	251139	250661	250757	250751

11.2.4 Both the commuting and the business trips have the biggest improvement, compared to their simple forms, from inclusion of NS-SeC and Income terms. The combination of NS-SeC, age and workstatus gives a much better model of commuting and business trip rates than the combination of income, age and workstatus. The goodness of fit for commuting and business would further improve slightly from inclusion of both income and NS-SeC variables.

11.2.5 For all purposes except “shopping” and “visiting friends”, including both income and NS-SeC will improve the model. However, if just one is to be selected, NS-SeC is a much better variable to include as the difference between the AIC of a model with NS-SeC and that with both income and NS-SeC is marginal.

11.2.6 In contrast, for “Shopping” and “Visiting Friends”, income is the variable which provides the best goodness of fit.

11.2.7 Overall, it can be concluded that the inclusion of socio-economic class has a stronger explanatory power than the inclusion of the income variable.

11.2.8 Income bands are statistically significant for the most of the travel purposes so that adding these into the model, which is built from a combination of socio-economic class and other basic variables, could improve the goodness of fit. However, this improvement is marginal which makes it difficult to justify making the model more complicated through the inclusion of an income term.

12 Analysis of Rail Trips

12.1 INTRODUCTION

12.1.1 This chapter describes the rail trip analysis that was carried out using a weighted Negative Binomial regression approach. Overall changes through time in rail trips based on the NTS dataset are discussed first, followed by a comparison of these NTS based trends with the growth trends observed from rail industry data. The findings from analysing NTS rail data are presented. Because the number of rail trips captured within the NTS data is relatively small, it is only for the only for the purpose commuting that home-based rail trips had a large enough sample to be distinguished from non-home-based. For business trips and for all other purposes combined, all home and non-home-based rail trips are combined together when estimating the trip rates.

12.2 TRENDS IN YEARLY RAIL TRIPS

12.2.1 This section examines the aggregate changes in the number of rail trips per person per annum estimated based on the NTS data. This is helpful in validating the results of regression analysis. A comparison with published rail industry data has been also provided in order to examine whether the NTS data is in line with that published in the National Rail Trends (NRT) Yearbook (2008). Figure 12.1 shows the comparison of annual rail trips per person from 1996 to 2006 based on NTS data and on the published National Rail Trends. The trend in long distance rail journeys (>50 miles) from NTS is also shown.

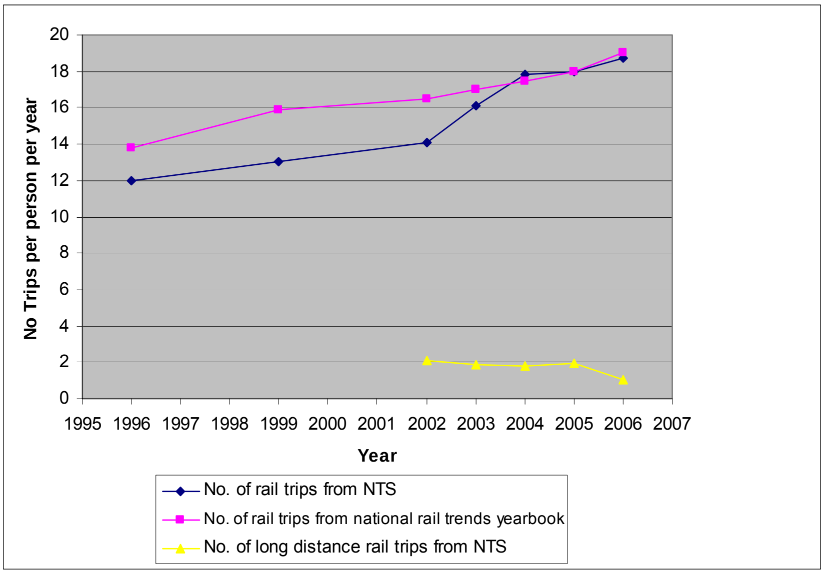

Figure 12.1 Comparison of trends in rail trips per person, based on NTS and National Rail Trends data

12.2.2 NRT data shows a reasonably continuous growth in number of trips over time. From 2004 to 2006 the NTS shows similar rail trip rates to those of NRT but the NTS rates are rather lower in earlier years. These differences may result from sampling errors in NTS caused by the relatively small sample available for the analysis of rail movements in any year, particularly in the period prior to the expansion in sample size in 2002. In reality a relatively small proportion of the population make intensive use of rail while a large proportion make little or no use of rail in a year. This heterogeneous pattern of behaviour increases the sampling errors for rail travel within the NTS. Sampling errors are likely to impact particularly on the long distance trips which may explain their sharp decrease from 2005 to 2006.

12.2.3 Figure 12.2 and Figure 12.3 show the split of all rail trips by purpose from both the standard NTS data and from the companion Long Distance Survey database.

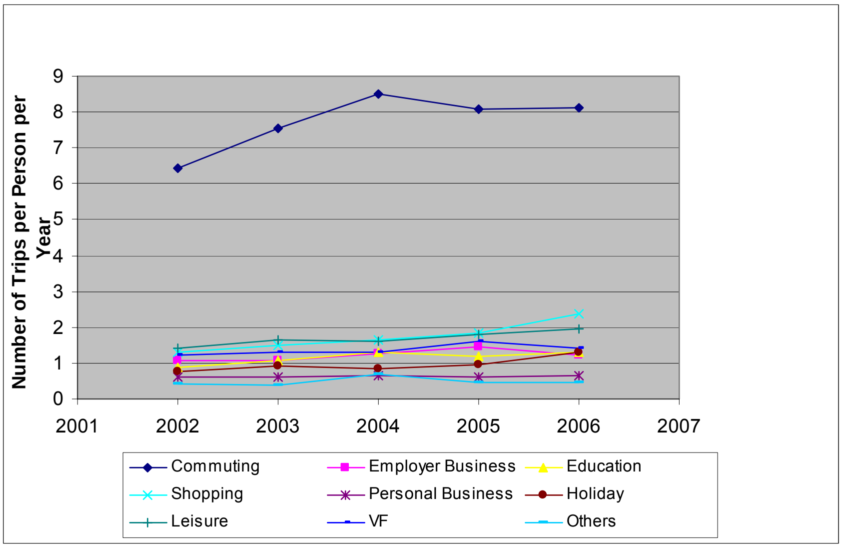

Figure 12.2 Number of rail trips by purpose (standard NTS data)

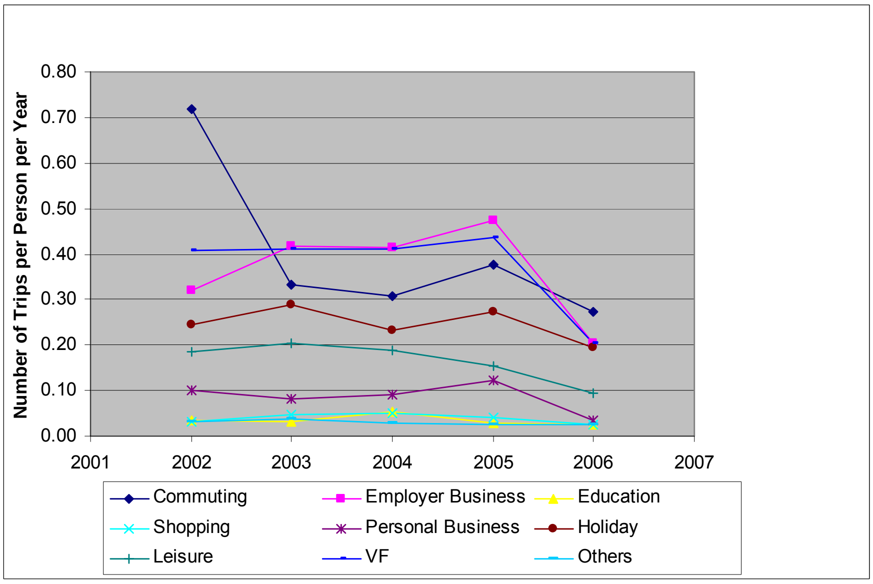

Figure 12.3 Number of rail trips by purpose (Long Distance Survey only)

12.2.4 It can be seen from Figure 12.2 that the greatest number of rail trips is made for commuting. The next most numerous rail purposes are shopping and then leisure, both of which exhibit a strong consistent increase through time in their number of rail trips. Commuting with slightly below 50% of all rail trips shows a rapid increase from 2002 to 2004 followed by a sharp decrease from 2004 to 2005 and a small change from 2005 to 2006.

12.2.5 If any trend over time can be seen from the long distance journeys in Figure 12.3, it is a reduction rather than an increase in the number of trips for all purposes. Employers business and visiting friends are the two purposes for which people make the greatest number of long rail journeys.

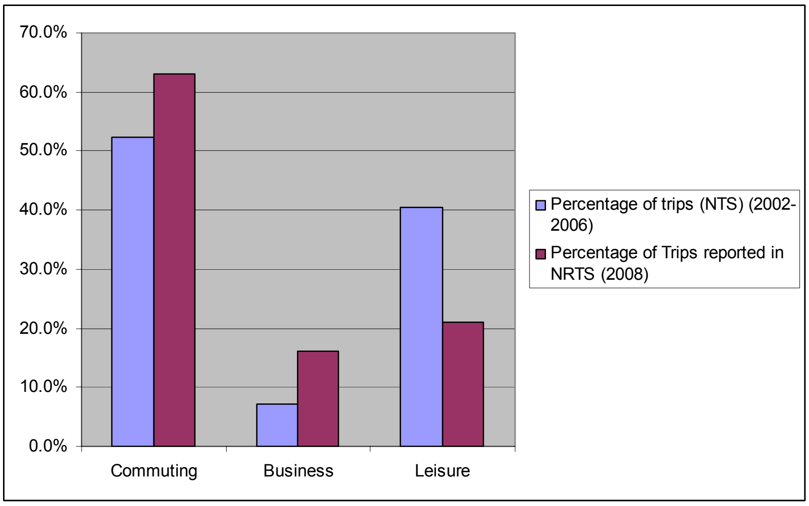

Figure 12.4 Comparison of percentage of trips by purpose, between NTS and NRTS datasets

12.2.6 Figure 12.4 compares the percentage of rail travel by purpose derived from NTS data with that reported in the 2008 National Rail Travel Survey Report. It can be seen that commuting and business trips are underestimated while Leisure (which includes all other purposes) are overestimated. Part of the explanation for this difference will lie in inconsistencies between definitions. Within NRTS, journeys were grouped into three main categories: commuting, business and leisure. The NRTS commuting category includes travel to work *plus to education*, which may explain why this NRTS proportion is larger, though the reasons for the differences in the other two categories are less obvious.

12.3 ANALYSING RAIL TRIP RATES BY REGRESSION

MAIN FINDINGS

12.3.1 A set of NB regression analyses have been conducted for three travel purposes (i.e. commuting, business and the rest) in order to examine the effect of a variety of variables, including survey year, accessibility factors and income/NS-SeC. Only standard NTS data is examined in this section; no additional tests have been conducted for long distance journeys.

12.3.2 The following are the key findings from the NB regression tests:

- As would be expected from the previous section, trend by year is not significant for rail trips for the HBW and employers business purposes. For the rest of the purposes combined together, yearly trend shows up as significant if considered alone but no meaningful trends can be seen when other variables are added in to improve the fit of the model. This implies that the changes over time could be captured adequately by these other variables.

- NS- SeC is generally a better estimator of number of trips by rail than income but when both were considered together, the model improved overall. These conclusions are based on the comparison of the models' overall goodness of fit as provided in Table 12.1.

Table 12.1 The AIC indicator of three forms of the rail model

	AIC of model with NS-SeC	AIC of model with Income	AIC of model with both NS-SeC and Income
Business	11522	11608	11398
Commuting	24298	24514	24164
Others	61934	62321	61821

.

- Area type in general is significant in explaining the number of rail trips. Residents of London tend to make a larger number of rail trips than people living elsewhere in the country.
- Accessibility variables, specifically walk time to rail stations and frequency of services, show a significant influence on number of rail trips. Considering that these variables were not as important in estimating the number of trips in total, this suggests that the effect of accessibility on rail mode choice is greater than on trip production rates in total across all modes.

12.3.3 In line with expectations gained from the aggregate analysis, testing a series of year specific dummy variables alone did not show any apparent trend over time. Interestingly and in contrast to the trend shown in Figure 12.2, even the difference in number of trips in 2003 and 2004 in comparison to the reference year of 2002 is not statistically significant. The reason is the large standard error due to the large variance in individuals' number of rail trips in a specific year. This large sampling error results from the availability only of a relatively small number of individuals within the NTS who make rail trips.

12.3.4 Extending the NTS database used in the estimation to include earlier years might help marginally on this front. However, it has the potential disadvantage of complicating the analysis through the inclusion of data prior to 2002 that was collected and processed in a somewhat different manner, including non-trivial differences in some of the data definitions in use.

ANALYSIS OF COMMUTING TRIPS

12.3.5 Figure 12.5 shows the model of commuting trips by rail where accessibility variables are included in conjunction with base socio-economic factors and area type. Contrary to the findings from the trip model of commuting trips for all modes combined (i.e. which included rail), where none of the accessibility variables were significant, the walk time to railway station and frequency of rail services are significant in determining the number of commuting trips by rail.

12.3.6 The category part-time old worker (i.e. 65+ years old) is considered as reference. It can be seen that among those under 65 with the exception of full-time people in NS-SeC categories: intermediate occupations (3) and managerial professional (1), all other part-time and full-time workers make either fewer or a similar number of rail trips for commuting compared to this reference category. It is primarily the full-time office based workers that use rail for commuting, as well as those residing in London.

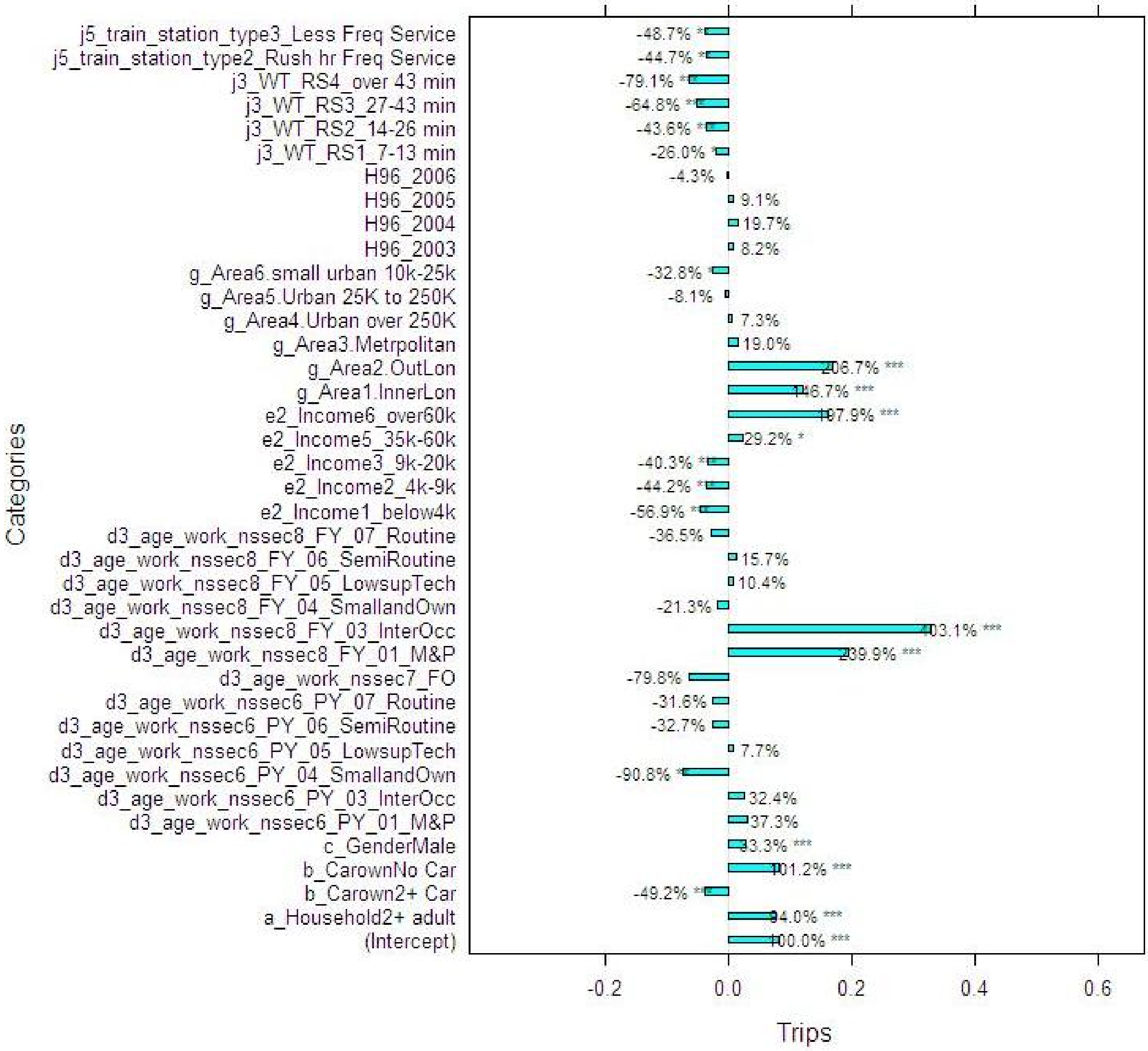

Figure 12.5 Model of commuting trips by rail for FT/PT workers

12.3.7 Three tests of NS-SeC alone, household income alone and both together were conducted. As can be seen in Table 12.2, in line with the findings for commuting trips with all modes combined, NS-SeC is a more important variable than income. Putting both income and NS-SeC into the model does not make a noticeable improvement in the overall goodness of fit.

Table 12.2 Comparing the AIC values of commuting by rail and overall commuting trips

	Income alone	NS-SEC alone	Income+ NS-SEC	Percentage improvement from adding income compared to NS-SeC alone
Commuting: all modes	205180	204142	203890	0.12%
Commuting by rail	24514	24298	24164	0.55%

12.3.8 Figure 12.6 shows the comparison between the two models of commuting trip rates for all modes combined and just by rail. Here for clarity of presentation the reference area type has been changed to be Inner London rather than the rural area type of Figure 12.5. Socioeconomic factors (i.e. household size, car ownership, age, work status and gender) have a larger impact on the number of commuting trips by rail in comparison to that on commuting trips overall. This is also true for variation by NS-SeC and income. A clear pattern of change in number of commuting trips by income group exists for rail, suggesting that those in the higher band make more rail trips but it is not as clear over all commuting trips, where the people in mid income categories make the greatest number of commuting trips.

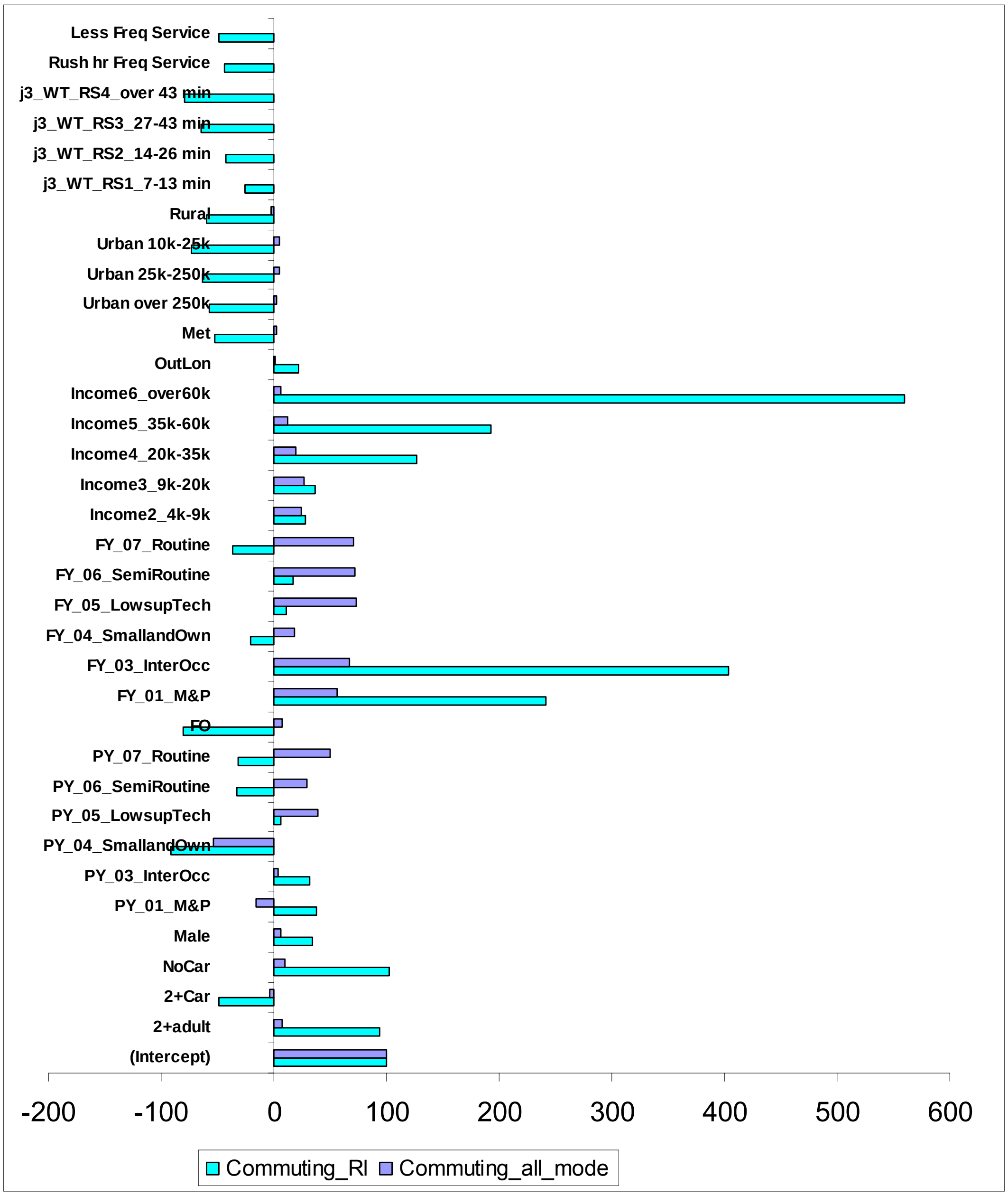

Figure 12.6 Comparing percentage impacts on commuting trip rates - model of all modes combined versus rail only

12.3.9 Unlike the approach taken for commuting where only home-based trips were examined, for both business and the rest of the purposes, all trips by rail were combined. Therefore, no comparison of rail trip rates versus number of trips for all modes combined is conducted.

12.3.10 The hypothesis of changes in trip rates across the years 2002-2006 was tested by conducting a regression analysis of year alone and also that in combination with other basic variables. None of these revealed any evidence of changes in trip rates over time. Figure 12.7 shows the marginal effect of various variables used in the model for all business trips by rail.

- Walk time to rail station is significantly important in describing the variation in number of rail business trips. A rapid drop in number of rail trips can be observed specifically when the walk time to rail station is more than 15 minutes. Similar to the findings of HBEB trips over all modes, the number of rail trips is also greater for those living in inner and outer London.
- Gender and household size are not significant variables for estimation of number of rail trips.
- Interestingly, although a large variation by NS-SeC group can be seen in Figure 12.7, none of its categories are reported as significant. This is primarily because of the large standard errors and variations associated with each category which results from working with a small sized dataset.

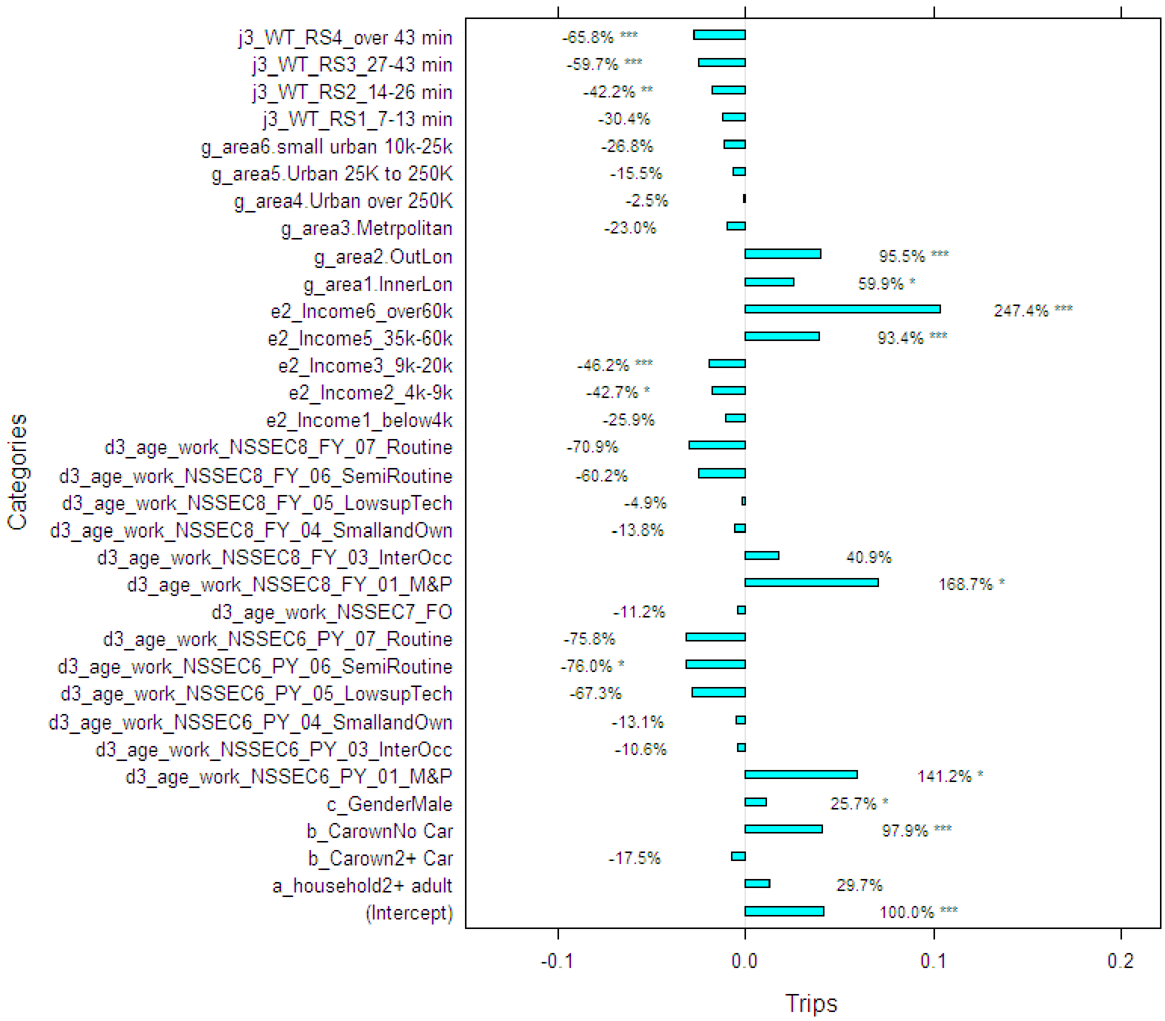

Figure 12.7 Model of employers business trips-by rail

12.3.11 One interesting finding is that when neither income nor NS-SeC is considered within the rail model of business trips, Individuals in 2+car households make a significantly greater number of rail trips than those in 1-adult households. The direction of this effect reverses when income and NS-SeC are introduced as can be seen in Figure 12.7 where people in no car households now make more trips by rail than those in 2+ and 1-car households. This suggests that in the former case, car ownership shows some of the effects which could have been captured by NS-SeC and/or income if they were included.

12.3.12 With regard to income bands, it can be seen that while people in lower income bands (i.e. household income below £20,000 per annum) do not show a large difference from each other in their number of rail trips, those in higher income bands make many more business rail trips than the reference group (i.e. £20-35k).

ANALYSIS OF OTHER PURPOSES COMBINED

12.3.13 Next we discuss trips by rail for all other (i.e. non work or business) purposes combined. Unlike business and commuting, testing year alone in a regression model across the years 2002-2006 shows an increasing trend in the trip rates over time. However, adding other standard explanatory variables into the model changes the pattern and significance level of number of trips over time. This suggests that the increase in trip rates could be explained by changes in the accessibility, area type or socio-economic characteristics of the individuals who were sampled in the NTS.

12.3.14 Figure 12.8 shows the marginal effect of each variable which was tested within the model. The following are the main findings for rail trips for the other trip purposes category:

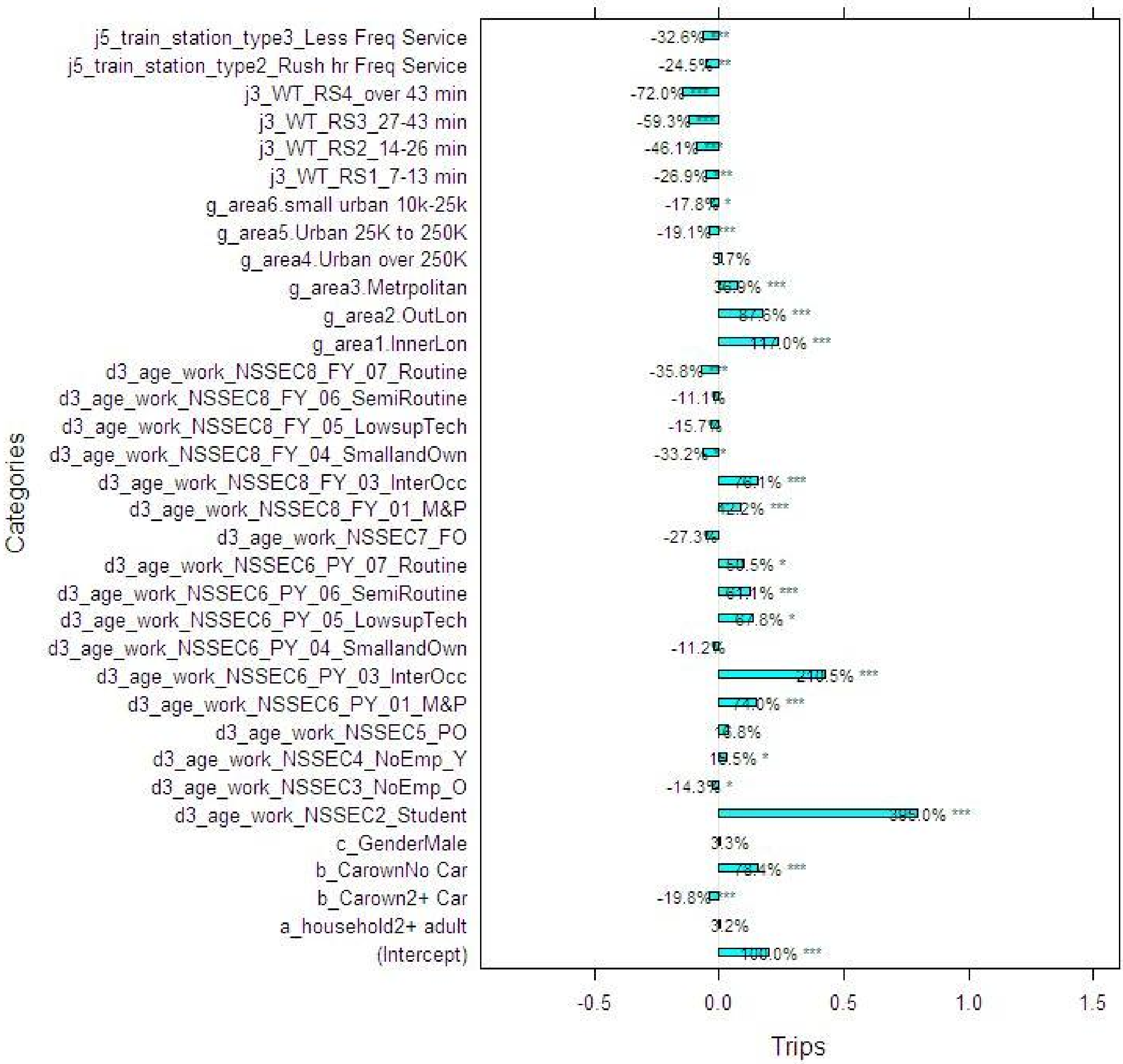

Figure 12.8 Model of all other trips-by rail

- People with no car make more rail trips than those with 1 or 2+ cars.
- Household size and gender are not important variables in determining trip rates by rail. This was also found when analysing business trips by rail.
- Students make the greatest number of rail trips for other purposes combined. Part of the reason could be because of relatively high proportion of education trips in comparison to all other purposes. Part time workers in general make the most number of trips, among which those in the managerial/ professional and intermediate NS-SeC categories tend to make the most trips.
- The trend by area type suggests that individuals living in Inner London, Outer London and Metropolitan areas make significantly more rail trips than those in the rest of country.
- With regard to accessibility, walk time to rail station has emerged as an important factor. Also those people living in places with less frequent rail services prefer not to make rail trips.

12.4 IMPROVING TRIP END GROWTH ESTIMATES BY RAIL

12.4.1 The rapid growth rate in trips by rail over the last decade has not been captured in the NTEM procedures that underlie the forecasts in TEMPRO. Based on our past modelling experience it is not a surprise that rail has been the most difficult to represent *within a stand-alone trip end model* for reasons now discussed in the context of commuting trips, though the issues are also relevant to other and to business trips, though with decreasing relevance.

12.4.2 Modes car and walk/cycle and to a lesser extent bus tend to be reasonably ubiquitously available within a residential area. In contrast, for rail there are major differences between local areas in their accessibility to rail stations, which implies that the rate of usage of rail is likely to vary much more substantially by location than would the usage of other more ubiquitous modes. Nevertheless, this differentiation at the residence end can be approximated within a trip end model through the use of suitably measured accessibility terms so that it is not an insurmountable problem within trip production modelling for rail. The estimation results presented previously have confirmed that variables denoting ease of access to a rail station play a much more substantial role for rail trips than for trips overall.

12.4.3 The real modelling problem arises at the workplace (attraction) end of a commuter rail trip rather than at the home-based (production) end. At the residence end, even when people reside in low density suburban or rural areas that are not close to a railway station, they have the option of park-and-ride or kiss-and-ride to enable them to access a suitable station at some distance from their home, provided that their household has access to car. The situation is different at the destination end of a trip, where unless there is a reliable bus or underground connection from the terminating rail station through to their ultimate destination, most travellers will be restricted to a relatively small radius adjacent to that station. This is an important explanation of the relatively high use of commuter rail by managerial/professional and intermediate occupations (see Figure 12.6), since even though these workers may often reside in villages at some distance from a rail station, many of their office jobs in cities such as in Central London, Leeds, Reading, Cambridge are congregated at high densities adjacent to rail stations. Through PPG13 the planning system has successfully encouraged this type of centralised office development in recent years. In contrast, jobs in industry and distribution have decentralised away from inner city areas so that most such workplaces are highly inaccessible to rail and are likely to remain so.

12.4.4 The maps of London and its surrounds that are presented in Figure 12.9 illustrate the points made above using journey to work data from the 2001 Census, noting that the same broad pattern was also exhibited by the 1991 Census. The proliferation of mid red districts (plus unitaries and boroughs) shown in the left hand map, indicates that a substantial proportion of the districts within a 40 mile radius of London have 5% or more of their *resident* labour force who travel to work by train or by underground. In the top right hand side which shows on the same scale the rail mode split by district at the *workplace end*, there is a general absence of mid red wards, except within London. There are only six districts surrounding London (Mole Valley and Worthing to the south; Watford to the north west; and Epping Forest, Brentwood and Basildon in Essex) in which more than 5% of the workplaces are accessed by rail/underground, compared to more than 60 such districts at the residence end. Even within London, the 18 boroughs with more than 30% of residents commuting by rail or underground reduce down to just 5 Central London boroughs at the workplace end.

12.4.5 This shows that workplace location is the main determinant of whether rail is used on a commuting trip, because very few workplace zones have a very high rail proportion and most have a minimal rail proportion. In contrast, the proportion using rail at the residence end shows differences between locations that are much less extreme.

District of residence | District of workplace

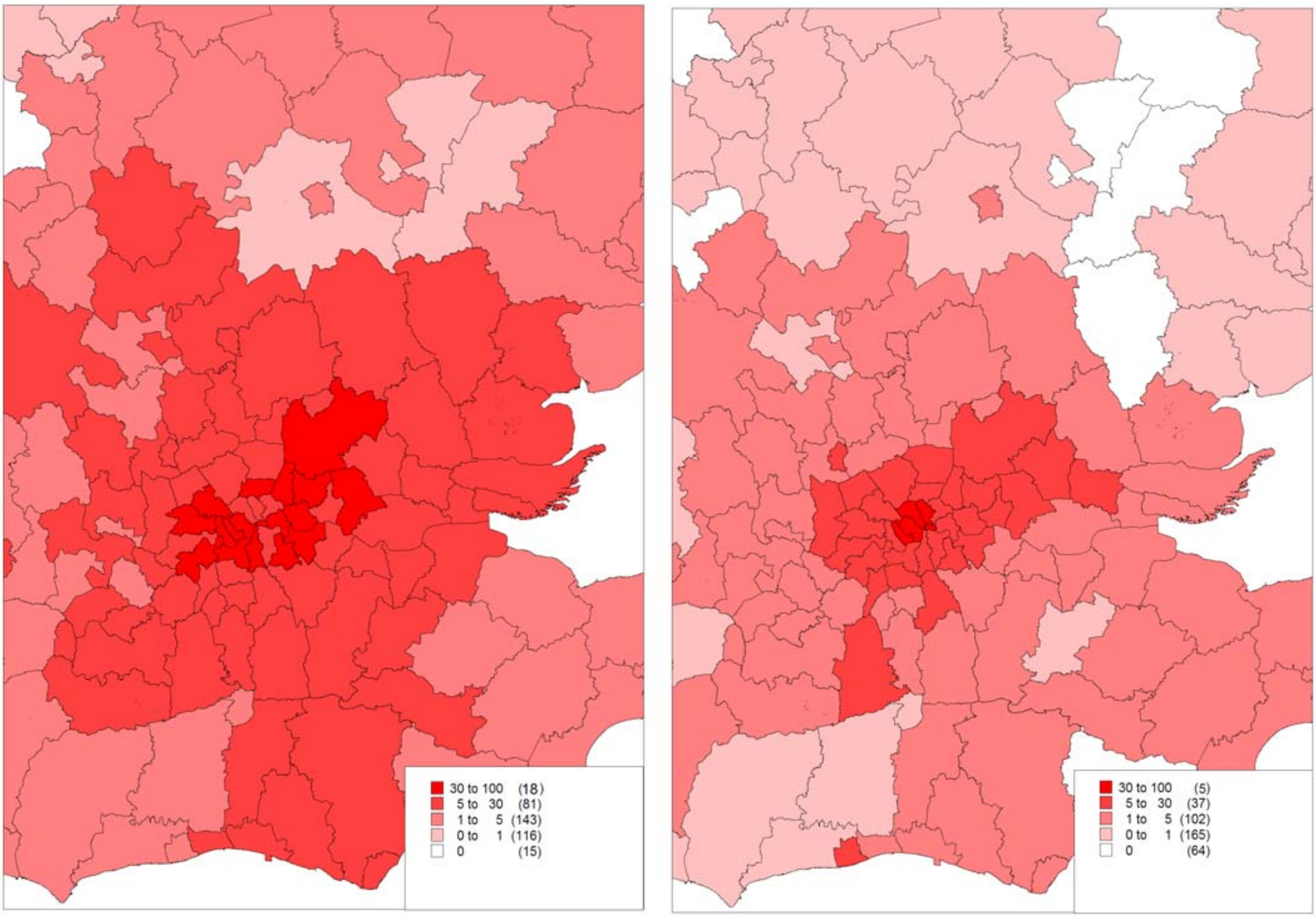

Figure 12.9 Contrast in commuter rail mode choice percentage between residence and workplace district for London and surrounds, 2001 Census

12.4.6 The general principle outlined above for commuting is that it is the characteristics of the workplace (attraction) end location more than those of the residence (production) end that determine the rail trip rate. This principle is also of relevance to trips by the aggregate purpose, other. Because of the longer than average journeys typically travelled on rail and the limited network connectivity of rail to orbital destinations and even to many radial destinations other than in London, the attractiveness of the usage of rail from a residential area is heavily dependent on the opportunities at the other end of the rail line and on the modal competition to reach these opportunities. Trips to shopping, recreation and leisure activities that are concentrated in those city centres that have expensive or scarce parking are likely to be attracted to rail, provided that these activities are accessible to the rail terminal.

12.4.7 The relative concentration of business services in offices in city centres adjacent to rail stations again has made rail a more attractive mode for longer distance business travel for professionals, provided that it is on radial routes to major urban centres that have fast and frequent services.

12.4.8 Along with any influence from recent improvements in rail services and general supply quality, five main factors have been identified in this Chapter that influence the recent more rapid growth in rail travel compared to other modes:

- More rapid population growth in London than in other parts of the country;
- Increased average income enabling the high fares on rail to be more affordable;

- Structural economic change leading to an increasing proportion of all employment falling within high income business and financial service occupations;
- A relative concentration in city centres in areas adjacent to rail terminals of the new office developments within which the business and financial service workplaces will locate; and
- Increased parking costs and restricted parking supply within central city areas.

12.4.9 The current NTEM only takes account of the first of these factors so it is not surprising that it has not captured the recent rapid growth in rail trips. In principle, using the results developed in this study the trip production models of NTEM could be extended by including income and NS-SeC variables to take account of the second and third factors.

12.4.10 However, the final two factors are also of major importance but necessarily fall completely outside the current framework of pure trip production models because by definition they relate to the relationship between the trip attraction and trip production zone. This is the province of more comprehensive travel demand models that combine the stages of trip production, distribution and mode split in order to provide a coherent behavioural framework to forecast the future spatial pattern of rail travel growth.

12.4.11 Models with these capabilities already exist, such as the LASER (WSP, 2002) model of London and the wider South East, which includes:

- segmentation by socio-economic group and car availability for commuting trips,
- distinction in business travel between managerial/professional movements (which may be attracted to rail) and travel by tradesmen (which normally wont);
- segmentation by income and car availability for other trip purposes;
- an explicit linkage between the location of workplaces by occupation type and local planning policies for commercial development;
- a service based representation of rail and London Underground supply (cost, time, frequency and capacity) throughout London and the wider South East, as well as networks for competing car, bus and slow modes.

12.4.12 In this way a travel demand model such as LASER that contains a full set of choice modelling stages and segmentation detail can represent all of the five factors above that are required in order to provide a sound behavioural framework for rail trip end growth forecasts.

12.4.13 In summary, a pure trip production model such as NTEM could be improved significantly through taking on board the extra segmentation detail by area type, income and NS-SeC that has been determined from the detailed estimation results of this study. Nevertheless, even when these improvements have been implemented, **the splitting of trips between rail and the set of other modes would remain as the weakest link within the trip production estimation process** unless there is a major shift in methodology to take full account also of trip distribution and mode split stages. This requirement to extend the trip end model to include these extra modelling stages is primarily needed for the estimation of rail trip ends for the reasons already explained in this Section. **This extended model functionality would not be necessary when splitting trip productions among the other more ubiquitous modes: car, bus and walk/cycle** for which a conventional trip end modelling approach is likely to be sufficient.

13 Analysis of Escort Trips

13.1 INTRODUCTION

13.1.1 This section describes the findings from modelling home-based escort trips. Escort trips, as given within NTS, are categorised into four groups: commuting, shopping plus personal business, education and others. Table 13.1 shows the number and percentage of escort trips by purpose, scaled by household and trip weights. It can be seen that the greatest number of escort trips are those for education which form 34.7% of total escort trips in the 2002-2006 NTS dataset.

13.1.2 It is clear from Table 13.1 that around one fourth of all escort trips are for shopping / personal business. This has been investigated further by splitting these by age. It shows that the largest proportion of shopping escort trips is by children, the majority of these presumably are accompanying their parents. These could be a conceptually different sort of trip from those we might categorise otherwise as escort trips. For this reason, in estimating commuting and shopping/personal business escort trips, children were excluded from the analysis.

Table 13.1 Number of escort trips by purpose and age group

Age Bands	Commuting	Shopping/Personal Business	Education	Employers Business/ Escort Home & Other Escort
0-16	2122	16039	7835	5196
16-65	11925	12162	33430	26445
65+	985	1959	1079	2918
All	15032	30160	42344	34599
%	12.3%	24.7%	34.7%	28.3%

13.2 HOME-BASED ESCORT TRIP ESTIMATION

13.2.1 As explained previously, the current NTEM has combined escort trips and the same purpose's home-based trips together as one group of home-based trips. For instance, when the number of home-based commuting trips is estimated in NTEM, the escort trips for commuting are embedded within it. This would essentially consider the characteristics of the escorters to be the same as those of the commuters themselves in influencing the number of home-based commuting trips. This approach could be criticised as not considering the characteristics of escortees, which are in most cases more important and are the main influences on the number of escort trips made.

13.2.2 This issue has been tackled in this research by considering the escort trips separately in order to test the effects of some extra variables, which are associated with the number of the potential escortees within a household. Table 13.2 shows the results of testing the additional variables with the exception of area types and the base socio-economic characteristics of the escorters.

Table 13.2 The main additional factors tested for escort trips in conjunction with the basic variables

Purpose	Extra Variable tested	Significance
Commuting	Number of workers in a household	Significant with a relatively large coefficient
Shopping/ Personal Business	Number of workers in household; number of children in a household	Significant with a relatively large coefficient
Education	Number of children by age (less than 6 , 6-11, 11-15)	Significant. All with relatively large coefficients
Employers Business /Escort Home & Other Escort	Number of workers in a household; number of children in a household	Number of workers was not significant; Number of children was significant with a relatively large coefficient

13.2.3 Figure 13.1 illustrates the findings from analysing each of four purposes. The variables which are highlighted in green are continuous variable for which direct comparison with the categorical variables within the model is not straightforward.

13.2.4 In general, it can be seen that no specific trend over time can be observed as the year specific variables H96xx are either not significant or do not show any trend of regular change (xx refers to years from 2003 to 2006).

13.2.5 The following are the main findings from these analyses:

- Number of workers within a household is an important factor to consider for commuting and shopping escort trips but not for other escort trips. An increase in number of workers tends to generate a decline in shopping but an increase in commuting escort trips.
- Number of children is an important factor to consider. Adults in households with a greater number of children make more escort trips for education, for shopping and for other purposes. In the case of education, this has been analysed by age, suggesting that those within families with younger children make significantly more education escort trips than other families.
- Except for “Education” and “Shopping&Personal Business”, area type is an important factor to be considered. For commuting and for other trips, those who are living in Inner London and Outer London make noticeably fewer escort trips than others.
- The effects of age and work status also vary noticeably within purposes. However, we conclude in general that young, not employed and part-time workers tend to make significantly more escort trips than other groups. This is presumably due to their available time or greater likelihood of having young children to escort.
- Other socio-economic variables (i.e. gender, household size and car ownership) also show different directions of effects for different travel purposes. For instance, males make more escort commuting but make fewer education and shopping escort trips.

13.2.6 In general, our findings show that escort trips should be considered separately from the main travel purpose and that the number of escortees within the household in each case can have at least as much importance as the characteristics of the escorters.

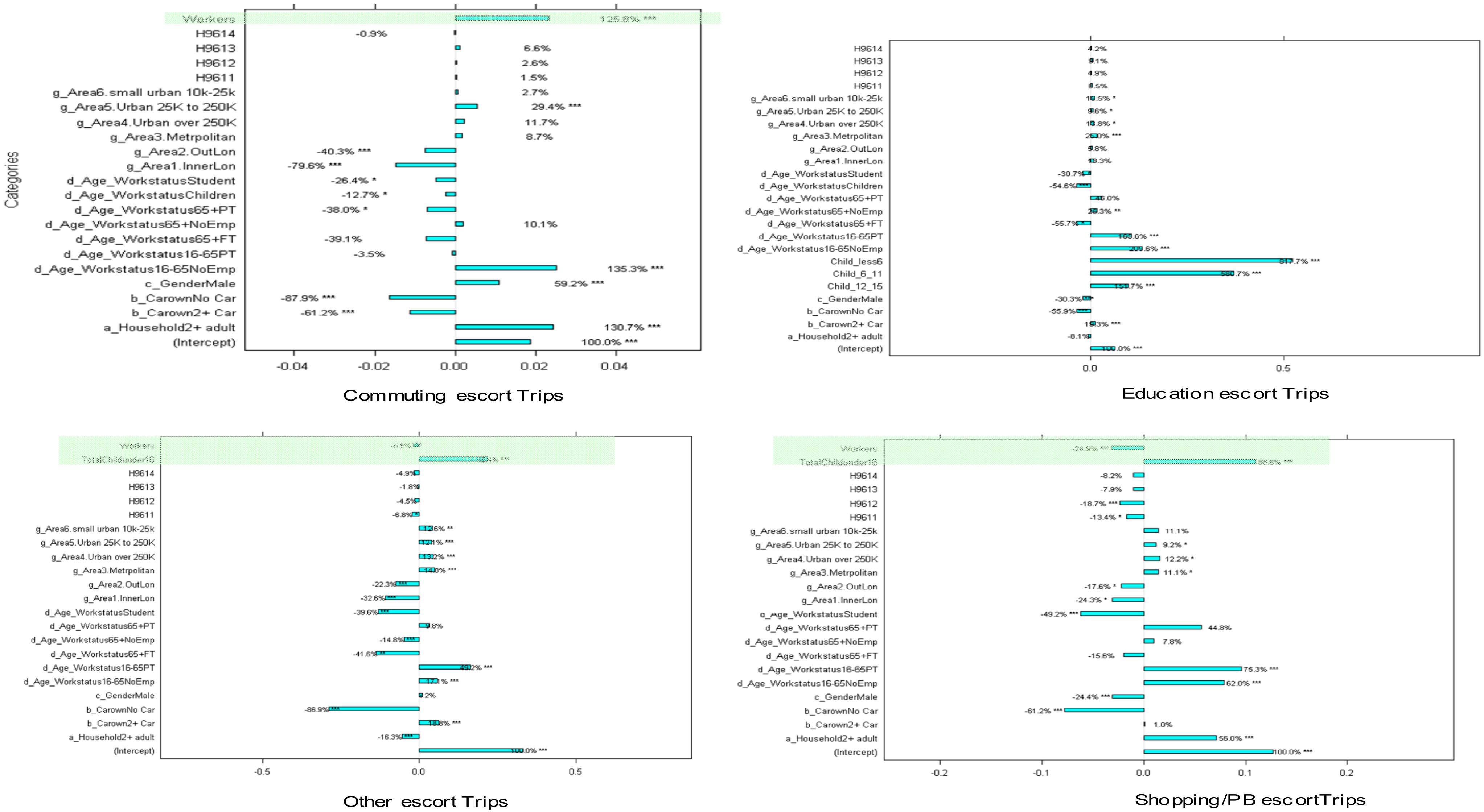

Figure 13.1 Results of modelling escort trips for four categories of trip purposes

`

14 Areas for Further Research and Development

14.1 INTRODUCTION

14.1.1 To enable researchers to examine in detail the large volume of trip rate model results that have already been produced in the course of this study, we have provided within Appendix E a comprehensive listing of the results obtained from estimating a number of model forms for each individual travel purpose. Although the main findings from our analyses of these models have been summarised in earlier Chapters, there are likely to be other findings of interest that could be derived through further examination of these model results by those starting from a different context or perspective.

14.1.2 This Chapter now outlines how some other ideas gained in the course of this study could be taken forward through further research and development activities. Firstly, a set of research tasks is outlined which would increase behavioural understanding of differences between different types of individuals or different types of contexts in their trip rates and more general travel behaviour.

14.1.3 This is followed by suggestions on how the results from this current study could best be used to update and improve the trip estimates in the NTEM that are published through TEMPRO.

14.2 SUGGESTED FURTHER RESEARCH

14.2.1 A number of topics have been identified where the understanding of the determinants of trip rates could be increased.

14.2.2 **Expand the analysis so as to estimate return home-based trips and non-home-based trip models** for each relevant trip purpose. These when combined with the outbound home-based trip and the escort trip models documented in this study would then provide a complete representation of all person movements (except those of series of trips movements - typically for deliveries - that are outside the scope of this work). This research would feed into the enhancements to NTEM that are discussed inn the next section.

14.2.3 **Investigate which factors are important in encouraging travellers living in one area type to make more or fewer trips than those living in other area types**. This will help to understand better the underlying factors which affect the variations in trip rates between area types. Examine accessibility variables in more detail by employing a GIS examination of reported trip origin and destination areas, to improve understanding of the alternative modes available and of their generalised cost in each area.

14.2.4 **Investigate the effect of socioeconomic and accessibility variables on travel distance / trip length, both by mode and overall**. Investigate the reason why average travel distance / trip lengths had increased consistently for some decades but appear to have stabilised in recent years. Is this likely to be a genuine behavioural change or is it largely a side-effect of methodological issues related to the survey data. One interesting feature of NB regression is that the overdispersion factor can be set to be a linear function of independent variables which can be estimated by R. This provides a potential means to work out which are the main factors within our model which most affect the variation of the dependent variable (i.e. trip rates) and so its standard error. In other words, which variables are the most responsible in increasing the overdispersion factor? This feature may prove helpful in distinguishing which accessibility and density variables might contribute most to explaining the differences in trip rates between area types.

14.2.5 **Investigate the possible interrelationship between trip rates across various purposes**. It has been found that people in higher income bands make fewer home-based commuting but more home-based business trips. The question is whether making more home-based business trips is one of the reasons for making fewer home-based work trips for this group of individuals. This could potentially be tested by adding both the number of HBEB trips and income into the model of HBW trips and then testing what would now be the relative importance and the directional effect of each income group? Likewise, the influence of non-home-based shopping trips on the number of home-based shopping trips should be tested, particularly to see whether this might explain the low rates of home-based shopping trips in Inner London.

14.2.6 **Investigate time budget influences**. It has been noticed that the people who are busier, such as full time workers and full time students, tend to make fewer shopping trips than those who have more free time, such as pensioners and those not in employment. It may also be expected that those with commuting trips of long duration might make fewer home-based trips during weekdays but perhaps more than average non-home-based trips. These effects could potentially be modelled by inclusion of the residual time available in the week or in a day as an explanatory variable or alternatively through imposing a constraint on the number of trips which can be made, based on their total cumulative duration.

14.2.7 It has been proved that NS-SeC in general is more important in defining trip rates than income. However, it is the NS-SeC and income of the individual that has been used in our tests. It would be of interest to test whether the models could be further improved through **inclusion instead of household income and / or of the NS-SeC of the head of the household.**

14.3 ENHANCEMENTS TO NTEM /TEMPRO

14.3.1 The current multi-modal version of NTEM was originally developed based on research and estimation methods at the start of this decade. Although the research with recent NTS data in the current study has shown that there is no substantial evidence that trip rates have changed significantly since the original NTEM, the improved estimation methods that are now adopted have enabled better estimates of trip production rates to be generated, through making use of a wider set of behavioural variables and improved model forms. The main changes that would need to be introduced to NTEM and TEMPRO are now summarised.

14.3.2 **Extra explanatory variables should be considered for inclusion**. In practice in order to avoid undue added complexity, it may be that not all of these extra variables are ultimately included. The grounds to support the decision to include a specific extra variable are:

- How easy is it to forecast this variable by year and zone into the future? Those variables that are difficult to forecast with any accuracy would prove problematic to

`

use. In this respect, the NS-SeC categories may prove less problematic than income categories as the former are already available at a spatially detailed zonal level from the 2001 Census. Furthermore, there are grounds for believing that spatial NS-SeC patterns only evolve relatively slowly in most areas, other than those few specific areas that might exhibit rapid gentrification.

- Is it a variable that has explanatory power for most trip purposes? The fewer the number of explanatory variables in total that need to be forecast the easier it will be to update the model.
- How great is its impact on trip rates? Because of the large sample size of the NTS as a whole, some of the variables that are statistically significant may not in practice make a noticeable numerical difference in the trip rates and so are candidates for exclusion.

14.3.3 **Separate out the escort trips by purpose from the other home-based trips** in order to improve the modelling of them both, as explained in Chapter 13.

14.3.4 **Improve modelling of trip attractions**. Because the NTS for the years 2002-2007 does contain the postcode of the trip destination, it would potentially be feasible to investigate ways in which trip attraction rates could be better modelled. In NTEM, the methodology for the estimation of trip attractions was developed prior to having such data available, so that trip attractions remain much less well estimated in NTEM than their companion trip productions. The other dataset of potential use for trip attraction modelling is the TRICS database of movements associated with different types of land use. This database has widened its scope in recent years so that an increasing number of its constituent surveys cover all modes of transport, rather than just road vehicle movements. Accordingly, for trip purposes such as shopping, for which TRICS has many surveys, it may now be the most suitable database on which to estimate trip attraction patterns within NTEM.

14.3.5 **Improve the split of trip productions by mode**. When setting up this study, specific resources were provided only for the investigation of rail mode trips, as their forecasts had proved problematic. Consequently, no investigations have been made of the wider question of improvements in the split across all of the other modes. Some further investigation of these other modes is likely to prove instructive in improving trip production forecasts in general.

14.3.6 **Improve the split of trip productions by rail.** Section 12.4 has presented a detailed discussion of the scope to improve the estimations of zonal rail trips and their forecasts through time. It shows that rail forecasts could be improved significantly through taking on board extra segmentation detail by area type, income and NS-SeC. However, without a further major shift in methodology to take full account also of trip distribution and mode split stages, the forecasts of the growth specifically of rail trips into the future though improved, would still remain somewhat uncertain.

15 Conclusions

15.1 INTRODUCTION

15.1.1 This Chapter summarises the main findings from this study of trip production rates. The aims of this research are to:

- Investigate the cause of the apparent decline in trip rates over time;
- Investigate whether the variables used in the current version of NTEM are valid for estimating trip rates;
- Provide advice on other factors which could improve estimation of trip rates in NTEM.

15.1.2 In general, the most important lessons which are learned are reported in each Chapter, while more detailed purpose by purpose charts of the estimation results for a variety of models can be found in Appendix E. These charts can be analysed and compared against each other in detail to develop a deeper understanding of the influences on trip rates.

15.1.3 The negative binomial regression approach used for this study is the most suitable method for modelling trip rates and the version available in the R package is not difficult to operate. It is a much more discriminating method for estimating influences on trip rates than the category analysis method used originally for NTEM in 1999. Also weighted rather than simple regression is used so as to make effective use of the household weights recently introduced into the NTS dataset to offset some of the effects of reductions in survey response rates.

15.2 KEY FINDINGS

15.2.1 The most important finding is that there is no convincing evidence of any substantial behavioural trend through time towards higher or lower trip rates per person for any travel purpose. Although summary analysis of the NTS trip rates appears to suggest a reducing trend in trip rates, more detailed analysis shows that this apparent trend is likely to be due to a mixture of other non-behavioural effects, including:

- Reductions in household response rates in the NTS, leading to a lower rate of representation of those people who make the most trips;
- Reductions in trips reported, especially short walk trips, probably in part due to the requirement to report the postcode of trip destinations;
- Changes in population profiles (e.g. a greater proportion of those in work being part-time workers).

15.2.2 This continuing validity of the original NTEM assumption that trip rates per person are reasonably constant through the years for each trip purpose implies that past forecasts produced by NTEM via the TEMPRO system should continue to be a reasonable representation of future travel demand levels.

15.2.3 Using the improved estimation methods, the individual variables which are in use in the current version of NTEM are shown to still have an important role in explaining trip rates for the majority of trip purposes. In addition, the inclusion of interactions between variables is shown to be important for some purposes, especially shopping. The interaction between car ownership and work status is the most important of the interaction terms to be considered in the model, in addition to the separate influence of each of these variables individually.

15.2.4 Regarding the effect of accessibility variables, the influences of area type and of population density are proved to be highly correlated. Between the two, area type plays a stronger role in explaining variations in trip rates. The other NTS accessibility variables, such as walking time to bus stops and to train stations, are not important for most travel purposes. These accessibility terms, however, are statistically significant in modelling trip rates by rail, suggesting that accessibility parameters play a bigger role in determining mode choice rather than in all mode trip production models.

15.2.5 The current NTEM model could be significantly improved by the inclusion of socioeconomic class and/or income variables. One interesting result is that income and socio-economic class are not highly inter-correlated within this model and act reasonably independently in influencing trip rates. However, if just one were to be selected for modelling purposes, the socio-economic class variable was shown to play a more important role than income group, for each individual trip purpose except shopping and visiting friends for which income is marginally better.

15.2.6 The splitting of trip productions between rail and the set of other modes remains as the weakest link within the trip production estimation process unless there is a major shift in methodology to take full account also of trip distribution and mode split stages. This requirement to extend the trip end model to include these extra modelling stages is primarily needed for the estimation of rail trip ends for the reasons explained in Section 12.4. This extended model functionality would not be necessary when splitting trip productions among the other more ubiquitous modes: car, bus and walk/cycle for which a conventional simpler trip end modelling approach is likely to be sufficient.

15.2.7 NTEM assumes that the same factors influence home-based trips and escort trips because it represents both within one common trip purpose. These assumptions have been tested, with results suggesting that this model could be improved by including a new term for the number of potential escortees within the household, along with the characteristics of the escorter. This suggests there may be benefit to replacing the current combined models of escort and non-escort trips used in NTEM by a pair of individual models of each.

16 References

Arentze, T. and H. Timmermans (2007). “Dynamic Model for Generating Multiday, Multiperson Activity Agendas: Approach and Illustration”, presented at the Annual Meeting of the Transportation Research Board, Washington D.C.

Barmby, T. and J. Doornik (1989). "Modelling trip frequency as a Poisson variable", *Journal of Transport Economics and Policy*, Vol. 23 (3), pp. 309-315.

Bartlett, R.S. and W.H. Newton (1982). “Goods Vehicle Trip Generation and Attraction by Industrial and Commercial Premises”, Transport Research Laboratory, TRL Report LR1059.

Berkhout, P. and E. Plug (2004). “A Bivariate Poisson Count Data Model Using Conditional Probabilities”, *Statistica Neerlandica*, Vol. 58 (3), pp. 349-364.

Bhat, C.R. and F.S. Koppelman (1999). “Activity-Based Modelling for Travel Demand", Handbook of Transportation Science, R.W. Hall (ed.), Kluwer Academic Publisher.

Boarnet, M.G., Nesamani, K.S. and S. Smith (2003). "Comparing the Influence of Land Use on Nonwork TripGeneration and Vehicle Distance Traveled: An Analysis using Travel Diary Data", Center for Activity Systems Analysis, Paper UCI-ITS-AS-WP-03-1. http://repositories.cdlib.org/itsirvine/casa/UCI-ITS-AS-WP-03-1

Cotrus, A.V., Prashker, J.N. and Y. Shiftan (2005), “Spatial and Temporal Transferability of Trip Generation Demand Models in Israel”, *Journal of Transportation and Statistics*, Vol. 8 (1), pp. 37-56. http://www.bts.gov/publications/journal_of_transportation_and_statistics/volume_08_number_01/html/paper_04/

Daly, A.J. (1997) "Improved Methods for Trip Generation". In: PTRC European Transport Forum, PTRC.

Daly, A.J. and S.P. Miller (2006). “Advances in Modelling Traffic Generation”, proceedings of the European Transport Conference 2006, Strasbourg.

Daly, A. (1982). "Estimating Choice Models Containing Attraction Variables", *Transportation Research Part B*, Vol. 16 (1), pp. 5-15.

Dasgupta, M., N. Raha and K. Sharman (1996). “Review of Trip Generation Studies”, Transport Research Laboratory, TRL Report 190.

Dey, S.S. and J.D. Fricker (1994). “Bayesian updating of trip generation data: Combining national trip generation rates with local data”, *Transportation*, Vol. 21 (4), pp. 393-403.

Ewing, R., Deanna, M. and S. Li (1996) “Land Use Impacts on Trip Generation Rates”, *Transportation Research Record*, Vol. 1518, pp. 1-6.

Fox, J., Daly, A. and H. Gunn (2003). “Review of RAND Europe’s Transport Demand Model Systems”, RAND Europe Monograph Report MR-1694-TRL. http://www.rand.org/pubs/monograph_reports/MR1694/

Goulias, K.G. and R. Kitamura (1989). “Recursive Model System for Trip Generation and Trip Chaining”, *Transportation Research Record*, Vol. 1236, pp. 59 – 66.

Greene, W. (2007). "Functional Form and Heterogeneity in Models for Count Data", *Foundations and Trends in Econometrics*, Vol. 1 (2), pp 113-218.

Golob, T.F. (1987). “Effects of Income and Car Ownership on Trip Generation: A Structural Equations Model and its Temporal Stability”, Institute of Transport Studies, UC Irvine, UCI-ITS-AS-WP-87-2.

Golob, T.F. (1999). "A Simultaneous Model of Household Activity Participation and Trip Chain Generation", Centre for Activity Systems Analysis paper UCI-ITS-AS-WP-99-3. http://repositories.cdlib.org/itsirvine/casa/UCI-ITS-AS-WP-99-3

Hupkes, G. (1982). ‘The Law of Constant Travel Time and Trip Rates’, *Futures*, Vol. 14(1), pp. 38-46.

Kitamura, R. and T. van der Hoorn (1987). “Regularity and Irreversibility of Weekly Travel Behaviour”, *Transportation*, Vol. 14 (3), pp. 227-251.

Koenig, J.G. (1980) “Indicators of urban accessibility: Theory and application”, *Transportation*, Vol. 9 (2), pp. 145-172.

Kouwenhoven, M., Rohr, C., Miller, S., Siemonsma, H., Burge, P. and J. Laird (2006). Isles of Scilly travel demand study, Cambridge, UK: RAND Europe, TR-367-CCC.

Larsen, O. I. (2003). “Estimating Independent and Simultaneous Trip Frequency Models for all Travel Purposes with Combined Logit/Poisson”, proceedings of the European Transport Conference 2003, Strasbourg.

Leake, G. R. and A.S. Huzayyin (1980) “Importance of accessibility measures in trip production models”, *Transportation Planning and Technology*, Vol. 6 (1), pp 9-20.

Miranda-Moreno, L.F. and M.E.H. Lee-Gosselin (2008). “A Week in the Life of Baby Boomers: How Do They See the Spatial_Temporal Organization of Their Activities and Travel?,” presented at the Annual Meeting of the Transportation Research Board 2007, Washington D.C.

Noland, R. B. and J.V. Thomas (2007). “Multivariate Analysis of Trip-Chaining Behaviour”, *Environment and Planning B: Planning and Design*, Vol. 34, pp. 953-970.

Ortúzar, J. de D. and L.G. Willumsen (2001). “Modelling Transport”, Published by Wiley, 3rd Edition.

Pendyala, R.M., Goulias, K.G. and R. Kitamura (1991). “Impact of Telecommuting on Spatial and Temporal Patterns of Household Travel”, *Transportation*, Vol. 18(4), pp. 383-409.

Polak, J. (2002) "Analysis of non-response in the LATS 2001 Pilot Household Travel Diary Survey" Paper presented at the 81st Annual Meeting of the Transportation Research Board, Washington DC. www.cts.cv.ic.ac.uk/documents/publications/iccts00045.pdf

Purvis, C.L., Iglesias, M. and V.A. Eisen (1996). “Incorporating Work Trip Accessibility in Non-Work Trip Generation Models in the San Francisco Bay Area”, *Transportation Research Record*, Vol. 1556 , pp. 37-45. http://www.mtc.ca.gov/maps_and_data/datamart/research/paper96.htm

Purvis, C.L. (1998). “Incorporating Land Use and Accessibility Variables in Travel Demand Models”, presentation at the ASCE Speciality Conference on Transportation, Land Use and Air Quality, May 1998. http://www.mtc.ca.gov/maps_and_data/datamart/research/ASCE_Paper_MAY98.pdf

Shafer, A. (2000). “Regularities in Travel Demand: An International Perspective”, *Journal of Transportation and Statistics*, Vol. 3 (3), pp. 1-32.

Schmöcker, J., Quddus, M.A., Noland, R.B. and M.G.H. Bell (2005). “Estimating Trip Generation of Elderly and Disabled People: Analysis of London Data”, *Transportation Research Record*, Vol. 1924, pp. 9-18.

Scott, D.M. and P.S. Kanaroglou (2002). “An Activity-Episode Generation Model that captures Interactions between Household Heads: Development and Empirical Analysis”, *Transportation Research Part B*, Vol. 36, pp. 875-896.

Sugie, Y., Zhang, J. and A. Fujiwara (2003). “A Weekend Shopping Activity Participation Model Dependent on Weekday Shopping Behaviour”, *Journal of Retailing and Consumer Services*, Vol. 10 (6), pp. 335-343.

Transport for London (2008) "London Travel Report, 2007"

WSP (2002) "LASER Enhancement Project: Final Report". Contract PPAD 9/65/71 for Department for Transport.
http://www.dft.gov.uk/pgr/economics/rdg/laserenhancementproject

Zahavi, Y. (1974). “Travel Time Budget and Mobility in Urban Areas”, Federal Highway Administration, Washington DC., U.S. Department of Transportation, May, NTIS PB 234 145.

Appendix A Mathematical Details of Poisson Probability Function

According to Equation 5-1:

$$P(y \mid n, \lambda/n) = \binom{n}{y}(\lambda/n)^y (1-\lambda/n)^{n-y} = \frac{n!}{y!(n-y)!}(\lambda/n)^y (1-\lambda/n)^{n-y}$$

Considering that we set n to be very large and by letting $p = \lambda/n$, then we have

$$\lim_{n\to\infty} P(y \mid n, p) = \lim_{n\to\infty} \frac{n!}{y!(n-y)!}(\lambda/n)^y (1-\lambda/n)^{n-y} = \lim_{n\to\infty} \frac{n!}{n^y (n-y)!}(\lambda^y/y!)(1-\lambda/n)^n (1-\lambda/n)^{-y}$$

Considering that $\lim_{n\to\infty}(1-\lambda/n)^{-y} = 1$ and $\lim_{n\to\infty}(1-\lambda/n)^n = e^{-\lambda}$ then we would have

$$\lim_{n\to\infty} P(y \mid n, p) = \lim_{n\to\infty} \frac{n!}{n^y (n-y)!}(\lambda^y/y!)e^{-\lambda}$$

If we call F= $\frac{n!}{n^y (n-y)!}$ then we can write that

$$\lim_{n\to\infty}(F) = \lim_{n\to\infty}\left(\frac{n(n-1).....(n-y+1)}{n^y}\right) = 1$$

$$\therefore \lim_{n\to\infty} F = 1$$

So we would have $\lim_{n\to\infty} P(y \mid n, p) = (\lambda y/y!)e^{-\lambda}$

Appendix B Cameron and Trivedi Test of Over Dispersion

`

Overdispersion is one potential source of error when using the Poisson regression.

The Cameron and Trivedi test aims to evaluate whether the value of the variance differs significantly from that of the mean of data. It is based on the assumption that under a Poisson model, $(y_i - E[y_i])^2 - E[y_i]$ has a Zero mean (i.e. the mean and variance of the dependent variable are the same).

The null and alternative hypothesis can then be defined as:

$H_0 : Var[y_i] = E[y_i]$ and $H_A : Var[y_i] = E[y_i] + \alpha g(E[y_i])$

where $g(E[y_i])$ is most often given values of $E[y_i]$ and $E[y_i]^2$

The simple linear regression model in which Z_i is regressed on W_i should be estimated (i.e. Z_i = bW_i). If b happens to be statistically significant in either the case of $g(E[y_i]) = E[y_i]$ or $g(E[y_i]) = E[y_i]^2$, then H_0 is rejected.

The definitions used are:

$$Z_i = \frac{(y_i - E[y_i])^2 - y_i}{E[y_i]\sqrt{2}}$$

$$W_i = \frac{g(E[y_i])}{\sqrt{2}}$$

Appendix C Negative Binomial Regression

`

Poisson regression makes a strong assumption that the variability of counts within a covariate group is equal to the mean (i.e. λ_i). If this fails to be true, the estimates of the coefficients can still be consistent using Poisson regression, but the standard errors can be biased and they will be too small. Negative Binomial is an extension to Poisson regression, which can account for greater than Poisson variation and is based on the negative binomial distribution.

As explained in section 5.5, the negative binomial distribution arises as a continuous mixture of Poisson distributions where the mixing distribution of the Poisson rate is a gamma distribution. Formally, this means that the mass function of the negative binomial distribution can also be written as:

$$P(y;r,p) = \int_0^\infty Poisson(y \mid \lambda).Gamma(\lambda \mid r,(1-p)/p)d_\lambda = \frac{\Gamma(r+y)}{y!\Gamma(r)} p^r (1-p)^y$$ Equation 16-1

$\frac{\Gamma(r+y)}{y!\Gamma(r)} p^r (1-p)^X$ is actually the probability mass function of a random variable with a NegBin(r, p) distribution. The mean of the distribution is:

$$Mean = \lambda = r\frac{1-p}{p}$$ Equation 16-2

And the variance would be equal to

$$Var = \lambda + (1/r)(\lambda)^2 = r\frac{1-p}{p^2}$$ Equation 16-3

As can be seen, this distribution would allow the variance to be different from the mean. In most statistical text books and packages including R statistical package, $\theta = 1/r$ is called the overdispersion factor and should be estimated in the estimation routine. As is shown below, when θ approaches zero and so the variance is the same as the mean, the mass function of the negative binomial distribution would be the same as that of the Poisson distribution, proving that Poisson is the limiting case of negative binomial.

By substituting $\lambda = r\frac{1-p}{p}$ and so $p = \frac{r}{r+\lambda}$, the mass function becomes

$$P(y;\lambda,r) = \frac{\Gamma(r+y)}{(r+\lambda)^X \Gamma(r)} * \frac{\lambda^X}{y!} * \frac{1}{(\frac{r+\lambda}{r})^r}$$ Equation 16-4

In the same way as Poisson, $E[y_i] = \lambda_i = EXP(\beta X_i)$ should be replaced in the probability function of the negative binomial and both parameters of β and r should be estimated jointly in a way to maximize the resulting log likelihood function.

Appendix D Interpreting the Charts of NB Regression Results

The results of the regression tests in this report and in Appendix E are shown in the format of bar-charts. This Appendix describes how to interpret these charts and provides some background on how the charts have been produced.

Detailed information on the estimated parameters from a Negative Binomial regression is output by the R package in the form illustrated in Figure 16.1.

```
summary of outputs

Call:
glm.nb(formula = formula1, data = .nb.data, weights = W2, init.theta = 1.19479560664939,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1375  -1.0859  -0.3038   0.2976   5.9407

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                   0.350024   0.016445  21.285  < 2e-16 ***
a_Household2+ adult           0.003854   0.011884   0.324   0.7457
b_Carown2+ Car               -0.084218   0.009547  -8.821  < 2e-16 ***
b_CarownNo Car               -0.098465   0.011754  -8.377  < 2e-16 ***
c_GenderMale                 -0.104865   0.008479 -12.368  < 2e-16 ***
d_Age_Workstatus16-65NoEmp    0.556359   0.012310  45.196  < 2e-16 ***
d_Age_Workstatus16-65PT       0.272598   0.013898  19.614  < 2e-16 ***
d_Age_Workstatus65+FT         0.128905   0.072532   1.777   0.0755 .
d_Age_Workstatus65+NoEmp      0.579169   0.012467  46.455  < 2e-16 ***
d_Age_Workstatus65+PT         0.479675   0.047218  10.159  < 2e-16 ***
d_Age_WorkstatusChildren     -0.801819   0.013309 -60.245  < 2e-16 ***
d_Age_WorkstatusStudent      -0.126145   0.023955  -5.266 1.40e-07 ***
g_Area1.InnerLon             -0.171245   0.023284  -7.355 1.91e-13 ***
g_Area2.OutLon               -0.013841   0.017061  -0.811   0.4172
g_Area3.Metrpolitan           0.104288   0.014293   7.296 2.95e-13 ***
g_Area4.Urban over 250K       0.125914   0.013811   9.117  < 2e-16 ***
g_Area5.Urban 25K to 250K     0.116493   0.012187   9.559  < 2e-16 ***
g_Area6.small urban 10k-25k   0.123697   0.015396   8.034 9.42e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.1948) family taken to be 1)

    Null deviance: 112107  on 94631  degrees of freedom
Residual deviance:  98787  on 94614  degrees of freedom
AIC: 317432

Number of Fisher Scoring iterations: 1


              Theta:  1.1948
          Std. Err.:  0.0111

 2 x log-likelihood:  -317393.7110
Wald test to test model fit
Wald test

Model 1: SumOfTrips ~ a_Household + b_Carown + c_Gender + d_Age_Workstatus +
    g_Area
Model 2: SumOfTrips ~ 1
  Res.Df    Df  Chisq Pr(>Chisq)
1  94614
2  94631   -17 8751.5  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Testing significancy of coeficients using robust variance covariance matrix

z test of coefficients:

                               Estimate Std. Error  z value  Pr(>|z|)
(Intercept)                   0.3500240  0.0180727  19.3675 < 2.2e-16 ***
a_Household2+ adult           0.0038542  0.0140289   0.2747 0.7835230
b_Carown2+ Car               -0.0842184  0.0104408  -8.0663 7.246e-16 ***
b_CarownNo Car               -0.0984646  0.0149904  -6.5685 5.082e-11 ***
c_GenderMale                 -0.1048650  0.0098327 -10.6649 < 2.2e-16 ***
d_Age_Workstatus16-65NoEmp    0.5563595  0.0132169  42.0946 < 2.2e-16 ***
d_Age_Workstatus16-65PT       0.2725977  0.0147425  18.4906 < 2.2e-16 ***
d_Age_Workstatus65+FT         0.1289053  0.0825848   1.5609 0.1185509
d_Age_Workstatus65+NoEmp      0.5791687  0.0132365  43.7555 < 2.2e-16 ***
d_Age_Workstatus65+PT         0.4796748  0.0419513  11.4341 < 2.2e-16 ***
d_Age_WorkstatusChildren     -0.8018187  0.0183024 -43.8095 < 2.2e-16 ***
d_Age_WorkstatusStudent      -0.1261453  0.0349191  -3.6125 0.0003033 ***
g_Area1.InnerLon             -0.1712454  0.0323662  -5.2909 1.217e-07 ***
g_Area2.OutLon               -0.0138410  0.0200451  -0.6905 0.4898854
g_Area3.Metrpolitan           0.1042877  0.0154937   6.7310 1.685e-11 ***
g_Area4.Urban over 250K       0.1259139  0.0148547   8.4764 < 2.2e-16 ***
g_Area5.Urban 25K to 250K     0.1164933  0.0126645   9.1984 < 2.2e-16 ***
g_Area6.small urban 10k-25k   0.1236972  0.0160420   7.7108 1.250e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 16.1 Typical output from running simple shopping trip rates

The most important outputs highlighted in Figure 16.1 can be described as follows:

- Tables numbered 1 and 4 are the results of estimating various categorical and continuous variables. The columns from left to right are:
  – the estimated value of the coefficient;
  – its standard error;
  – its z score (or t ratio), calculated from the ratio of the coefficient to its standard error;
  – the probability that the coefficient is not significantly different from the assumed null value, zero;
  – the significance level code as indicated at the foot of the table. For categorical variables it denotes whether the specified category has an impact that is significantly different from the impact associated with the omitted reference category for that variable.
- The difference between Tables 1 and 4 is that a robust variance covariance matrix is used in the latter, which is a more conservative approach for estimating error terms. There is no difference between the two approaches in their estimated coefficients but only in their error estimates. To be on the safe side, the results based on the robust standard errors in Table 4 are used throughout this Report when creating the bar charts and interpreting results
- Tables numbered 2 and 3 show the overall measure of goodness of fit (AIC) of the model and the overdispersion factor (θ), respectively.

Although the estimation theory applies equally to continuous and categorical variables, in practice the analysis of trip rates is largely concerned with category variables. This has some important consequences for the interpretation of the coefficients, which we now discuss.

Whereas with a continuous variable we typically have a single estimated coefficient, indicating the (linear) effect of changes in the variable, for category variables we expect a separate coefficient for each "level". However, because of linear dependence, it is only possible to estimate (n–1) coefficients for a category variable with n levels. It is therefore standard practice to designate one of the levels as the "base", and to specify the estimated coefficients for the other levels so that they represent the effect of moving from the base to a different level. This is done by means of specifying the independent variables as "dummy" (0/1) variables for all levels other than the designated base.

When the equation contains multiple category variables, then the base levels of each category contribute to the overall model constant, referred to as the "Intercept". Thus in order to interpret the intercept, it is necessary to understand how the base levels are defined.

The estimated coefficient (β_i^k, say) for any category variable k at level i measures the effect of that level relative to the base. Since the model is multiplicatively defined, it increases the base level by a factor of $\exp(\beta_i^k)$, and since, for small values, $\exp(\beta_i^k) \approx 1 + \beta_i^k$, it follows that, for example, a coefficient value of +0.1 implies an increase in trip making of approximately 10%.

`

The example shown contains 5 category variables: “household”, “car ownership”, “gender”, “age/workstatus”, “area”. By reference to the missing levels of each category, we can define the base to which the intercept coefficient refers – this is: persons in households with 1 person, 1 car, female, aged 16-65 in full time work, and living in a Rural area with the population of less than 10K. The intercept coefficient is 0.35, so that the average number of weekly trips for this base category is exp (0.35) = 1.42.

For a person who was male but otherwise identical in base categories, the coefficient “c_GenderMale” of –0.1049 indicates an average trip rate of approximately 10% lower: the average number of weekly trips for this category is exp (0.35 –0.1049) = 1.28.

For presentation purposes, we have converted the information to charts. The resulting chart in Figure 16.2 demonstrates the impact of various factors on trip rates. It presents two components: percentage change and significance code. The former measures the relative impact of the specified category on the trip rate, compared to the impact of the omitted reference category for that variable. The latter measures the statistical significance of that category of the variable within the model.

The percentage change is calculated based on the estimated coefficient for that category. It denotes the difference between any category of one categorical variable from the reference category of that variable, which is omitted from the model (i.e. has the coefficient of zero and so its effect is reflected in the intercept). This difference is conditional on all other categorical variables being set to their reference value and all continuous variables included being set to the value zero[5]. Equation 16-5 shows how the percentage difference is calculated.

$$Per = \left(\frac{e^{\beta_i + Intercept}}{e^{Intercept}} - 1\right) * 100 \qquad \text{Equation 16-5}$$

Where Per is the percentage difference and β_i is the coefficient of variable i and $Intercept$ is the intercept value estimated for the model.

The reference category will not be shown in the charts/models as it is assumed that the reference has a zero coefficient and its effect is embedded within “constant”. The coefficient of all other categories of one variable does actually show how different they are compared to the base one. Whether this difference is statistically significant is shown by the significance codes. Percentage here shows how different each category is compared to the base one when all other variables are constant. For instance in Figure 16.2, the variable “b_Carown” has three categories as “No Car”, “1 Car” and “2+ Car”. “1 Car” is chosen as the reference and omitted from the chart. This chart then shows both other categories are significantly different from “1 Car” and also that travellers in “2+ Car” are making 8.1% fewer trips and those in “No Car” are making 9.4% fewer trips than those in “1 Car” when everything else is constant.

[5] If the continuous variables cannot legitimately be set to zero, then the interpretation of percentage change will be more complex. In the models estimated in this study, all continuous variables can be set the value of zero.

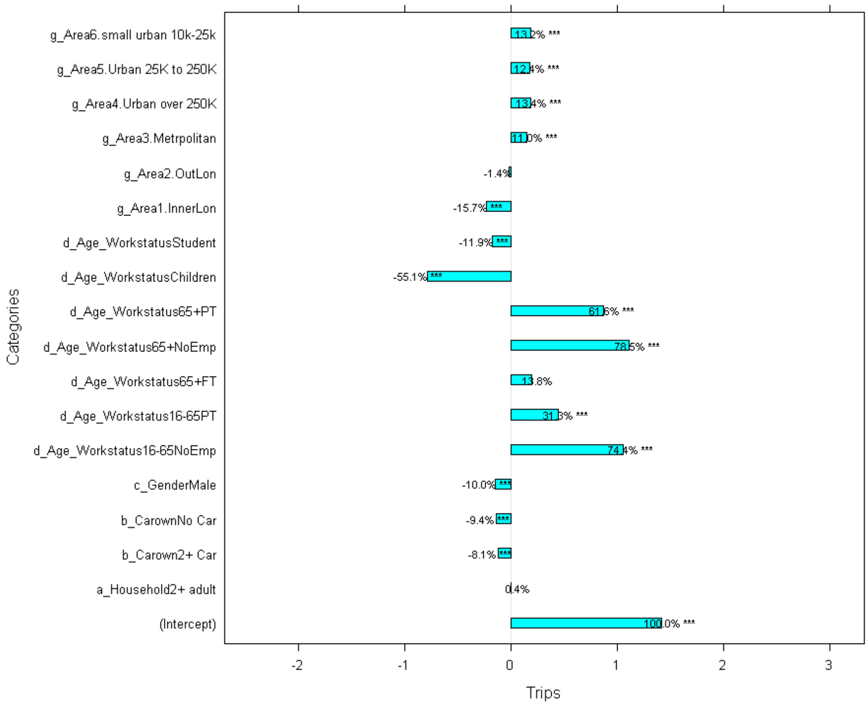

Figure 16.2 Typical Chart output from running the simple shopping trip rates

Percentage change, however, does not provide a meaningful interpretation for continuous variables. To explain this, assume that two variables of "age" and "Car Ownership" are modelled as continuous variables in a model. The coefficient of the variable "age" which has the range of values from 0 to 100 cannot be directly compared with that of "car ownership" which lies within the range of 0 to 10. In this case, we need to take account of the scale of the continuous variable

It should be noted that sometimes some categories which show a big percentage difference are not reported as being significant. This happens when the standard error associated with that variable is large because of a large variance in the behaviour of this population. This can happen when the size of population associated with that specific group is small or simply if the people in that group behave heterogeneously which suggests that the differentiation by that group is not useful.