Traffic Demand and Land-Use in the UK: An Econometric Analysis Using the TRICS Database

David C. Broadstock
Surrey Energy Economics Centre
University of Surrey
Guildford
UK
June 24, 2008
Abstract

The main research question motivating the research in this thesis is whether or not land zone placement (as defined in accordance with Planning Policy Guidance definitions) is a genuinely important determinant of trip generation behaviour. In order to answer this question three unique data sets have been assembled. The key source is the UK Trip Rate Information Computer System (TRICS), particularly the site information relating to (i) Office Developments, (ii) Food Superstores and (iii) Residential Developments. These site types were chosen for scrutiny as they have for various reasons been the focus of enduring media, environmentalist and Government scrutiny in the light of the increasingly mainstream acceptance of sustainable development principles.

For Office Developments and Food Superstores single equation trip attraction models, based on the tenets of a standard derived demand modelling framework, are estimated; whereas for Residential Developments a system is estimated encompassing the generation of trips in residential sites and how they interact with levels of car ownership. Due to recurrent small sample problems and issues of heteroskedasticity all models are estimated using a semi-parametric regression model.

The results support the contention that land zone features, as a group of indicators, should be accounted for in trip generation models for office developments. For Food Superstores household economies of scale and scope are identified for individuals visiting these sites and public transport services are also not found to be kinked with reductions in car traffic. Conversely, the elasticity of public transport provision is significant and positive. Household trip generations are primarily determined by car ownership levels and that car ownership is in turn primarily determined by
employment levels. It is found that land zone placement is a valid extension to the trip generation model.

Key Words;

Trip Generation, Car Ownership, Traffic Impact, Traffic Demand, TRICS, Environmental Impact, Urban Design, Car Use, etc Retail, Office employment, Residential, Traffic Modelling, Sustainability.
Acknowledgements

This thesis is dedicated with love to the memory of my father, Peter Charles Broadstock, who was sadly diagnosed with, and passed away from cancer during my time at Surrey. He was a truly understated person whom I owe more than I will ever be able to repay.

Thanks must also go to my mother, Daphne, my two brothers, Robert and William, and my sister, Deborah for support over the years. In particular, if it wasn’t for my mum helping me with my first piece of business studies homework in my GCSE I may have never found the desire to truly expand my knowledge. Thanks mum...

There are a great many other people who I would like to thank for many different reasons, too numerous to mention. Aynur has been a pillar to me for the last three years, particularly through my father’s cancer, and this will never be forgotten. Julie has been such a kind and generous friend, and Lily has made it all fall into perspective.

Unfaltering support from Lester, Alan and Colin has allowed me to develop a long way from when I began this thesis, both personally and academically.

Support from the Economic and Social Research Council (Project Number PTA-035-2004-00035) is gratefully acknowledged, as is permission from JMP Consultants Limited and the UK TRICS Consortium, for use of the TRICS database.
Contents

1 Introduction ........................................... 12
   1.1 Aims of the research ............................ 13
   1.2 Urban Design, Transport planning and Smart growth .... 15
   1.3 The Trip Rate Information Computer System .......... 21
   1.4 The Research Questions .......................... 24
   1.5 Structure of the Thesis .......................... 27

2 Planning and Policy Context ......................... 28
   2.1 Introduction .................................... 28
   2.2 Traditional responses to traffic demand ............ 31
      2.2.1 Road Capacity, Junctions and Intersections .... 32
   2.3 Contemporary responses to traffic demand .......... 37
      2.3.1 The True Costs of transport ................. 40
      2.3.2 Traffic Impact Analysis ..................... 43
      2.3.3 Transportation Demand Management (TDM) ....... 48
      2.3.4 TDM and Government Guidance Notes .......... 51
      2.3.5 Soft factors versus Hard Factors ............ 54
   2.4 Summary of Planning and Policy Context ............ 55
   2.5 Appendix 1 - Soft factors versus hard factors .... 58
      2.5.1 Introduction ............................... 58
3 Modelling Traffic Generation

3.1 Introduction ........................................ 72
3.2 Trip Modelling ........................................ 73
   3.2.1 Microeconomic Foundations .................... 74
   3.2.2 Empirical Modelling Approaches ............... 77
   3.2.3 Summary of Trip Modelling Approaches .......... 106
3.3 Car Ownership Literature .......................... 108
   3.3.1 Microeconomic Foundations .................... 109
3.4 Empirical Modelling Approaches .................. 115
   3.4.1 Pure Car Ownership; Time-Series Models ........ 115
   3.4.2 Models of Car Ownership and Use ............. 125
   3.4.3 Summary of Car Ownership Modelling Approaches ... 132
3.5 Overall Summary of Modelling Chapter ............ 134

4 Data and Methodology ................................. 136
4.1 The Data ............................................ 136
   4.1.1 The TRICS database ........................... 137
   4.1.2 Office developments data ..................... 150
   4.1.3 Food Superstore Data .......................... 154
   4.1.4 Residential developments data ................. 158
4.2 Methodology ........................................ 161
   4.2.1 The general semi-parametric model .......... 162
   4.2.2 The Bootstrap procedure. ..................... 166
   4.2.3 The Achieved Significance Level (ASL) ....... 167
CONTENTS

5 Office Block Trip Generation Models ...................................... 171
   5.1 Introduction .................................................................... 171
   5.2 Modelling Strategy and Estimation .............................. 174
     5.2.1 Missing Data and Imputation Methods .................. 176
   5.3 Results ................................................................... 192
     5.3.1 Model Cross Comparison: Preferred imputation method and
          overall model specification ......................................... 196
   5.4 Concluding Remarks ................................................ 200

6 Trip Generation Models at Food Superstores .......................... 205
   6.1 Introduction ................................................................ 205
   6.2 Modelling Strategy and Estimation .............................. 208
   6.3 Results ................................................................... 210
   6.4 Food Superstores and ‘Food Deserts’ ......................... 222
   6.5 Concluding Remarks ................................................ 225

7 Residential Site Car Ownership and Use ............................. 227
   7.1 Introduction ................................................................ 227
   7.2 Modelling Strategy and Estimation .............................. 230
   7.3 Results ................................................................... 234
   7.4 Summary and Conclusions ......................................... 245

8 Summary and Conclusions of the Thesis ............................. 247
   8.1 The Research Questions Revisited .............................. 248
     8.1.1 Policy questions .................................................. 254
   8.2 Closing comments and discussion ............................... 257
List of Figures

1.1 The Four Step Model (FSM) ........................................ 17
1.2 Research Structure: A Conceptual Framework ................. 24

2.1 The demand for petroleum motor spirits and diesel fuel in the UK 1960-2005 (Data source: UK Department for Transport and Office of National Statistics online data collections) ................................................................. 29
2.2 Total arrivals and departures for an average Friday at a commercial warehouse (TRICS 20004b) .............................. 34
2.3 Junction choice graph from Rogers (2003) .................... 36
2.4 UK Travel and Economic Activity 1960-2004. (Data source: UK Department for Transport and Office of National Statistics online data collections) ................................................................. 39
2.5 Economists' view of Transport Planning in 1982 ............ 52
2.6 Policy implementation - A precursor. ........................... 63
2.7 Effect of soft factor policy intervention on target group. . 65
2.8 Effect of hard factor policy intervention on target group. . 66
2.9 Sequential rebound within a spatial region. .................... 68
2.10 Effect of policy interventions on non-target groups. ....... 69
LIST OF FIGURES

3.1 Car ownership and total distance travelled in the UK. *(Data source: UK Department for Transport and Office of National Statistics online data collections)* .......................... 108

3.2 Example of a Logistical Curve of Car Ownership ............................ 117

3.3 General approaches to transport modelling, and the TRICS database ................................. 135

4.1 TRICS location definitions ....................................................... 143

4.2 Total car traffic arrivals by time of day - Offices ........................... 152

4.3 Total car traffic arrivals by time of day - Food Superstores .................. 157

4.4 Total car traffic arrivals by time of day - Residential developments ................................. 161

5.1 Employee proximity to Office development .................................. 174

5.2 MaxiMin tables of results ......................................................... 198

5.3 Sensitivity of trip rate estimates to changes in imputation method used .......................... 201

6.1 Effects of Key Variable Changes upon Food Superstore Traffic Flow ......................................................... 218

6.2 Estimated Elasticities on Policy Variables, Models I(a), I(b) and I(c). ......................................................... 220

6.3 Estimated Elasticities on Land Zone Accessibility, Models I(a), I(b) and I(c). .......................... 221

7.1 UK Car Ownership and Use 1952-2005. *(Data source: UK Department for Transport online data collection)* .......................... 228
List of Tables

1.1 TRICS research reports .................................. 23

2.1 The Full Range of Costs associated with Transport Use. .... 41

3.1 Microeconomic properties of car ownership ................. 112

4.1 Descriptive Statistics for Office Data .................... 151
4.2 Descriptive statistics for Food superstore data ............ 154
4.3 Descriptive Statistics for Residential Data ................. 158
4.4 Dichotomous variable classification schema ................. 159

5.1 Spec I results: $T_{m,o,d} = f(\delta, \gamma, \lambda)$ ............ 193
5.2 Spec II results: $T_{m,o,d} = f(\delta, \gamma)$ .................. 194
5.3 Spec III results: $T_{m,o,d} = f(\delta, \lambda)$ ................. 194
5.4 Spec IV results: $T_{m,o,d} = f(\delta)$ ....................... 194

6.1 GFA Model Estimates .................................. 213
6.2 RFA Model Estimates .................................. 214

7.1 Single-equation model results ............................ 236
7.2 Kernel Distributions of ‘Household Type’ variables from Uni-
    variate Car Use Model .................................. 237
7.3 Two-Stage Least Squares model results .................... 240
8.1 Results for questions SQ1, SQ2 and SQ3 251
Chapter 1

Introduction

Recent years have seen massive changes in the way development control and transport planning is conducted in the UK, other developed economies and also in many other Less Developed Countries (LDCs). Many economies have been keen to follow in the UK’s footsteps, see for example Rabinovitch (1996), and in the light of rapid growth for instance in the Chinese and Indian economies as well as the expected associated traffic impact\(^1\), the need to explore all available empirical evidence is obviated.

In the past transport planning was very much perceived (and practiced) as a tool used primarily to ‘predict and provide’ required road space for expected traffic flows, particularly for new developments. However, in recent years (the past two decades or so), planners, as well as other stakeholders involved in the overall planning process (i.e. network users), have begun to realise the necessity to consider the social

\(^1\)Dargay, Gately & Sommer (2006) for instance predict that car ownership could rise in China from present levels of around 50 million, to a staggering 400 million by the year 2030. The associated externalities are not directly predicted, but it is obvious that, in the absence of careful planning and management, this is not good news for the environment.
impacts of their actions, resulting in a shift towards Transit Oriented Development (TOD) and Travel Demand Management (TDM). The precursor for this shift in behaviour of transport planning practitioners has come about through the more obvious effects of climate change which have been largely related to increased energy use, and particularly petroleum demand. Evans (2006) provides a satirical overview exemplifying the attention that climate change has received in recent media.

The majority of activities in the modern developed/Western economies of the world are made viable with the aid of non-renewable (finite) energy sources, such as Coal, Gas, Nuclear and (particularly for the Transport industry) Oil. Although it has been recognised for a long time that these fossil based and/or non-renewable sources of energy are finite in supply, it has only become a prominent issue over the last decade or so that energy reserves are beginning to dwindle away. It is unclear exactly how long oil reserves will last, but this question is now largely overshadowed by how transport can be sustained with renewable energy sources, and also how overall transport demand can be more efficiently managed.

1.1 Aims of the research

Given the discussion above, this thesis uses economic concepts and econometric techniques to better understand demand for car use using a nationally recognised dataset, namely the Trip Rate Information Computer System (TRICS) database. This analysis is conducted with a view to providing empirical guidance on helping to account for variations in key characteristics of vehicle traffic as well as to assess the TRICS database more rigorously than has hitherto been done. In particular, 

\footnote{Although the fuel price crises in the 1970's provide earlier evidence of the importance of fuel prices to the economy.}
CHAPTER 1. INTRODUCTION

the focus of this research is to contribute to the empirical analyses of land-use-traffic linkages and interactions, in so doing, supplementing the existing corpus of literature on activity related trip making behaviour. The breadth and depth of the TRICS database provides scope for considering geographical factors in arguably a more consistent manner than many earlier studies on this theme.

These geographical features are examined in several different ways. Firstly a systematic review of traffic demand at various sub land-use types is conducted for Supermarkets, Office developments and Housing, to establish differences in the determinants of traffic demand. Secondly the effects of geographical location type, for example ‘town centre’ or ‘out of town’, for each land-use type are featured into the traffic demand models. These issues have been recognised and considered in the past, albeit employing a range of widely differing methodologies and dimensions to datasets, see for example Boarnet, Nesmani & Scott-Smith (2003).

This is conducted in a series of ‘activity-based’ derived demand models, which are formalised in Chapter 4.2 (and Chapters 5, 6 and 7), though in short, the level of trips to a site is derived from the desire to participate or purchased the good or service which that site offers (McNally (2000), Vovsha, Petersen & Donnelly (2003)). Essentially it is considered that there are a number of demand inducing factors, which should generally reflect the production and attraction side of the demand for a trip. Production side induction factors will include variables which broadly affect the ability to consume, as determined through microeconomic demand relationships. Attraction side induction factors will involve site characteristics such as floor space, employees, parking provision etc.\textsuperscript{3}

\textsuperscript{3}The models developed in this thesis focus purely upon vehicle traffic, particularly passenger vehicles (as opposed to public service or delivery vehicles). This decision to focus in car trips only is rationalised in Chapter 4. Multi-modal transport modelling and/or solutions are not considered
1.2 Urban Design, Transport planning and Smart growth

In analysing transport it is prudent to identify some of the key concepts that are faced by transport practitioners and urban landscapers, given the intrinsic link between the two. The following section introduces some of these concepts and provides a very broad overview to indicate how they are interlinked, and therefore implicitly considers some of the linked externalities associated with transport and development.

Urban Design

Urban design and landscaping is one of the many roles which the modern transport planner must fulfil, and it is an integral part of designing and planning for sustainable communities, through 'smart growth' etc. The concept of sustainable communities is a complex one which is given further attention in Chapter 2, however the crux of the concept is the development of communities which are, or at least on the way to, becoming independent of non-renewable energy sources. Developments should be built into the existing fabric of society, not merely placed on top of it, so as to ensure integration, and to this end, a moderate understanding of urban design is a pre-requisite in contemporary planning.

Urban design is subsequently highly co-related with land-use, and there exist numerous (though not entirely extensive) analyses of the interaction between land-

---

4With carbon emission management being fundamentally critical in such communities
CHAPTER 1. INTRODUCTION

use and the demand for transportation/travel, see for example Banister, Cullen & Mackett (1990) or Litman (2004). Many of these studies consider interactions at the level of the city (as opposed to individual sites), such that planners consider the way the distribution of land-use patterns across a wider geographic region affect the flows and dynamics of transportation for that area, i.e. mode choice, trip route, trip length etc.

Transport Planning

Transport planning is increasingly a combined function of urban design, though further encapsulates downstream (ex-post) travel demand management. A comprehensive discussion of the role of transport planning is offered in section 2, what is pursued here is a brief introduction to the Four Step Model (see Figure 1.1), which encapsulates the key aspects of fundamental transport planning. McNally (2000) presents a retrospective discussion of the overall transport modelling process.

The overall aim of a transport model is to attempt to define and predict four fundamental choices made by individuals:

- Whether or not to make a trip?
- What route to take across the existing transport network?
- What mode of travel to go by?
- Are the choices made in the previous questions the optimal set of choices available?  

\footnote{The fourth of these issues is likely to be more of a subconscious decision than the other three for the individual, though is arguably the most important for the analyst. This however is a subjective comment, as should hopefully become apparent in the following discussion.}
It is possible to consider these four issues as quite individual elements. Taking each of the different elements in turn;

The choice ‘whether or not to make a trip?’ relates to the analysis of trip generations or attractions and forms the first stage in the general Four Stage Model (FSM). The FSM is widely accepted in the transport modelling literature, though is not deemed to be due to any particular author(s), nor is it considered an exact solution, merely a framework McNally (2000).
CHAPTER 1. INTRODUCTION

The decision of *which route to take*? is a further key issue in the analysis of journeys. Route choice is a fundamental determinant of congestion levels on transport networks and consequently affects capacity choices and implementation of traffic reduction schemes. Congestion in return is a key determinant of route choice and is a by-product of the overall level of trips generated in a zone. As a result, there is a generally accepted sequential nature in the analysis of individuals transport decisions, i.e. in order to understand the distribution of travellers it is necessary first to know how many people are travelling, where they are travelling from (origin) and where they are heading to (destination). These are all issues which are considered in many trip generation models, for example Kockelman & Krishnamurthy (2004), Parsons Brinckerhoff Quade & Douglas, Kittelson & Associates (1995).

The next issue in the process is that of modal split, *what mode of travel to go by*? By which means of transport are commuters getting from one point of the network to another point, and why is that mode being used as opposed to another? The primary aim of researching into this issue being to help guide efforts to move people away from single occupancy travel in multiple occupancy vehicles towards multiple occupancy travel and public transport, bikes, walking etc.

The final issue commonly dealt with in transport modelling is that of trip assignment. *Are the choices being made by individuals the best they can make given the finite set of alternatives?* (although finite, the actual number may be unfeasibly large to provide any comprehensive response to this issue). This final stage draws upon information from the prior three stages in an effort to assess the existing position of the transport network and also to understand areas that need to be focussed on in order to alleviate congestion problems, reduce growth in trip rates and to encourage greener modes of transport. Thus, this stage not only acts as a
form of conclusion on the current standing of the transport model, but essentially adds a feedback loop to the system, where plans made in the assignment stage feed directly back onto the generation, distribution and modal split stages.

**Smart Growth**

Smart growth is the brand name given to sustainable development\(^6\) and can be thought of as the process which takes urban design and transport planning and marries them with the tenets of sustainable Growth. The definition of 'sustainable growth' has failed to reach an agreed point, however the Brundtland definition (World Commission on Environmental Development 1987) is a widely accepted benchmark;

"Humansy has the ability to make development sustainable - to ensure that it meets the needs of the present without compromising the ability of future generations to meet their needs" *(Page 24)*

This statement has now become synonymous with a school of thought surrounding aspects of sustainability and is not necessarily considered to be a stand-alone definition\(^7\). The general idea is that growth and development can be accommodated within a community/economy but that this must not be coupled with increased usage of or reliance upon finite resources, particularly energy sources. Encapsulated within this general notion of sustainability is the concept of externalities, which are costs associated with an action (or purchase) which are not borne by the consumer directly. Careful management of externalities implicitly results in the more efficient


\(^7\)Hanley, Shogren & White (2001), pp 133-139 provide an extended discussion of the definition of sustainability and some alternative, but consistent viewpoints.
CHAPTER 1. INTRODUCTION

use of resources and can therefore play an important role in achieving sustainability.

The modelling work contained within this thesis tackles to some extent the issue of externalities, though not directly, by providing empirically founded evidence on the root causes of the demand for transport. Transport is fraught full of externalities (which are further explored in Chapter 2) and one of the most prominent examples would be that of congestion.

Congestion is considered to be a transport related externality because it creates concentrated pockets of pollutants, causes people to spend unnecessary time in traffic queues (time which could otherwise have been used working, relaxing or fulfilling some other activity) etc. Bamford (2001) explains that it it is an all too common a problem resulting from the constrained (under-)capacity of the road system and further that the monetary value of congestion costs in 1999 were 20 billion a year, having risen from 15 billion a year in 1989. Another negative effect of congestion which receives less attention is the effect upon public transport systems, as buses are equally likely to be caught up in traffic queues as many other vehicles. Bus lanes\(^8\) for instance help to overcome such difficulties but impose constraints on available road space prohibit the use of these in all situations.

Though not directly tackling the management of congestion, the models determine some of the key drivers of the decision to use a car for certain trip types. Hence, policy discourse based on the results of these models will indirectly aid the management of congestion. These discussions are further complemented with

---

\(^8\)Bus lanes are not featured within the empirical analysis as these facilities are more general highway features that will not normally be located at site access/egress points. Rather they will normally only appear on significant traffic corridors where demand flow to capacity ratios are high. They are not recorded in the database and will not be an explicit feature of the sites considered.
a conceptual discussion of the spatial scale at which policies should be implemented.

The discussions thus far have identified that transport issues are still in a seemingly low level of maturity, with priorities that seem to change frequently over time. Moreover the goals of transport, as discussed here, have shared priorities with other social and political agendas. Chapter 2 discusses these in further detail, relating these issues to some of the more specific policy documents and other relevant literature. The next section of this Chapter introduces the TRICS database, and provides an overview of the policy based research which the organising consortium has engaged with in the past.

1.3 The Trip Rate Information Computer System

The TRICS consortium was established several decades ago by members from six counties from the south of England (East Sussex, West Sussex, Hampshire, Dorset, Surrey and Kent) and developed the database as a means to collate the wealth of information which was being generated from development applications. In order to ensure that the research work contained within this thesis is of value to the TRICS consortium, their direct users, and also the academic community, all previous research reports commissioned by the consortium under the TRICS label were first reviewed (see Table 1.1). As can be seen from the Table 1.1 they have been actively involved with producing research of direct and practical relevance to the profession since the late 1980’s. The consortium remained in place, after the databases initial conception, to ensure that the data was collected in a consistent and robust manner and also that it developed in line with the ever evolving requirements of planning officers/transport planners. To this end the data collection process has evolved to feature far more information, both qualitative and quantitative, than was required
in the industry over 20 years ago. Examples of this would include the movement away from automated traffic counts towards manual counts, a greater emphasis on multi-modal surveys and the recent expansion of the database to incorporate detailed site-specific travel plan information.

The list of research reports exemplifies the attitude of the consortium towards generating transportation research. The reports provide empirically guided evidence, mostly researched by transport practitioners, focusing largely upon politically important subjects, e.g., the journey to work, shopping journeys and school travel. There are a number of methodologies used in the reports, however the evidence they contain is largely derived from simple quantitative analysis (i.e. descriptive statistics) and often uses only small sections of the data held in TRICS. Some of the reports combined the database with data from other sources, largely through questionnaires, though none of the reports to date have attempted any formal statistical analysis. This thesis therefore contains the first application of econometric techniques to the TRICS database, the next section of this chapter defines the core research questions by which the remainder of this thesis builds upon.
<table>
<thead>
<tr>
<th>Research Report</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRICS Research Report (89/2)</td>
<td>Traffic Generation Studies</td>
</tr>
<tr>
<td>TRICS Research Report (89/3)</td>
<td>Review of Parking Standards</td>
</tr>
<tr>
<td>TRICS Research Report (89/4)</td>
<td>Traffic Generation from Industrial Estates</td>
</tr>
<tr>
<td>TRICS Research Report (89/5)</td>
<td>Traffic Generated from Marinas</td>
</tr>
<tr>
<td>TRICS Research Report (92/1)</td>
<td>Assessment of Parking Demand</td>
</tr>
<tr>
<td>TRICS Research Report (92/2)</td>
<td>Traffic Implications of Sunday Trading</td>
</tr>
<tr>
<td>TRICS Research Report (93/1)</td>
<td>Parking and Public Transport - The effect of Mode Choice</td>
</tr>
<tr>
<td>TRICS Research Report (94/1)</td>
<td>Long Term Changes in Superstore Traffic Generation</td>
</tr>
<tr>
<td>TRICS Research Report (95/1)</td>
<td>Parking and Public Transport - The effect of Mode Choice (2)</td>
</tr>
<tr>
<td>TRICS Research Report (95/2)</td>
<td>Pass By and Diverted Traffic - A Resume</td>
</tr>
<tr>
<td>TRICS Research Report (95/3)</td>
<td>Traffic and Parking at Food Retailing</td>
</tr>
<tr>
<td>TRICS Research Report (95/4)</td>
<td>Traffic and Parking at Food Retailing (Appendices)</td>
</tr>
<tr>
<td>TRICS Research Report (97/1)</td>
<td>Transport Characteristics of Non-Food Retail Parks</td>
</tr>
<tr>
<td>TRICS Research Report (99/1)</td>
<td>Transport Characteristics of Foodstores at Retail Parks</td>
</tr>
<tr>
<td>TRICS Research Report (99/2)</td>
<td>Research into Trip Rate Variation</td>
</tr>
<tr>
<td>TRICS Research Report (02/1)</td>
<td>Relationships Between Accessibility and Parking at New Developments</td>
</tr>
<tr>
<td>TRICS Research Report (02/2)</td>
<td>Seasonality Research</td>
</tr>
<tr>
<td>TRICS Research Report (03/1)</td>
<td>The Effectiveness of Travel Plans</td>
</tr>
<tr>
<td>TRICS Research Report (04/1)</td>
<td>Does Historic Site and Survey Data Remain Valid to Use?</td>
</tr>
<tr>
<td>TRICS Research Report (04/2)</td>
<td>Hospital Travel</td>
</tr>
<tr>
<td>TRICS Research Report (05/1)</td>
<td>Mixed uses</td>
</tr>
</tbody>
</table>
1.4 The Research Questions

Given the various discussions provided within this introduction, and the gaps identified in the literature in the previous section, the specific questions which this thesis attempts to provide answers to are summarised in Figure 1.2 as follows:

*Figure 1.2: Research Structure: A Conceptual Framework*

The first layer of the conceptual diagram offers the main research question (MQ), which is;

*(MQ1) What determines vehicle trips?*

Which is naturally supported by the second question main question;

---

9These adhere primarily to the original remit of the Research Proposal submitted to the Economics and Social Research Council. Whilst perhaps not as many different land-uses are covered as may have been originally intended (due in part to data limitations), extensions have been made to various goals in an effort to help provide empirical guidance on local level policy with a previously unused data source.
(MQ2) *What is the best way to model vehicle trips?*

The second of these main research question will be answered in the light of model applicability at the level of the TRICS data, rather than purely upon the strength of their theoretical foundations. It is reasonable to consider that this may perhaps be a question that should be answered prior to (MQ2), however it remains secondary to this question on the basis that models are a tool and not a reason. Therefore this order reflects that it is most important to understand the contextual issues, and why the models may be estimated, prior to understanding how they should be estimated. However even with this philosophy in mind, the answers to these first two questions will undoubtedly support each other, hence it coming into Figure 1.2 from the side as opposed to within a systematic order.

The second layer introduces the core sub-questions (SQ);

(SQ1) *What determines Trip rates at Office blocks/Employment sites?*

(SQ2) *What determines Trip rates to retail sites?*

(SQ3) *What determines Trip rates in residential zones?*

(SQ1) and (SQ2) provide clear clear scope to expand analysis to consider sub-groups within the context of the major land use. Layer three of figure 1.2 identifies this possibility and introduces two sub-sub-questions.

§Q1 *What significant differences exist within the various individual segments of the*
residential sector, of the determinants of vehicle trip making behaviour? i.e privately owned accommodation versus rented accommodation.

Following this methodical and scientific approach to this study allows the final main research question to answered;

(MQ3) What Qualitative differences are observed in the determinants of trip making behaviour across the main land-use types? e.g. Retail versus Office versus Residential.

There are a number of policy related issues that have been identified as being interrelated and the research in this thesis provide scope to answer several policy questions (p);

(p1) Prior to answering (p2), what are the key issues in contemporary transport planning?

(p2) How can the information from (SQ2) help guide development of work based travel plans?

(p3) What insights do the results from (SQ1), (SQ2) and (SQ3) offer in relation to spatial planning policy? i.e. how does land-zone choice (e.g. town center or free standing) affect vehicle trip rates?

(p4) How suitable is TRICS for this type of research to help inform local level policy?
1.5 Structure of the Thesis

In order to answer the research questions above, this thesis proceeds by reviewing the relevant parts of the literature, with the review broken into two key parts. Firstly, Chapter 2 reviews the most fundamental aspects of the policy based literature, to provide a broad underpinning of the most important concepts. Chapter 3 provides an overview of traffic/transport modelling literature is given, focussing attention on trip generation/attraction models at various levels of data composition. This will also include a review of car ownership modelling techniques.

In Chapter 4, the Data is defined, firstly via a general overview of TRICS, and then subsequently by looking at the specific data for each of the substantive sections of this thesis. This chapter also includes a discussion of the general modelling strategy employed within the three analysis chapters.

Chapters 5, 6 and 7 of the thesis contain the analytical results. Firstly, Office block (employment site) trip generations are estimated in Chapter 5. Subsequent to this will be a chapter focussing on the estimation of trip attraction models for Food Superstores. Chapter 7 will then estimate trip generation models for residential developments, including a joint model of car ownership and trip generations.

The thesis will then culminate in Chapter 8 by re-iterating the key findings of the substantive analyses and directly answer the specific research question outlined in the previous section. In so doing it will attempt to draw a scientifically grounded conclusion on the relationship between land use and travel behaviour.
2.1 Introduction

From a point of view of economics, the policy context for this thesis could be generally explained by the desire to better manage the distribution/allocation and consumption of scarce resources. In the context of the transport industry this largely entails, at least at the present, the management of oil consuming and emission generating agents as most transport is energised by either oil or diesel. Although there are signs that transport policies enacted over recent years are beginning to reduce the growth in demand for oil, it can be seen from Figure 2.1 that there has historically been a strong increasing trend. It should be noted that this is seemingly associated with a substitution away from petrol towards diesel and some petering off in demand in recent years, though no clear sign of reduction.

The context of planning for transport in general and the associated instruments through which policy are enacted have clear evolved over time in line with increasing public debate and concern owing to the health of the global environment. For example, there has been a movement away from a ‘predict and provide’ strategy towards
Figure 2.1: The demand for petroleum motor spirits and diesel fuel in the UK 1960-2005 (Data source: UK Department for Transport and Office of National Statistics online data collections)

demand management, and current Department for Transport guidance increasingly supports the inclusion of environmental costs into project appraisal frameworks. The evolution in environmental plight is closely linked with the demand for petrol¹ and diesel. This thesis does not tackle these issues directly, however the intrinsic link between travel demand and oil demand serves as a precursor to the subsequent discussion.

¹The terms Petrol, Motor spirit and Gasoline are used interchangeably here, the contention in the UK is Petrol, whereas internationally the term Gasoline is more widely applied.
ture within the governments policy literature, especially in light of their relative absence from guidance literature to date. This chapter therefore briefly reviews some of the highlights in the development of responses to travel demand problems in the context of the UK. In so doing it implicitly reveals a historic attitude which shows little concern for transport related externalities, with greater focus having been afforded to the benefits transport can bring. This chapter is not intended to be a thorough review of the policy context as such, rather it is an introduction into the core historical and contemporary aspects. Policy oriented discussion is thus broken down into two constituent parts;

- Traditional responses and
- Contemporary responses

This generally follows the evolution in the planning practice over time, where initially ‘planning’ was a relatively little thought of process, with transport externalities being massively outweighed by the gains from unfettered provision of transport (which helps sustain growth). Over time however the marginal externalities become significantly more obvious, with the marginal returns to increasing transport infrastructure rapidly diminishing. This became the precursor to the spawning of the ‘transport planner’ who’s principal purpose was to develop transport infrastructure in a far more considered way, focussing heavily on the efficient use of existing systems. The following chapter discusses some of the key features presented by the literature, surrounding aspects of the implementation of both the old and new schools of thought in travel demand management.
2.2 Traditional responses to traffic demand

Transportation studies and practices in the UK and elsewhere in the world were traditionally the realm of highways/civil/traffic engineers, who have the skills to develop and maintain the necessary infrastructures. Berry (1960), in a general discussion of the scope of traffic engineering, defines its primary goals by quoting (an undated definition by) the Institute of Traffic Engineers;

"...that phase of engineering which deals with the planning and geometric design of streets, highways, and abutting lands, and with traffic operations thereon, as their use is related to the safe, convenient and economic transportation of persons and goods..."

Berry (1960) then goes on to discuss two alternative approaches recognised within the profession to fulfilling the remit contained in the definition above namely: The 'constructive' and 'restrictive' approaches. Each of which having slightly different goals and both of which now made obsolete by an overarching transportation goal. However it serves a purpose to introduce and discuss them, as they embody the traditional approach of an engineer, not a transport planner. Hence they highlight the approaches to planning that have been tested and deemed to have failed as stand-alone methodologies.

The 'constructive approach' is essentially a 'predict and provide' methodology where engineers build a system that will meet future demand expectations. The 'restrictive approach' conversely, is effectively soft-engineering\(^2\), whereby traffic regulations and other traffic-control devices are implemented in a bid to maximise

\(^2\)Soft and hard engineering factors refer behavioural demand management policies and physical demand management policies respectively for further discussion see Cairns, Sloman, Newson, Anable, Kirkbride & Goodwin (2004b)
the efficiency of existing networks. Therefore the constructive approach is merely a supply-side policy response to increasing demand, with no obvious concern for the knock on effects of increased traffic volumes in absolute terms. The restrictive approach is more of an optimisation scheme, seeking neither to affect supply nor curtail demand per se, but rather to ensure the existing systems are used without any inefficiencies.

Discussion on Transportation Demand Management (TDM) will essentially highlight how these two schools of thought should be thought of more in terms of complementary approaches as opposed to substitutes. Prior to moving on to discuss TDM, some of the finer details of the engineering aspects will be reviewed, in particular the relationship between road capacity and vehicle flow at junctions and intersections.

2.2.1 Road Capacity, Junctions and Intersections.

A key feature in the efficient use of roadspace is the correct choice of junction to enable throughput from one road to the next in the most (cost) effective manner. The literature on junction design revolves around a number of important aspects, and its discussion is of high importance as the development of a new site has a number of knock on effects, primarily via the demand generated to that site\(^3\). The affect of increasing demand directly impacts upon the flow of cars on the road and consequently the available capacity of the road system.

\(^3\)This is of course not considering other economic effects such as the boosts to the local economy and social costs/gains etc. from the provision of the products/services made available at the new development.
CHAPTER 2. PLANNING AND POLICY CONTEXT

decision, namely (i) the flow of traffic on the major road and (ii) the flow of traffic on the minor road as can be seen in Department of Transport (1981), O’Flaherty (2002) or Rogers (2003). However, although the decision is relatively straightforward when these flows are known (assuming for simplicity that there are no land and cost constraints), it is very unlikely that these flows will be known prior to development. Estimates must be used, derived in the appropriate manner for the task in hand. Rogers (2003) presents up to date discussion on approaches to ‘Deriving design reference flow from baseline traffic figures’ for new junctions. The data is not in place to be able to conduct this junction choice analysis, however the guiding principles based on Rogers (2003) approaches are as follows:

Step 1 - Modelling expected traffic. A traffic modelling process (refer to following chapter for discussion on modelling approaches) is applied to new junctions in order to establish traffic flow forecasts for a predetermined future design year. These estimates should be either 12, 16 or 24 hour link flows. If required, these values are factored up to create Annual Average Daily Traffic (AADT) flows, which are then divided by 24 hours to get Annual Average Hourly Traffic (AAHT) Flows.

Step 2 - Peak demand profiling The second step in the process requires an understanding of the daily ‘dynamic distribution’ of traffic at the junction in question. This is necessary so that AAHT values can then be accurately factored to present the highest hourly flow. In practice a flow profile is required to observe the peaks in traffic, thus providing empirical proof of the dynamic distribution. Figure 2.2 reports an aggregated flow profile for an average Friday at a commercial warehouse. The figure provides a crude example of a flow profile and helps highlight some of the important points in their construction. The profile must be representative of an average day (thus calculated across several days worth of pooled data or multiple profiles for individual days, as is
most appropriate), and must be relevant to the development in question. E.g. it is not appropriate to understand the flow profile at a petrol station if the proposed development is a school.

Figure 2.2: Total arrivals and departures for an average Friday at a commercial warehouse (TRICS 20004b)

**Step 3 - Final calibration** Following steps 1 and 2, tidal flows considered and turning proportions are guestimated. This will help to define the likely capacity at various points of the junction at peak periods, as well as during periods of average flow.

**Step 4 - Junction Choice** For completeness the very final stage should be done both visually in relation to the graph, and then numerically with the aid of the equations outlined in Department of Transport (1981).
It is important to remember that (in the general case), roads have two contra-
flow lanes and that these must both be factored into the junction decision process,
especially when determining right turning traffic. Ashley (1994) refers to this junc-
tion manoeuvre as the 'cutting' of traffic and presents a useful set of diagrams to
identify their relationship with oncoming traffic flows. Further, Ashley (1994) makes
a key link to accident rates and explains how control junctions have an increased
likelihood of accident occurrence when compared to signalised junctions or round-
abouts. This is true regardless of the level of traffic flow, though increases in severity
and frequency with increased levels of traffic flow.

These updated guidelines for junction choice are almost unchanged from those
offered by the Department of Transport (1981), in fact the work done in 1981 appears
to be very much at the cutting edge of junction design and the more fundamental
aspects surrounding it by even today's standards. One of the key notes of interest
is the inclusion of environmental impacts as an important consideration in junction
design, an issue which is given more dedicated focus in PPG13 (1992). O'Flaherty
(2002) adds a little more depth to the discussion of junctions, relating choice to a
design process that encapsulates vehicle characteristics, drivers requirements and
accident prevention guidelines (which are heavily influenced by visibility). However
the same relationship as above prevails, first one must understand the expected de-
mand patterns, then one must seek to provide an adequate solution.

Martin Rogers (2003) presents Figure 2.3 within his book\(^4\), which shows the relation-
ship between traffic flows and junction choice, this is essentially just a tidier/neater
version of the original Department of Transport (1981) graphical representation,

\(^4\)Labelled Figure 2/2 in the book, hence the reference to Figure 2/2 in the diagram contained
in this dissertation.
further literature also presents a series of equations to facilitate the junction choice process, these can also be found in DoT (1981).

When referring to Figure 2.3 it is important to be mindful of potential land constraints, as they add a third dimension to the graph. In the absence of land constraints, the relationship is exactly as the Figure shows, however if we introduce the concept of land constraints, then the choice problem becomes decidedly more intricate.

The considerations thus far focus on the ways and means by which engineers traditionally respond to traffic demand levels. In light of the changing face of the transport planning process as a whole, and the contemporary focus on integrated planning solutions, it is prudent to expand further the contemporary views on how
to manage travel demand.

2.3 Contemporary responses to traffic demand

The previous section provided a very brief overview of some of the more historic attitudes towards planning for transport. This section subsequently outlines the more prominent issues associated with modern day regional and local transport policy within the UK, thereby contextualising some of the problems which the results contained in this thesis aim to attend to. The relevant literature is expansive to say the least, and the lack of a single document which clarifies the key attributes of UK transport policy as a succinct whole serves to be a hindrance to policy development/application in this field. The DfT's website for ‘Regional and local transport’ (http://www.dft.gov.uk/pgr/regional/, 2007) proceeds with the following statement outlining their key objectives:

"Transport plays a key role in supporting regional and local prosperity, economic growth and enhancing quality of life. To that end, the Department works in partnership with both regional and local stakeholders towards the same shared priority of:

‘Improving access to jobs and services, particularly for those most in need, in ways that are sustainable: improved public transport - reduced problems of congestion, pollution and safety.’

At a regional level, the Department encourages the close integration between regional transport, housing, economic development strategies and spending decisions. We do this to ensure that transport can best support
the delivery of wider Government objectives on the economy, the environment and social inclusiveness, and we work with partners to ensure that strategies are well evidenced and supported in their delivery.

At a local level, the Department encourages modernisation of local transport as part of our strategy for a sustainable and integrated transport system, improving accessibility and reducing congestion. We oversee local transport planning and expenditure, helping local authorities and transport operators improve local roads, bus, taxi and light rail services, and walking and cycling facilities.”

These statements underline that the notion of sustainability is a recurring theme, further that it is also a firm member of the governments stated policy mix. Moreover it is evident that the government’s attitude towards transport is that it is a key feature of national activity (measured in terms of economic growth), but that growth itself does not necessarily need to be coupled with increased transport requirements. In fact the national accounts in the past two decades or so have seen an as yet unexplained but consistent disembodiment of the link between economic activity and transport (see Figure 2.4). One potential reason for this disembodiment may be to do with the World Commission on Environmental Development (1987) which gave rise to the renowned ‘Brundtland definition’ of sustainability. From Figure 2.4 it seems plausible that this meeting and the integration of the notion of sustainability into government policy mixes (with the associated increase in spending, for which data is unfortunately not available), may have been the precursor to what appears potentially to be a structural break in the relationship between transport and economic activity.

5 “Humanity has the ability to make development sustainable - to ensure that it meets the needs of the present without compromising the ability of future generations to meet their needs”
Further to this attitude towards economic development which does not come at the cost of increased transport use (and hence increased externalities), the DfT has a proactive, though less explicit, attitude towards the reduction of externalities. This is evident through phrases such as "...enhancing quality of life.,” “Improving access to jobs and services, particularly for those most in need, in ways that are sustainable” and “...wider Government objectives on the economy, the environment and social inclusiveness,...”. It is instructive to review what these wider costs may be, so as to develop a useful understanding of the potential range of stakeholders in transport projects. These will therefore be discussed in more detail in the following section, including an overview of both the direct and indirect costs associated with
2.3.1 The True Costs of transport

Whitelegg (1993), in a more comprehensive fashion than most, see for example Bruinsma, Koets, Rietveld & Vreeker (2002) or Calthrop & Proost (1998), provides an identification and explanation of the range of costs attributable to transportation. The clarification of these help to better understand the factors which a transport planner and/or policy maker/setter should have in mind when making day to day decisions. It is a combination of optimisation of these costs (through minimising negative and maximising positive eternalities), intuitively rather than through some predefined math based model, that a planner proceeds. The specific costs identified by Whitelegg (1993) are given in Table 2.1.

Full exposition of the range of costs considered in Table 2.1 are given in Whitelegg (1993), however a few particular points should be discussed here. The costs given indicate that there are direct (i.e. affects the person using transport) and indirect (i.e. affects people who are not using the transport service) which can further be associated with upstream and downstream costs. Upstream costs would include extraction, production and subsequent delivery of petroleum and diesel etc., whilst downstream costs would include many of the 'Hidden costs' given in Table 2.1. Therefore there is an implicit multiplier in savings associated with the reduction of travel, although it is not easy to place a direct value on this due to the multiplex of stages (and stakeholders) involved in production and consumption of transport services.

Bruinsma et al. (2002) also provide a general classification of externalities broadly consistent with those given in Table 2.1, though the notion of land-use costs (both direct and indirect) are more explicitly considered.
Table 2.1: The Full Range of Costs associated with Transport Use.

<table>
<thead>
<tr>
<th>Individual costs</th>
<th>Direct costs</th>
<th>Hidden costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ownership costs</td>
<td>1. Highway and road expenditures</td>
<td>1. Destruction of Farmland, urban green space and habitats</td>
</tr>
<tr>
<td>2. Operating costs</td>
<td>2. Interest on provisional debt due to previous highway spending</td>
<td>2. Excessive energy costs from making and using the car</td>
</tr>
<tr>
<td></td>
<td>3. Government spending on the environment (i.e. pollution control/clean up)</td>
<td>3. Damage to air and water from mining</td>
</tr>
<tr>
<td></td>
<td>4. Road safety</td>
<td>4. Air and soil pollution and contamination from smelting</td>
</tr>
<tr>
<td></td>
<td>5. Health care</td>
<td>5. Air and water pollution from drilling and processing petroleum</td>
</tr>
<tr>
<td></td>
<td>6. Policing</td>
<td>6. Water pollution from the production of petroleum-based chemicals</td>
</tr>
<tr>
<td></td>
<td>7. Court costs</td>
<td>7. Damage from the transport of petroleum on land and water</td>
</tr>
<tr>
<td></td>
<td>8. Subsidies to companies</td>
<td>8. Acidification of land and water from emissions and industry smokestacks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9. Damage to plant and crop growth from elevated levels of ozone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10. The growing environment and health costs of global warming</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. Damage to water and vegetation from use of salt and oil as dust suppressant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12. Damage to air, land and water from disposal of cars and component parts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13. Damage to human health from discharge of waste into lakes and rivers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14. Respiratory damage from elevated levels of $SO_2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15. Respiratory damage from elevated levels of ozone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16. Impaired co-ordination and heart damage from CO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17. Neurological damage from elevated levels of lead</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18. Damage to skin and eyes from ozone depletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19. Loss of time due to overcrowded highways</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20. Stress and decline of quality of life</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21. Unique costs of transportation disadvantaged</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22. Financial costs due to lost productivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23. Emotional damage to victims and families</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opportunity cost of car dependence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1. Lack of R&amp;D for rural and public transit and alternative fuels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Growing inflexibility of the economy</td>
</tr>
</tbody>
</table>
Similar to the concept of sustainability, the notion of externalities in the context of transport is not well defined. Although the general crux of the term is well understood, its measurement and valuation is so complex as to make it too ambiguous to apply a full evaluation. Given however that these (negative) externalities are an implicit function of travel, then they can be indirectly managed by managing the demand for travel.\textsuperscript{7}

As alluded to in the introduction, congestion is considered to be a transport related externality because for instance it creates concentrated pockets of pollutants, causes people to spend unnecessary time in traffic queues (time which could otherwise have been used working, relaxing or fulfilling some other activity) etc. Bamford (2001) explains that it is an all too common problem resulting from the constrained (under-)capacity of the road system and further that the monetary value of congestion costs in 1999 were 20 billion pounds a year, having risen from 15 billion in 1989. Another negative effect of congestion which receives less attention is the effect upon public transport systems, as buses are equally likely to be caught up in traffic queues as many other vehicles. Bus lanes and similar initiatives help to overcome such difficulties but constraints on available road space prohibit the use of these in all situations/locations.

As a result of these increasing monetary and social costs upon the various members of society, it is necessary to consider ways in which congestion can be alleviated. Bamford (2001) identifies some possible solutions as;

\textsuperscript{7}More direct methods for dealing with externalities include direct taxation (Calthrop & Proost 1998) or road pricing (Hau 1992). Given the nature of the empirical work contained in this thesis, these approaches are recognised but discussed no further.
CHAPTER 2. PLANNING AND POLICY CONTEXT

- Making Better Use of the Road Network
- Building More Roads
- Improving Public Transport
- Increasing the Cost of Urban Travel (for Motorists)

What has therefore been identified within this section is that there is not only a rational set of reasoning for reducing congestion, but also that there already exist some potential guidelines/frameworks for achieving reduction. The remainder of this chapter provides further discussion over the impacts of traffic and how to manage the demand for travel.

2.3.2 Traffic Impact Analysis

Recognition of the need to understand the impacts of new developments is increasingly well documented in the literature, and references to the notion stretch back as far as 1970 in the US urban transportation policies (see Meyer (1999) for more detailed discussion). Suggestions are made that the implementation of such approaches occurred more due to a lack of funding for infrastructure than a desire to create more socially focussed sustainable transport solutions. Regardless of this Meyer (1999) clearly presents some of the reasons why the full dynamics of the planning dilemma (i.e. being mindful of the need to plan for the future period’s needs and requirements) must be considered. Some of the basic points Meyer presents include environmental consequences such as pollution effects and clean air, and the supply/demand of energy resources. Also discussed is the need for planners to bring some focus on creating efficient solutions.
CHAPTER 2. PLANNING AND POLICY CONTEXT

One response to managing the impact of demand from a new development is Transportation/Travel Demand Management (TDM), (see for example Meyer (1997)) which is further explained below. This is an approach that needs to be tied in with traditional highway engineering for an effective solution to be created. The more fundamental highway engineering solutions stipulate what is and what is not feasible within a road network and why, whereas the TDM approach sets down a more idealistic forward thinking response. The problem is difficult as the TDM must remember that the engineering solutions are relatively non-linear in their responses, however the engineers must take into account the need to accommodate for TDM solutions. When land constraints do not exist, it is much easier for a balance to be found between the two schools of thought, however when land (and perhaps more importantly capital) is not abundant, decisions can become more complex.

As mentioned previously, Bamford (2001) identifies the role of congestion in modern day society as a source of increased monetary and social costs to all parties, not just those who use the transport system directly. Similarly, Bergendorff (2003) for instance, identifies the prominent impact which downstream effects have in the total (social) cost structure of infrastructure changes. These costs serve as an impetus for the uptake of new and more controversial policies, as local, regional and national authorities realise that change is needed to the way we live, and it is needed soon.

As previously discussed in the true costs of transport, there are more concerns surrounding transport than recent media attention gives rise to. Whitelegg (1993) has provided a comprehensive overview, though Litman (2004) short-lists the following main impacts as some of those attributable to transportation;

- Urban design,
- Development and Public Service costs,
CHAPTER 2. PLANNING AND POLICY CONTEXT

- Sustainability impacts,
- Equity Impacts,
- Aesthetic impacts,
- Cultural preservation,
- Environmental impacts

The concept of urban design has already been introduced in the introduction, along with the concept of aesthetic impacts and the need to design places that people want to live in. Development and public service costs have not really been given much attention in the discussion within the thesis hitherto, since this is not a direct focus of the work herein. Equity impacts are an increasing concern in the research with projects such as AUNT SUE (Sustainable Urban Environment) project look into the transport needs of largely excluded members of the community. One problem which arises from transport is the inertia towards already 'transport rich' areas, due to the cost effective nature of the markets in these areas. This can to some extent increase the divide of wealth between rich and poor members of the community determined by access to goods and services. As such there is a potential inequitable distribution of wealth effect that may occur with the advent of transport system innovations.

Cultural preservation and environmental impacts overlap considerably, though both do have unique features. Cultural preservation relates to the desire to retain the qualities of the aesthetic nature of the environment, i.e. the preservation of green field space. In the UK this has given rise to policy initiative intended to discourage and/or completely avoid development on Greenfield sites in preference of re-development of Brown field sites. This is not always achievable but is a policy
implemented with the intention of maintaining the physical fabric which defines the countryside as it is. This cultural preservation can also apply to the intangible features of an area that are captured within the physical fabric.

Environmental concerns relate alternatively to the preservation and sustenance of the ecosystems within the area. Every development has an impact on an ecosystem, with both positive and negative effects. A recent practical example would be the development of 'The Worlds' in Dubai in which the construction of ultra expensive housing (on individual man-made islands perhaps) resulted in destruction of the sealife and ecosystem in the water around those islands. The developers of this site then were forced to invest substantial sums of money to clean the water and re-introduce life into the waters in order to uphold the value of the properties. This brings to the fore an interesting point, as it was recognised by the developers that it was financially more attractive for them to maintain the environment during their development process.

However, although all of the full range of upstream and downstream costs and concerns surrounding the use of transport are undeniably important, this thesis constrains discussion primarily to the direct traffic flow impact at a development only. It is clear that transport must be clearly targeted, as to ignore it would be to ignore the source of approximately a third of the UK's total emissions. Recent government literature such as Planning and Policy Guidance note number 13 (PPG13 1992) clearly presents the official viewpoint in relation to this debate. Sustainability is one of the key objectives presented in this document, outlining a number of responses

---

8 And also greenfield sites within urban conurbations, as these provide character to highly developed zones.

9 This is justified on the basis that to concentrate attention in this area is to implicitly provide meaningful discussion relating to the core driver of all the upstream and downstream effects.
that will be shown to coincide with the key goals of TDM.
2.3.3 Transportation Demand Management (TDM).

The History of TDM is a little ambiguous, in that it is an approach that was historically adopted on numerous occasions, but never under a unified name. Essentially the badge was given to the concept in the US in the 1970's, when state and local efforts to manage congestion problems were taken into federal management, and further linked to the common goals of meeting future requirements while being mindful of environmental aspects. Meyer (1997) presents an in depth discussion on TDM and its role within congestion and mobility alleviation and in his own words;

"In its broadest sense, transportation demand management (TDM) is any action or set of actions aimed at influencing people’s behaviour in such a way that alternative mobility options are presented and/or congestion is reduced"

It was previously mentioned that TDM is a complementary process/tool that should ideally be used in conjunction with orthodox highway engineering techniques, however it is worth noting that it is also to some extent necessarily sequential to any highway engineering response. The reason being that a firm understanding of the engineering response is required before it can be known how much to alter the mobility patterns of user groups. In practice this will be able to be established prior to any physical construction has occurred, but the TDM specialist must understand how the engineers are thinking, and any restrictions imposed upon them.

Direct links are made between land use and TDM in Meyer (1999) with some important points being raised. The spatial distribution of land and in particular its specific type of use are noted is important factors in a relationship between volumes, patterns and modal split of trips. As a result, an understanding of and control over the trip generation characteristics of a site provides scope for an efficient allocation
(or at least steps towards) of transport within a network, that can help shape the characteristics and behaviour of an entire community.

In implementation TDM tackles three key areas Meyer (1997);

- To facilitate the switch to alternative modes of transport.
- Encourage travel outside of peak hours
- Encourage Teleworking and telecommunications in an effort to avoid unnecessary journeys.

There are many ways in which the above options can be addressed, its essentially an 'insert your own idea here' framework, however the idea/policy instrument must be robust enough to withstand scrutiny, and realistic enough so that it can actually be implemented. Aside from that there are no real restrictions on how to go about achieving these three goals.

In a discussion paper, Cansult Limited (2000) make the important assertion that TDM represents a demand-side response to changes in the level of demand, that is to say, a TDM solution must necessarily avoid increasing supply-side factors. By holding supply constant (or as near as is feasibly possible given individual network/problem characteristics), the TDM is enforcing a change in the demand profile, as a result individual users of the network may be forced away from unconstrained utility maximisation towards a form of constrained behaviour, with an intention to maximise the population’s utility in the long-run.

This concept works on the notion that demand is elastic, and consequently susceptible to changes in the supply/demand relationship, such that when demand is
increased, but supply remains constant, it should in effect find its own level again. Or a reduction in the provision of supply will cause demand levels to change. However, this is not always the case as was found at the University of Wisconsin-Madison by Fair, Frisk, Oakleaf & Renner (2000). Where efforts to reduce the utility derived from the demand for parking by influencing the cost of supply had no effect, leaving the authors somewhat perplexed as to how to proceed. Regardless of the seemingly inelastic nature of the demand with respect to price in this instance, the recommended action was to impose a form of two-part tariff allowing for two different pay rate structures which by their own admission is likely to fail! This example highlights just how difficult a position transport planners can find themselves in, though also highlights the naivety of planners in certain cases, where the approaches taken rely on putting all the eggs in one basket. For instance Fair et al. (2000) make no real efforts to present alternative solutions such as park and ride\textsuperscript{10} or even tele-working (though the potential for this is not great in this instance), just rely on pricing strategies to enforce demand.

In an interesting conclusion Meyer (1997) remarks that it is unlikely that TDM by itself will have any major impact upon car usage, but that it presents ways in which traffic demand responses can be linked to wider political issues in an attempt to sway the public towards a more 'sustainable' attitude. This by itself does not detract from the importance of this discussion, moreover it highlights the consensus in this field of research of an uphill struggle against public opinion and attitudes.

\textsuperscript{10}Park and ride facilities are not normally located at the site types considered within this thesis, further these facilities generate trips to other locations which are not by car, hence will not feature in the data used. Therefore no further discussion is given on such features.
2.3.4 TDM and Government Guidance Notes

The traditional role of the transport planner was considered very much to be a linear process in which the goals and objectives were chosen from the outset, then a number of processes were sequentially dealt with until a final solution was met. Namely, understanding traffic flows, formulation of traffic re-assignment plans, evaluation of the effects of re-assignment and then the execution of the plan (see Figure 2.5, taken from Button (1982)). However in recent years the role of the contemporary transport planner has branched out to incorporate the roles of other professions too, such as engineers, urban designers, psychologists, architects, economists etc.

The government produces a wealth of literature/guidance on transport planning the related issues of development. Of greatest notoriety within the transport planning profession are Planning Policy Guidance 13: Transport (PPG13 1992) and also Planning Policy Guidance 6, which is unfortunately no longer accessible. At the time of writing there were 18 separate Planning Policy Guidance notes available from the Office of the Deputy Prime Minister's website (ODPM).

PPG12 (1992), looks specifically into the area of Development Plans, considering both their initial conception and also what considerations should be borne in mind when updating or renewing them. The note initiates with a comment on the prudency of instigating a rigorous and well organised management structure to facilitate improved performance. Further to this, it is recognised that a 'carrot and stick approach', whereby government assign local authorities targets, which should help to enhance the performance of the planning sector, as local planners have clear targets to work towards. Also made clear is the necessity to consult with all potential stakeholders, on more than one occasion, such that development can fall in line with the interests of the community who will intentionally benefit from it.
Figure 2.5: Economists' view of Transport Planning in 1982

- Decide Goals and Objectives
  - Inventory of existing patterns of travel
    - Develop mathematical models of local supply and demand relationships in transport
      - Formulate alternative plans
        - Forecast physical effects of plan on local travel patterns
          - Economic evaluation of the alternative plans
            - Execute plan
Focussing upon Development Plan Reviews, PPG12 (1992) clarifies that the following factors should not only be considered, but they should also be continually monitored:

- The principal physical and economic characteristics of the area
- the size, composition and distribution of population (Whether resident or otherwise) of the area, and
- the communications, transport system and traffic of the area

where statutory power exists to ensure that this can be done. This therefore recognises that travel behaviour is derived from a series of complex inter-relationships, to consider the effects of development without monitoring the full range of effects would result in an incomplete understanding and hence less effective development plan. PPG12 further re-emphasises the importance of the concept of sustainable development.

The other PPG notes stress the various different issues important in dealing with the development of a specific land use type, for instance PPG 21 considers ‘Tourism’ and PPG 17 looks at ‘Planning for Open Space, Sport and Recreation’. Alternatively a number of the guidance notes look at site specific characteristics which may influence the planning process, i.e. PPG 25 ‘Development and Flood Risk’ and PPG 10 ‘Planning and Waste Management’. Essentially this clarifies that the planning guidance notes offered by the ODPM try to encapsulate the complexities and readily observable differentiables which have bearing on the effectiveness of the planning process. To this extent these guidance notes are necessary reading for any planner.
An increasingly widely documented and applied approach to travel demand management is the application of smarter choice style interventions within the planning process in order to support development in a sustainable manner. The following section (and supporting Appendix) distinguishes between what a soft factor and hard factor is, as well as identifying the way they work and their potential effectiveness.

### 2.3.5 Soft factors versus Hard Factors

There are two ways, as already alluded to in some of the previous discussion, that transport policies can essentially be implemented. These are termed as soft factors and hard factors where a soft factor works by adjusting the behavioural profiles of individuals in the market for car travel, hence an effective soft factor intervention will work by shifting the demand curve. A hard factor on the other hand imposes restrictions on car use therefore there is no adjustment to the demand schedule in the market for car use, but supply will be reduced.

Due to Cairns et al. (2004b), evidence suggests that sensible application of soft factors alone could reduce national transport levels by a minimum of 2% up to a maximum of 11%. It is noted that the lower end of this spectrum is the more realistic end of the target, however it should not be thought of as a small change. Due to the sheer scale of the demand for travel, a 2% change would be a vast reduction in the number of trips/distance travelled in the UK. Care should be taken to emphasise that these are projections, and as such should not be relied upon wholly, however a number of cases studies are presented in their supplementary work Cairns, Sloman, Newson, Anable, Kirkbride & Goodwin (2004a), 24 in fact, supporting the point that there is realistic scope for these ‘soft’ policies.

It can be shown under certain assumptions that the decision to apply a soft
factor approach to travel demand reduction, provides increased scope for rebound behaviour (See Appendix I at the end of this chapter). To some extent this would imply that hard factors may be more preferable, as they reduce/eliminate the potential for backfire associated with the implementation of site specific travel reduction policies. However this does not take into account social and psychological aspects, which are harder to determine, measure and control. In particular a soft factor approach provides individuals with a sense of empowerment, in that they have the choice to change behaviour and know that they will benefit the wider environment if they do. i.e. the positive externalities of soft factor responses may potentially provide individuals with a significant positive utility. Hard factor approaches simply constraint the set of available options open to an individual, and thus would likely leave individuals feeling to some extent that they have had some of their liberties removed. The extent to which this may be true is unknown, however it is obvious that the application of transport policies is not an entirely easy process.

2.4 Summary of Planning and Policy Context

The resources currently used to facilitate the vast majority of personal transit come from non-renewable sources, which for instance according to Monbiot (2003) are forecast to last until only around 2037 (prediction based on reportedly ambiguous U.S. Department of Energy forecasts).\textsuperscript{11} Regardless of the exact terminal date of oil supplies, there is a clear need to focus on the future. This chapter identified the contemporary link between achieving transport objectives which do not come at the expense of either;

1. Economic growth and stability

\textsuperscript{11}There are in fact numerous predictions on the reserves of oil, with some disparity in predictions, however there is broad agreement that time is running out.
2. Environmental stability for now and for future generations

Ensuring economic growth and stability is a pre-requisite for any economy for many obvious reasons, and in the past transport has been considered an integral feature in procuring and maintaining efficient economic activity. However as has been shown, this seeming reliance upon transport as a determinant of economic productivity is less substantiated in the past decade than has been the case in recent history (refer back to Figure ??). As yet however it is not entirely certain what the drivers of this structural change in behaviour are, and this is openly identified by the national research councils as an issue which needs immediate and thorough exploration. Though this is however well beyond the scope of this thesis.

The link between economic activity and environmental stability is made obvious through the way in which the majority of energy demand is derived from economic activity. This is complemented with the fact that many traditional and incumbent energy sources in (particularly though not exclusively) developed countries are high emitters of gasses and pollutants which are harmful to the environment. However environmental concerns extend beyond the mere use of energy (and the management of growth in energy use) and the finite nature of many resources, with their increasingly rapid dissipation, is rekindling fears over the extent of humanity's impact on the world.

All of this points towards the development of policy mixes and instruments to ensure that energy and resource use is more efficient and less destructive than has been the case in recent years. This is spurred also by forecasts of rapid growth in densely populated developing economies such as China and India, where if expecta-

\footnote{The link was documented in the economic literature as early as Smith (1776), who remarked how efficient joint management of transport can provide mutual benefits.}
tions of growth are fulfilled by the same means that western economy’s used, could be immeasurably costly to the global environment.

This thesis does not attempt to directly extract policy measures designed to help respond to the above mentioned concerns at a national level. But through the models which are subsequently developed and estimated offers a disaggregated insight into the key determinants of car based traffic for certain trip types. Through this it shall be how behavioural influences have historically determined the decision to use a car to complete these trip types. From this information, i.e. knowing what’s important in determining car use, it will then be possible to consider how policy measures can be manipulated, or new instruments developed, to manage car use more effectively. The externalities associated with car use are well documented and it is clear that more efficient management of car use at any level of aggregation will serve to benefit the economies wider goals of achieving sustainability and reducing other negative externalities.

Given the desire to better understand and develop policy instruments to mange these deleterious effects, the following chapter discusses the major techniques applied to model the flows of traffic and general travel behaviour.
2.5 Appendix 1 - Soft factors versus hard factors

2.5.1 Introduction

Demand for transport is derived from the desire and need to consume goods and services, whereby transport is merely a necessary tool to access these goods and services. If one accepts a stock definition that a lifestyle is characterised by the activities conducted by an individual (and the time allocated to those activities), it is then clear that lifestyles hinge on the availability of transport and also that the demand for transport depends on lifestyle choices. Such a connection is in some ways intuitively simple and requires little explanation whilst in other ways it is fraught with undesirable complexities and caveats. Current projects such as the UK Economic and Social Research Council’s 2.9 million Research on Lifestyles Values and the Environment (RESOLVE) stands testament to the need to better define the link between lifestyle choice and the inception of sustainable lifestyles.

Contemporary lifestyles are significantly influenced by technological advances, through, for instance, media such as the internet and the increased communication capacity afforded by the invention and inception of mobile telecommunications (see for instance Office of Communications (OfCom) (2004)). These advances have allowed international economic activity to pick up pace, where for instance websites such as ‘e-bay’ allow for conventional barriers to market entry to be significantly redefined, thus stimulating competition. This increased demand for activity translates as an increase in the demand for energy, however it is well recognised that ‘conventional’ energy sources are finite in supply, particularly those primarily used for transport. This serves to be a further catalyst for the recent rapid expansion of the sustainable communities policy drive, where sustainability is commonly expressed using the Brundtland definition on Environment & Development (1987):
"Humanity has the ability to make development sustainable - to ensure that it meets the needs of the present without compromising the ability of future generations to meet their needs".

However it is arbitrarily defined that sustainability for the road transport sector (with the exception of freight transport which is an entirely unique sector), focuses primarily on the reduction of car use, particularly single/low occupancy journeys. This appendix\textsuperscript{13} therefore contributes to the discussions surrounding the level of 'spatial aggregation' at which policy interventions for reducing the demand for car travel should be implemented.

2.5.2 Travel planning in the early 21st century

Stakeholder engagement has internationally become a key issue in the transport industry in recent years (see for instance European Conference of Ministers of Transport (2002)), more so than had been previously considered appropriate. Travel Planning and Travel Demand Management (TDM) have thus become the core tools of the present day transport professional due to their direct involvement with the network users, whereas the pure 'Predict and Provide' approach of yesteryear (see Berry, 1960), i.e. simply increasing network capacity, has now been largely foregone. Such approaches (i.e. Travel planning and TDM) tend to influence behavioural change at the origin end of the journey and increase mode choice availability, rather than to impose constraints on car use.

\textsuperscript{13}The information contained within this appendix may also be found in:

The range of instruments available to policy setters for intervening in, and adjusting behaviour towards the setter's objective goals is potentially infinite. As a result it is more convenient to define a general target for what policies must achieve, policies can then be defined in any way so long as they meet these criteria. The SMART criteria are adopted here;

**Specific** There must be no ambiguity in the output

**Measurable** The policy target(s) can be set against directly observable output(s)

**Achievable** The policy must be feasible (Rocket science should be avoided)

**Realistic** Targets should be within reasonable bounds and not too optimistic

**Time bound** The output of the policy should be observable over a pre-determined time frame

Such SMART policies can subsequently be implemented in a number of ways, the present appendix considers the following alternatives; Site Specific Travel Plans and Area Travel Plans as will each be briefly discussed;

**Site Specific Travel Plans (STP's)** The TP can be broadly defined as the process whereby a specific site (i.e. an office block) evaluates its existing travel requirements/demand (or predicted requirements in the case of new developments), then through the implementation of an appropriate mix of policy instruments, seeks to reduce that travel demand to some agreed target level. The DfT offers expanding amounts of advice and guidance on TP implementation and evaluation and the Association for Commuter Transport is also very
active in disseminating advice to practitioners around the UK.

TP’s are a highly proactive approach to managing travel requirements encouraging a process of planning, implementing, monitoring and evaluating demand. Monitoring is a key element of the TP whereby successfully implemented TP’s are encouraged to be continually monitored to ensure they remain successful, and careful monitoring during the implementation helps to reveal what is causing the TP to be more or less effective at certain points in the process. The site specific nature of TP’s means that there will be parts/users of the wider network that are potentially not directly controlled for in the TP policy interventions.

**Area Travel Plans (ATP’s)** Area Travel Plans are a natural extension of the site specific Travel Plan (TP), where the planning process is not instigated by the local authority, but by the stakeholders themselves, so as to take proactive TDM actions. However the ATP is trying to manage travel across a much wider network and not constrained to implementation at discrete points of time.

ATP’s offer high levels of community engagement, and are a dynamic process seeking community feedback with the introduction of each new development in an area to ensure they are managed appropriately. This affords ATP’s the ability to instantaneously adjust the development of a region to account for its constantly changing dynamics. Thus community engagement is far more frequent, thus allowing network users to adjust behaviour instantaneously. Furthermore the manner in which the ATP decentralises organisational power away from the government, and further provides stakeholders with the ability
to influence the development of their local network is seen as an advantage over alternative statutory transport planning instruments.

2.5.3 Supply, Demand and Economic Rebound

The concept of rebound behaviour (Otherwise known as ‘Offsetting behaviour’, ‘backfire’ or ‘Takeback effect’, is generally characterized by an increase in the demand for a commodity, in this instance car travel, induced by an indirect reduction in the price of that service. The notion, stemming largely from the pioneering research of Khazzoom (1980) and Brookes (1978), is often characterized by price reductions induced by technological advancements, in particular implied price reductions which come about through increased efficiency. The ‘technology’ in this case is the policy intervention, which is implemented to reduce car use and therefore reduce the level of associated externalities attributed to an area of the network.

An interesting question is ‘why might rebound behaviour occur in the transport industry?’ The response is intuitively simple; a reduction in the cost of travel, holding the price of all other goods constant, reduces the relative price of all goods and services and hence increases the amount of disposable income available. This indirect increase in disposable income allows individuals the opportunity to engage in a greater number of activities (e.g. shopping or leisure), indirectly generating increased demand for travel. Therefore a reduction in the price of travel causes an indirect increase in the demand for transport. Rebound however likely occurs not only because individuals have the ability to increase consumption, but must necessarily be supplemented by a desire to increase activity. The law of diminishing marginal utility would imply for instance that rebound is more prominent for lower income lifestyles, suggesting that rebound is a direct (inverse) function of implied quality of life.
CHAPTER 2. PLANNING AND POLICY CONTEXT

STAGE 1
Observe the scope for policy. Is change needed?

STAGE 2
Define target group and also recognise impacts on non-target group.

STAGE 3
Define a selection of alternative policy intervention schemes comprising soft and hard-factors.

STAGE 4
Evaluate impact of policy interventions on target and non-target groups.

Figure 2.6: Policy implementation - A precursor.
A general framework by which policies are implemented could be summarised as in Figure 2.6. This framework begins by first assessing the scope for policy change, i.e. what is the current consumption attitude for car use? Following this, the relevant stakeholders should be identified and segmented into the target and non-target groups. Stage 3 sees the definition of specific policies which might be considered for reducing car use, whilst stage 4 subsequently evaluates the effects of the policies defined in stage 3 (upon the target and non-target groups). Once stage 4 is complete, planners will have the necessary knowledge to compare the overall effects of alternative policy schemes and choose the one which best suits their needs.

Following for example Cairns et al (2004a) and the 'smarter choices' literature, policies can be considered to be either soft or hard factor interventions, whereby soft factors are interventions such as tele-working, car-sharing and travel awareness campaigns whilst hard factors are those such as reducing parking provision at a site or removing lanes on a multiple lane carriageway. Supply and demand analysis will now be used to consider the effects of soft and hard factor interventions on the target and on-target groups in turn (analogous to policy implementation through a site specific TP). All figures, unless otherwise stated, represent the demand for car travel, further, the demand curves are Marshallian or compensated demand curves (see for example Varian (1992) or Mankiw (2004)) and therefore implicitly incorporate mode switching (substitution) behaviour. The demand curve in Figure 2.7a is represented by the downward sloping curve denoted 'Demand' (D in the subsequent Figures) indicating that at high prices consumers only desire a small quantity of goods, whilst at low prices much higher quantities are demanded. The Figure also features an upward sloping line denoted 'Supply' (S in the subsequent figures) which indicates that supply is high when prices are high (as more revenue can be skimmed.
Figure 2.7: Effect of soft factor policy intervention on target group.

from the market) and vice versa.

Figure 2.7a shows the initial supply and demand schedules in the market for car travel with the equilibrium quantity consumed being \( q_1 \) with price \( p_1 \). Soft factor interventions therefore work by adjusting the behavioural profiles of individuals in the market for car travel, hence an effective soft factor intervention will work by shifting the demand curve (with no change in the supply schedule) to the left with the following implications (see Figure 2.7b);

- Demand shifts inward from \( D_1 \) to \( D_2 \), thus indicating a lower quantity demanded for any given price.

- Given that the target group is aware of the introduction of the policy intervention, quantity of car travel demanded instantly moves to the new equilibrium and decreases demand from \( q_1 \) to \( q_2 \)

- The price of car travel also decreases, moving from \( p_1 \) to \( p_2 \)
CHAPTER 2. PLANNING AND POLICY CONTEXT

Figure 2.8: Effect of hard factor policy intervention on target group.

Figure 2.8 shows the effects of the implementation of a hard factor policy intervention upon the target group. Hard factor’s take an alternative approach to TDM (relative to Soft factors), and instead of working to adjust attitudes towards car use simply impose restrictions on car use. Therefore there is no adjustment to the demand schedule in the market for car use, but supply will be reduced with the following effects (see Figure 2.8b);

- Supply shifts inward from $S_1$ to $S_2$, thus indicating a lower quantity available in the market at any given price
- Given that the target group is aware of the introduction of the policy intervention, quantity of car travel demanded instantly moves to the new equilibrium and decreases from $q_1$ to $q_2$
- The price of car travel also increases, moving from $p_1$ to $p_2$

One important point revealed by Figures 2.7 and 2.8 is that hard factors increase
the generalised cost (price) of car use for the targeted individuals, whilst soft factors reduce it. This would imply that soft factor interventions provide greater scope for rebound behaviour within the group that the policy intervention is directly targeted upon. Now, the implications of Soft and Hard factor policy interventions upon individuals for whom the policy is not directly intended to alter the travel behaviour of (i.e. the non-target group), will be considered. One common feature presides in the following discussion, which is that policy interventions manifest as a reduction in the price of car travel for the non-target group as follows:

- A reduction in the quantity of car travel on a given sub-section of the road network reduces the generalised cost of travel through that part of the road network (the 'area of implementation') via reduced congestion

- Sections of the road network adjacent to this 'area of implementation' will be able to substitute their existing route for a new route which passes through the 'area of implementation', paying a lower overall price for the trip

- There will subsequently be a knock on effect of traffic redistribution across the whole network due to further route substitution. This comes about as the areas bordering the section of the network immediately adjacent to the 'area of implementation', experiences a similar cost reduction as is inferred in the previous bullet (See Figure 2.9)

As this continues across the entire network, the overall result will be that residents and network users will be able to sustain the same level of consumption (for all goods and services) as enjoyed before the policy intervention, but with lower travel costs.

As the price of transport falls (assuming for simplicity that transport costs are fixed across all goods), the budget line for all goods and services shifts proportion-
CHAPTER 2. PLANNING AND POLICY CONTEXT

Figure 2.9: **Sequential rebound within a spatial region.**

ately outwards. This comes about as the effective costs of all goods have fallen by a fixed amount. Given that transport is a complement for all other goods and services, the demand for transport therefore also increases.

Figure 2.10 shows the implications of these two effects (price reduction and demand increase). In the instance of a Soft factor measure (see Figure 2.10a);

- Price reduces from $p_1$ to $p_2$ moving the individual out of equilibrium behaviour. This comes about as the reduced number of cars on the roads eases congestion and hence reduced the generalized cost (price) for all transport network users

- The consumer re-evaluates their budget constraint and translates the reduced price of transport as a general reduction in the cost of living. This allows them to increase overall consumption and achieve a higher utility curve. Demand increases from $D_1$ to $D_2$
Figure 2.10: **Effect of policy interventions on non-target groups.**

- Quantity demanded increases from $q_1$ to $q_2$ whilst price adjusts to the new equilibrium at $p_3$

Therefore the implementation of the travel demand management policy might result in an increase in the demand for travel by the non-target group when using soft factors. The general result remains the same when using Hard factor policies but for the following exception (see Figure 2.10b);

- Depending on the type of Hard factor used, supply in the overall market for car travel may reduce. This reduction in supply will reduce quantity demanded from $q_2$ to $q_3$ and raise price from $p_2$ to $p_3$

As a result, the true effect of a hard factor depends on the degree to which supply is constrained. If the reduction in supply exactly equals the increase in demand, then the quantity demanded will remain unchanged, however the price associated with that level of demand will be higher (as the same amount of road users will be
competing for a smaller amount of road space).

The distinction between two uniquely observable groups, namely (i) the permanent users of the spatial regions travel network being directly targeted by the policy instrument (the target group), and (ii) the users who are not (the non-target group). For the target group it was observed that policies will have the desired effect of reduced quantities of car use. However for the non-target group it can be seen that undesirable increases in the quantity of car use are both rational and feasible. These two equal and opposite effects mean that careful spatial consideration is required in the implementation of policy.

When using a site specific TP, the net effect of a policy intervention is ambiguous and depends on the proportion of society which is engaged by the policy and the relative ability of those not directly targeted by the policy to increase car travel. Therefore it is evident that the greater the proportion of a ‘spatial region’ which is incorporated into a travel plan process, the less possibility there is for rebound behaviour to occur and hence the more effective the policy intervention will be. The ATP by definition includes the whole spatial region and therefore minimises the potential for rebound as far as is practicably possible.

This appendix has exploited the theoretical potential for a situation such that if policy interventions are implemented at an inappropriate level of spatial aggregation, rebound behaviour may occur. These policies are implemented so as to reduce the impact of a site’s negative externalities upon the general public however it can be seen that externalities may reach higher levels than had the policy not been implemented at all. These effects, if not at least recognized by policy planners and implementers, may significantly hamper progress in sustainable development.
Empirical evidence is necessary to see whether or not this theoretical phenomenon is prevalent in reality and the magnitude to which it may be masking the effectiveness of policy interventions. Further it is recommended that the time has come for development control officers to stop being SMART and to start being SMARTeR; (Specific, Measurable, Achievable, Realistic, Time bound and excludes Rebound). The ATP is considered as one such instrument with the ability to implement this 'SMARTeR' approach to sustainable development.
Chapter 3

Modelling Traffic Generation

3.1 Introduction

This chapter discusses the core methodologies and literature surrounding models of trip generations, including discussion on models of trip distribution (and how they feedback into the overall ‘four-step’ transport model). As a precursor to the more complex analysis of vehicle trip generations by residential sites (see Chapter 7), this chapter also includes a short review of joint models of car ownership and use, thus keeping with the theme of trip generations. The chapter focusses on formally presenting the core methodologies, drawing upon a selection of key references whilst providing a commentary on references relating to issues presented by these key texts in order to explain and underpin the approach taken in Chapters 5, 6 and 7.

The literature surrounding aspects of transport analysis is far greater than any single thesis has the scope to cover, as a result the review offered here is a focussed overview of the core literature relevant to the task in hand. The main review is separated into two constituent chapters, the present one which covers the core modelling approaches and the following one which discusses some of the key policy issues.
Further literature is discussed in the subsequent chapters, though these references are borne of specific empirical circumstances and hence dealt with outside of the general review.

It is prudent to identify in advance that this thesis is unable to directly consider the estimation of price elasticities as (i) the relevant price information is neither accessible at the appropriate level of aggregation retrospectively and (ii) such data would not necessarily be directly reflective of the prices that the trip maker pays. This is introduced now, rather than in the data discussion as it preempts the lack of a formal discussion of price elasticities of demand derived from previous studies. Moreover, there are numerous studies which have generally reviewed the general literature on price elasticities, see for example Glaister & Graham (2000), Glaister & Graham (2002) or Dahl & Roman (2004) for recent discussion.

The present chapter proceeds by reviewing the role transport plays in relation to general consumer behaviour within a microeconomic framework. Although this bears no direct influence upon the empirical work in this thesis, as it is not directly tested, the theoretical precursor enables a more thorough understanding of the policy implications of the empirical models.

3.2 Trip Modelling

The following 3 subsections outline the microeconomic foundations of travel activity behaviour, then discuss the core approaches to modelling this behaviour empirically. Subsequently, a summary of the methods presented is provided and related to the general goals of the thesis. This will inherently involve an allusion to the data prior to the formal discussion in the data chapter, though this will be at a very general
level of discussion, rather than the specific features of the data used in this thesis.

3.2.1 Microeconomic Foundations

It is widely understood that the demand for transport primarily stems from the desire to participate in activities or purchase goods (see for example Ettema & Timmermans (1997), for more comprehensive discussion on the role of activity in transport analysis). Transport is indeed a necessary component in the production, delivery and consequently consumption of any good or service, thus its demand is nearly always derived from the demand for these other goods (Ortuzar & Willumsen (1994), Hensher & Button (2000) and Hibbs (2003)).¹ In what follows, a fundamental theoretical framework supporting for travel demand is offered using standard microeconomic foundations, based closely on the work of Jara-Diaz (1998).

The treatment of quality and income as functional determinants of the level of utility are intuitive and commonplace in economic literature. A standard framework serves as the starting point for analysing consumer behaviour in a world containing a finite bundle of goods (which can be consumed continuously, not just in single units). It is assumed that a rational individual will wish to maximise his/her utility subject to their budget constraint and will consequently choose the mixture of goods that achieves this, noting that any finite bundle will have a finite number of potential permutations offering different levels of utility.

Lancaster (1966) further extends this general framework by identifying that consumers are not so much interested in the good itself, but alternatively by the char-

¹There is some literature surrounding aspects of direct demand for travel, see for instance Mokhtarian, Salomon & Redmond (2001), though due to their marginality these issues are not given further attention.
characteristics which that good has, such that utility is derived from the characteristics a good offers (e.g. ride quality, air conditioning or safety features in the context of transport). Consumers therefore purchase the goods available on the market which offer them the characteristics they desire, thus achieving utility gains, as a result the demand for a good is a function of the quality of that good. The purchase of goods is constrained by the individual's level of income.

Denoting the (qualitative) characteristics of the good by the behaviour of an individual is defined as;

\[
\max U(X, Q_j)
\]

Subject to

\[
\sum P_i X_i + c_j \leq I
\]

Where \( j \in M \), i.e. a subset of all alternative travels modes in the set M. This relationship implies that the individual will jointly maximise the utility derived from the bundle of goods and the characteristics those goods offer, subject to a standard budget constraint stating: The level of income \( (I) \) must be equal to or greater than the costs incurred in consumption e.g. you cannot consume more than you can afford. Where consumption of good \( i \) is measured by the price of the good \( (P_i) \) multiplied by the quantity of the good consumed \( (X_i) \) less the travel cost\(^2\) \( (c_j) \) incurred in obtaining good \( i \) using travel mode \( j \).

\(^2\)Often represented in extended versions of this framework by 'generalised costs' which incorporate more explicitly the direct and indirect costs of travel, including the value of time spent travelling (represented via opportunity cost) and a time constraint.
CHAPTER 3. MODELLING TRAFFIC GENERATION

This introduces the important recognition that the standard budget constraint procedure in classical consumer theory should incorporate both the direct and indirect costs involved in purchasing a good and that this will have implications on the utility gains achieved. Furthermore, it can be recognised from this relationship that applied modelling can become very time consuming and data intensive very quickly. Even if analysis is restricted to just a single type of good, there are many alternative travel modes available, each requiring individual calculation (and the issue of generalised travel costs are not even covered here).

The problem is solved for $X$ conditional upon $J$ such that conditional demands are produced:

$$X_i(P_i, I - c_j, Q_j)$$ (3.3)

This can then be placed into the original utility function to produce an indirect utility function of the form:

$$X_i(P_i, I - c_j, Q_j) = V_j$$ (3.4)

representing the maximum attainable utility using travel mode $j$. It can therefore be seen that the mode of transport chosen will be dependent on the costs of each mode, and the chosen mode will be the mode which offers a higher utility level than all other modes;

$$V_j > V_i, \forall i \neq j$$ (3.5)

Not all arguments in $V$ will necessarily influence travel mode choice. The portion of $V_j$ which determines the mode choice (i.e. the discrete comparison) is a truncated utility $\bar{U}$. This framework clearly identifies that the demand for transport is derived, and not demanded for its own sake, moreover it is treated as a cost which
Given this general utility maximisation approach to representing the demand for travel and activities, the most common methods employed in modelling these traffic flows and behavioural demand patterns will be discussed in the next section.

### 3.2.2 Empirical Modelling Approaches

This area of modelling (i.e. trip modelling as opposed to distribution modelling for instance) represents a large proportion of the transport literature and focus of transport models as it forms the basis from which all full transport models evolve. This in itself is a point requiring further expansion, as the goals of the research will have some bearing on the model used to predict and/or forecast trip rates. This point will be given more attention in the concluding remarks to this chapter.

The concept of a trip is intuitively simple though it is worth stating the obvious to ensure all ambiguity is removed. A ‘trip’ is the movement of any individual (be it by themselves or as part of a group) from one point to another, extending from this is the notion of a ‘journey’, which is a series of trips that starts and ends at the same origin. McLeod & Hanks (1986) among others recognise that a journey will often have a specific purpose (perhaps even several purposes in extended ‘trip chains’).

The term trip or journey does not directly infer any specific mode of travel, though it may be natural to consider trips to include only vehicular movements. Travel by foot or by push-bike still constitute as modes of transport and are thus encapsulated within the term ‘trips’. However, empirically they often receive less attention in the literature (or at least have done historically) as movement by foot/bike is often less well recorded, resulting in less data for meaningful analysis of these

is therefore inversely associated with the overall level of utility which can be achieved.
modes of transport having been provided by the literature. Furthermore it may be argued that the implications of any trips that can and are made by foot/bike are of less consequence in terms of the political agenda of the wider transport planning process, and can to some degree be justifiably overlooked. It is however important to recognise recent developments such as the TRAVL (www.travl.org.uk) database is working hard to help dispel this myth and is consciously expanding records on movement of foot travellers.

Ortuzar & Willumsen (1994) point out that there are two fundamental ways to analyse trip rates; (i) there are models which consider the relative frequency of trips for a specific purpose, which attempt to understand the determinants of trip frequency, often for a certain purpose such as shopping, (ii) there are models that consider the overall number of trips originating from a certain zone. This second type of model is often preferred in the literature as it provides scope to extend the results more comprehensively into later stages in the four step model, such as distribution and assignment. Models of trip frequency often find themselves incapable of offering insights into the distribution of trips due to the inherent flaw in their structure. This is to say that the narrow focus of trip frequency models to create better understanding of trip rate patterns for individual purposes almost necessarily removes the ability to offer substantial insights into the overall transport model.

Trip frequency are as necessary as wider reaching origin/destination models, as both provide information at different levels, and scope for alternative extensions them serve only to benefit each other. It should not be thought that a model incapable of accurately predicting trip distributions over a sizable proportion of a network is of no use. As will be discussed further below, such models may lack in extension in relation to the FSM, but provide more accurate understanding of the
determinants of trips by purpose. The result of which enables planners to provide
deeper insights into development control and policy effectiveness at existing sites.

Due to the relative lack of robust multi-modal data and the fact that the ma-
jority of travel is conducted by car, the focus of most trip models remains primarily
upon defining vehicle trips. Though most research will at least recognise this as a
deficiency.

It is necessary in this thesis to recognise the various different types of trips, and
the notations that are given to them. Trips can either be ‘produced’ or ‘attracted’,
and are generally regarded as Home Based (HB) or Non-Home Based (NHB). The
base end of the trip relates most directly to trip production modelling. Such models
are concerned with the factors determining the need/desire to leave the origin zone,
where the origin zone will be from the home (HB), or from another place (NHB).
It should be noted that recent models, such as the UK National Trip End Model
(NTEM) consider (HB) trips as any trip that originates from a house, not necessar-
ily your own.

Extensions to the (HB) and (NHB) notations are often made in models consid-
ering larger proportions of a network, such as Household Based Shopping (HBS) or
Household Based Work (HBW), as well as Non-Household Based Shopping (NHBS)
etc. These are notations that often appear in the transport modelling literature
and provide a neat way of identifying the type of trip under analysis. Further these
classifications clearly support the desire to disaggregate travel analysis by the activ-
ity which that trip is intended to fulfil.

\[3^3\] For a practical example see the DfT’s national trip modelling software TEMPRO available at
www.tempro.org.uk
Trip generation modelling sometimes overlooks the use of the notations mentioned in the previous paragraph as they are models that concern themselves with the interactions between the determinants of trip attractors and the total number of trips attracted to a site (trip frequency), and are often less concerned with the origins of the occupants.

In terms of modelling trips, irrespective of the trip purpose, there are a number of commonly accepted methodologies which will subsequently be discussed in this chapter including:

- Cross Classification/Category analysis
- Regression analysis (standard linear regressions as well as discrete choice models)
- (Non-Parametric) updating methods (Bayesian updating and Growth factor modelling)

Growth factor analysis is an approach that finds some use in the transport literature, helping to identify consistent/emerging patterns in changes in the level of trip productions or attractions over time, or to help predict future values for ‘hard to model’ pieces of information. This concept will be introduced prior to discussing the UK National Trip End Model (NTEM) in which such methods are applied.

Cross Classification analysis

This approach is most well known in the UK due to the pioneering work of Wooton & Pick (1967), in which a large dataset of individual trip surveys is analysed in an
effort to better understand trip generation relationships. This form of analysis fo-
cuses on exploring and emphasising the relationships observed in relation to different
classifications or categories. Hence the two alternative names. Cross-Classification
is generally reserved for the UK, whilst Category Analysis is the preferred term in
the USA, both refer to the same fundamental modelling technique.

The original model takes the following form,

\[ t^p(h) = \frac{T^p(h)}{H(h)} \]  \hspace{1cm} (3.6)

Where the term \( T^p(h) \) measures the average number of trips by purpose \('p'\), and
by members of household/individual type \('h'\). The number of types \('h'\) used in the
analysis is dependent on the researcher's preferences, and is determined often by
classifications recording the level of car ownership, and the household size. For now
the presentation of this model will retain focus upon household analysis, methods
considering individuals will be considered later.

\( T^p(h) \) is the total number of trips observed by purpose \('p'\) and group type \('h'\)
and \( H(h) \) is the total number of households by group type \('h'\).

The model can be written as;

\[ O_i^{np} = \sum_{h \in H(h)} a_i(h) t^p(h) \]  \hspace{1cm} (3.7)

Which states that the total number of trip origins \( O \) in zone \( i \) by group type \( n \)
for purpose \( p \), is equal to the sum of the total number of households of type \( h \) in
zone \( i \), \([a_i(h)]\), and the total trips for purpose \( p \) by household type \( h \), \([t^p(h)]\). The
term \([h \in H(h)]\) states that each household type \( h \) is a subset of all households of
type \( h \) with person of type \( n \). This model therefore creates trip origin values (e.g.
trip generations) that differentiate first by household type, then by trip purpose.

Ortuzar & Willumsen (1994) define the 'art' of this method as choosing categories for analysis such that the standard deviations of the frequency distributions for each group are minimised (similar to tree based regression analysis Washington (2000) and iteration methods for solving general equilibrium models. Thus to some degree, there is potential for data-mining to produce meaningful results, this is a point that will be discussed in more detail below.

This sets out the original form of the Cross-Classification model, key advantages and disadvantages are recognised by Ortuzar & Willumsen (1994);

Advantages;

- This analysis method removes itself from the restrictions of incumbent/existing zone boundaries. That is to say, this approach defines its own groups for analysis, and focuses on more fundamental relationships, that an analysis which restricts itself to zonal boundaries may be unable to elucidate. It should be noted that due to the data requirements, this approach is still able to create an origin destination matrix (based on pre-determined zones), and can therefore re-create incumbent/existing boundaries. Thus this approach is able to provide a level of information that both overlooks and recognises the impositions of set zones.

- There are no prior restrictions imposed upon the relationships arising from this method. To this extent it is happy to let relationships be what they appear to be from the model output.

- Relationships are able to differ from class to class. This approach has the
ability to differentiate between group types, and observe the relative change in relationship compared to other group types.

The third advantage presented above serves to be one of the key attributes of the Cross classification approach. Data requirements have not yet been mentioned, but it seems evident that they are unduly large. In order to be able to create multiple classifications and be able to draw meaningful conclusions from their trip rate levels (recognising that the data is separated by group type and then by purpose type to produce a great many classifications) requires a substantial amount of data. This is as much an advantage as a disadvantage.

The relative simplicity in which categories are analysed and trip rates produced provides for quick and easy differentiation of relationships by different groups. Dependent on the attributes defined to produce the initial categories for analysis, this provides some degree of bearing for policy discussion. By analysing the model output to observe which attributes have the largest impact upon trip rates, policy makers are able to forge a better understanding of how to best impact upon traffic levels in the analysis area.

Another data point which should be raised is the omission of time dynamics in this original approach. It is a relatively strong (and untestable) assumption which is made by this model that the trip generation rates remain stable over time.

The major disadvantages of Cross-Classification analysis are;

- The model provides no allowance for extrapolation in order to expand the analysis beyond the scope of the initial dataset.
- It is not possible to measure the goodness-of-fit for this model
CHAPTER 3. MODELLING TRAFFIC GENERATION

- As mentioned above, this model requires large inputs of data to produce meaningful results, as enough data must be present for each category analysed (at least thirty, preferably over 50).

- Attributes and characteristics are chosen arbitrarily as this model provides no accurate way of establishing the most appropriate attributes for analysis.

Some of these issues have been responded to by systematic extensions of the basic model.

So far the model has neglected to introduce any element of time dynamics. Wooton & Pick (1967) recognised this and defined an amended approach using the parameters of the calibration data (the initial dataset) in order to expand the dataset. The subsequent method provides scope for expanding the data in breadth (e.g. using the attributes of areas outside of the geographical zones of the calibration data to create trip rate data for those zones) and over time e.g. applying growth rates to the attributes of zones within the calibration dataset in order to create forecasts for future periods. Note that based on the assumption given earlier that trip rates remain stable over time, Wooton & Pick (1967) concentrate on predicting the number of households in each category.

The method used for predicting households in the future is as follows; Firstly, the probability density functions for the chosen attributes in the model are assessed (e.g. car ownership and household size levels). This information for each attribute is then used to create the joint probability functions belonging to each household type; \( h = (a_1, a_2, a_3) \) where \( a_1, ..., a_n \in A \), and \( A \) is the set of attributes used in the analysis. Wooton & Pick (1967) and Ortuzar & Willumsen (1994) recognise that the set \( A \) should remain quite small due to the increase in number of categories for analysis that would arise from the addition of any new variables (if Wooton & Pick
(1967) added one more variable to their study, the number of categories would have tripled from 108 to 324).

Multiple Classification Analysis (MCA)

Stopher & McDonald (1983) pioneered MCA in the transport literature as an improvement to the standard Cross-Classification analysis using a technique that has been used in the social sciences for some time.

This approach carries several advantages over the basic model in that its formulation overcomes the following difficulties:

- MCA is better equipped to handle small sample sizes without the need for expensive stratified samples being conducted.
- The MCA method provides more scope for sound statistical assessment of the model results.
- MCA works on a framework that provides the ability to statistically differentiate between independent variables and establish groups from within each independent variable.

Assuming the model contains a continuous dependent variable and $x$ discrete independent variables (such as trips for the dependent, and car ownership or household size as the independents), the method works in the following manner:

1. Compute the 'Grand Mean' for the data. - This is the mean value of the dependent variable using the entire survey data.
2. Estimate the 'Group Means' for the rows and columns of the cross-classification matrix. These should be expressed as deviations from the 'Grand Mean' value.

3. Cell values are estimated using the 'Grand Mean' and 'Group Means' - The deviations given by the 'Group Means' (for the relevant row and column) are added to the 'Grand mean' for each cell in the matrix.

It is clear that step three enables this model to overcome some of the flaws of small sample sizes by considering mean values as opposed to absolute values. This technique is not ideal, as the sample used may be misrepresentative of the population, leading results to be questionable still. However by incorporating the means in an alternative way to standard Cross-Classification analysis and increasing the amount of information used for each cell, some steps are made in overcoming data shortages.

One of the biggest advantages of MCA in fact, stems from this ability to overcome data shortages, as the process of using 'Group Means' enables the MCA method to provide some information for cells that would appear empty in a standard Cross-Classification matrix. Ortuzar & Willumsen (1994)(pp116-117) provide an example of this.

Parsons Brinckerhoff Quade & Douglas et al. (1995) in developing an overall transport model for the state of Oregon, USA, identify the use of MCA in the trip generation stage of their model, along with various curve fitting techniques used to account for data inadequacies.

A note on Regression Techniques in Cross Classification Analysis.
One of the flaws of the Original model by Wooton & Pick (1967) is its inability to clearly identify the attributes which should be used to classify the groups in the analysis. In response to this standard regression techniques have been applied in order to assess the degree of relationship between the dependent variable and potential independent variables. The level of association relating to each variable is assessed statistically in order that the most statistically preferred variables can be used for the Cross-Classification analysis.

The regression system is only used to help the researcher decide upon which attributes to include in the Cross-Classification of MCA analysis. Once these attributes have been ascertained, the analysis is then conducted in the standard way as set out above.

**Person-Based Category Approach.**

In likeness to the household based Cross-Classification analysis of Wooton & Pick (1967), Supernak (1979) and Supernak, Talvitie & DeJohn (1983) took the analysis one step further by focussing on individuals as opposed to households. Some data concerns are raised relating to how household cost and income structures are incorporated into an individual level model, though Supernak et al. (1983) respond to this by highlighting the ambiguity of such variables in the original model.

Supernak et al. (1983) suggest that the Person-Based approach offers deeper behavioural insights, and that it is less data intensive. Furthermore it is able to offer better contributions to the overall transport model as individuals can be better traced across the transport network, and individual travel patterns better understood. This much is true, however, the relative strength of the Person-Based
CHAPTER 3. MODELLING TRAFFIC GENERATION

approach is dependent on the goals of the individual research, and the level of data intensity is an issue of choice.

In Person-Based models the following relationship is considered;

\[ T_i = N_i \sum_j a_{ji} t_j \] (3.8)

Where \( T_i \) captures the total number of trips in zone \( i \), and is computed as the total number of individuals in zone \( i \) multiplied by the sum of the percentage of individuals of type \( j \) in zone \( i \) multiplied with the total number of trips by type \( j \).

In application Supernak et al. (1983) apply a regression analysis first, so to identify the most appropriate attributes for analysis and then go on to conduct Classification analysis following the framework provided above. Note that this model, although not derived here, retains the ability to differentiate by trip purpose.

**Regression Analysis**

The body of literature relating to purely to econometric techniques is extensive and good examples include Greene (2003) , Gujarati (2003).

Regression techniques in the field of transport analysis generally use relatively straightforward regression techniques, at least that is, in relation to trip rate/generation modelling see for example ITE (1979) or ITE (2001). Models will generally be multiple regressions including variables measuring factors such as household composition, income, car ownership etc. A selection of site specific variables will often be included also.
CHAPTER 3. MODELLING TRAFFIC GENERATION

It should be recalled from earlier that trip generation models can look into two fundamental relationships, namely 'trip productions' and 'trip attractions'. Different forms of regression technique can be applied in looking at each type, due to the substantially different type of relationship and data requirements that are often applied.

In analysing trip productions, the aim of the model is to establish relationships in the decision to make a trip. Consequently 'trip diary' type data is generally preferred in order to understand how many trips are made by each household/individual in a given time frame (trip frequency). In such analysis it is common to observe the use of discrete choice models to establish what affects the propensity to make a trip at the household/individual level. It is also still feasible to continue to use standard regression models to model trip productions instead of discrete choice models, though it is not so suitable to use discrete choice models for trip attraction modelling.

Trip attraction models require data on the rate of trips observed at a given site across a certain period of time. Although this is once again a form of trip frequency, the data offered to analysts often provides no way to differentiate between individual households, therefore it is not generally possible to include household level factors that differentiate across one geographical zone. In general trip attraction models will apply standard regression techniques using Ordinary Least Squares (OLS) estimators.

At the zonal level, Ortuzar & Willumsen (1994) propose a model of the general form;

\[ Y_i = \theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i} + \ldots + \theta_k X_{ki} + E_i \]  

(3.9)

Using total levels for the area (e.g. trips per zone and cars per zone etc.). This
can be re-written using aggregate level variables such as cars per household and then takes the form:

\[ y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_k x_{ki} + e_i \]  

(3.10)

Where \( y_i = \frac{Y_i}{H_i}, x_i = \frac{X_i}{H_i} \) and \( e_i = \frac{E_i}{H_i} \) with \( H_i \) being the total number of households in zone \( i \).

This specification introduces an issue of scaling, which is not essential to keep uniform throughout a model specification. For instance a model of the following form could be specified using zonal levels for the dependent variable and household levels for the independent variables;

\[ Y_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i} + \ldots + \theta_k x_{ki} + e_i \]  

(3.11)

Such a model requires more careful interpretation than the pure zonal or pure household level models (as coefficients need to be re-scaled appropriately for interpretation), though this does not reduce the accuracy of the model.

However, due to the ease in interpretation offered by using a uniform scale across the variables, this approach is generally preferred where feasible. Ortuzar & Willumsen (1994) explicitly state, "What is certainly unsound is the mixture of means and aggregate variables in a single model" (pp107). In terms of the models defined later in this research, it can be shown that they are pure household level models, as no variables are entered in a purely zonal level.

The analysis of variables in a trip rate regression at the household level can help to overcome problems arising from heteroskedasticity, which is done implicitly by the removal of the influence of zone size upon the variability observed among the
regressors. This is to say that by reducing the scale of variables down to the common household within each observation area, will likely remove the effects of bias in the standard errors arising from the comparison of large versus small areas (in aggregate terms). The simple process of dividing through by the number of households reduces the variability in total terms, whilst retaining the individualism of each observation.

Regression techniques have also in the past included pure household level studies. This is different to the notion presented thus far, in that these models do not even attempt to acknowledge geographical zones. This is somewhat analogous to part of the discussion presented relating to Cross Classification Analysis models in the previous section, in which the methods of Wooton & Pick (1967) were evolved to analyse primarily the individual level relationship between observed interactions, and then aggregate these up to geographical zones later. Similar methods have been applied in regression analysis, where the early 1970's saw the recognition of households as the most appropriate unit for analysis see Ortuzar & Willumsen (1994), pp108. The reasons behind this being that the geographical zone may act as a constraint in terms of data collection that could potentially hinder the ability of a model to measure and accurately understand the fundamental relationships in trip rates. Alternatively it was also recognised that to concentrate on each individual may prove to be more complex than is practical for the needs of the transport industry. There are a number of wider issues that each household is subject to, however these responsibilities are often distributed among the individuals within each household. Consequently, the responsibilities of each individual varies in relation to the size of the household and other complex social issues. As a result it is prudent (and less costly in terms of time and monetary expense) to collect data at the household level and not concern oneself with the complexities of the individual interactions within. At any rate, no feasible model exists to analyse the interactions of the individuals
within, at least not in any way conducive to the wide-scale analysis of transport systems.

One of the drawbacks encountered in regression analysis is the issue on non-linearities in observed relationships. These can arise due to a number of factors, and can either be quite easy or highly complex to overcome. Fortunately, in models of trip generations the most prominent source of non-linearity will arise from factors which can be 'dummied out' via, for example, interactive dummies. Such that a dummy variable is created to capture qualitative differences relating to differing areas of the dataset. For instance the data may allow the ability to differentiate between households with less than one car (i.e. 0 cars) and households with one or more cars (i.e. greater than 0), in which case a dummy variable can be created to observe any change in trip rate patterns by the group with access to a car. The dummy variable in this instance would take the value of zero for households with no car and one for households with a car, and works by adding a factor to the intercept coefficient (or the slope through interactive dummies), shifting the entire model to a point either above or below that observed by the base group, hence providing the ability to capture non-linear effects. The use of dummy variables creates a number of potential pitfalls for the modeller via the dummy variable trap. It is fairly easy to specify an incorrect number of dummies, and the relationship between these dummies and the constant term in a regression model is of paramount importance. In practice however, most packages will realise any such pitfalls, and the package used will simply not estimate the model.

Household level analysis has one further advantage in that it allows for the ability to aggregate the data to obtain totals for predefined geographical zones, thus allowing comparison to pure zonal models.
CHAPTER 3. MODELLING TRAFFIC GENERATION

Discrete Choice Models

An alternative regression technique applied to the analysis of trip rates involves a subset of econometric models referred to as discrete choice models, which focus on the analysis of finite outcome, or non-continuous variables.

Before developing a basic framework for discrete choice analysis, it serves well to discuss the difference in terms of data requirements in these models. As mentioned above, the term 'discrete' refers to a variable which is non-continuous in nature, which means that the estimation technique used must be altered appropriately. The dependent variable will, in the instance of trip generation models, measure the amount of trips made by an individual household on a given day for instance and place them into categories. For simplicity of theoretical exposition, the number of categories will be kept to two;

1. the household makes no trip on a given day,
2. the household makes at least one trip on a given day.

Though this could for instance be expanded out to provide a more detailed analysis. The discrete choice analysis framework then proceeds to use a set of independent variables, which can be either continuous or discrete in nature, and applies the maximum likelihood estimation technique to assess the maximum likelihood of observing an individual that has made 0 trips, given the data observed in the set of independent variables.

The standard model takes the following form;
\[ y^* = X\beta + \mu \] (3.12)

With

\[ y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \] (3.13)

Defining;

\[ P_i = Pr(y_i = 1) = Pr(y_i^* > 0) \]
\[ = Pr(u_i > -x_i'\beta) \]
\[ = 1 - F(-x_i'\beta) \] (3.14)

\( F(x) \) is defined as the cumulative density of \( x \). The following statement is provided for any symmetric distribution;

\[ F(x) = 1 - F(-x) \]

so that,

\[ P_i = F(x_i'\beta) \] (3.15)

as

\[ 1 - F(-x_i'\beta) = F(x_i'\beta) \]

Assuming a logistic distribution for the model (noting that the normal distribution is a commonly used alternative choice for the distribution) the following is specified;

\[ F(x_i'\beta) = \frac{1}{1 + \exp(-x_i'\beta)} \] (3.16)
For this 'logit' model (as it is often referred to), the log-odds ratio is thus defined as:

\[
\log \left( \frac{P_i}{1 - P_i} \right) = x_i'\beta = \Gamma_i \tag{3.17}
\]

The estimator \( \hat{\beta}_j \) can now be written,

\[
\hat{\beta}_j = \frac{\partial \Gamma_i}{\partial x_{ij}} \tag{3.18}
\]

This estimator captures the effect a change in \( x_{ij} \) has upon the log-odds ratio.

Interpreting this model is more intuitive than the mathematics behind it suggests. Essentially, the probability of a value of one occurring (which in our simple two value dependent variable case would mean the probability that an individual makes a trip) is equal to the probability that the observed value of the dependent variable is greater than one. One drawback of such models is the relatively strict imposition of functional form. Ben-Akiva & Lerman (1985) provide a comprehensive text defining the theoretical and practical application of discrete choice methods, with a particular focus on transport related studies.

Upd.ating methods.

The methods to be discussed below are focussed very much upon observable and forecastable dynamic relationships. Growth factor modelling, as its name implies, aims to determine the rate of growth for any given dependent variable, based on historic data. Methods for doing so can be either simple or complex. Ortuzar & Willumsen (1994) discuss the need to have such tools available and outline the method presented below for growth factor modelling. The need to forecast information in the transport industry is of increasing importance, not only due to the
need to guide policy for central as well as local governments. Also, there is a need to be able to predict the future effects of current trends, such that if current growth rates continue, it is possible to begin to quantify the likely impact upon society, thus providing information as to when policies need to be placed into effect, and to what extent they need to be enforced. Models applying Bayesian updating methods for parameters stem primarily from game theoretic applications, which was the first discipline within economics to apply Bayes theorem. The intuition behind Bayesian updating is that individuals observe behaviour over time and slowly adapt expectations. These expectations change over time, hence the term updating, as with every forward movement in time the information set available to the individual changes, causing them to have a slightly different outlook on the world. As a result the patterns linking determinants to behaviour changes over time.

Applications of Bayes theorem are numerous and diverse. Game theoretic applications are not considered here, only econometric type applications are introduced. Relationships between Bayesian updating and general growth factor models will be highlighted, with the recognition that both approaches are fulfilling the same fundamental task, though with different degrees of complexity, providing differing results.

Following the discussion on updating methods the UK National Trip End Model will be introduced and briefly discussed to indicate the general state of the art of modelling practice at present. In this discussion it will become clear how many of the methods discussed here feature in some way within the large and complex model used to understand travel behaviour in the UK.

**Growth Factor Methods**
The following method is taken from Ortuzar & Willumsen (1994) and provides one approach to modelling growth factors. Many alternative approaches exist, ranging from the simplicity of linear extrapolations through to the numerous macro-econometric growth models, utilising a wide array of theoretical backgrounds and estimation techniques.

What is presented here is a general, and relatively simple, growth factor model, which steps a level beyond the naive simplicity of some linear extrapolation techniques (e.g. an average growth rate across the entire history of the dataset), though does not go as far as the complexities of for instance the Bayesian updating methods discussed later in this chapter. This method is however still considered a somewhat crude approach, though it does at least make initial efforts to create a more theoretically sound model.

What is presented here is essentially a linear extrapolation model, though it is advanced enough to consider more information than only the previous period’s level of growth. The basic model is;

\[ T_i = F_i t_i \]  
(3.19)

Where \( T_i \) is the future level of trips, \( F_i \) is the growth factor and \( t_i \) is the current level of trips are relating to zone \( i \). The key to this model is estimating the growth factor.

\[ F_i = \frac{f(P_i^d, I_i^d, C_i^d)}{f(P_i^c, I_i^c, C_i^c)} \]  
(3.20)

Equation 3.20 is generally applied, stating that the growth factor is a function \( f \) of the design year’s population \( P_i^d \), income level \( I_i^d \) and car ownership levels \( C_i^d \) divided by a function \( f \) of the current years population \( P_i^c \), income \( I_i^c \) and car own-
ership levels $C_t^c$. The form of this function is however not clearly defined. Ortuzar & Willumsen (1994) suggest that this can simply be a direct multiplicative function for instance, thus not requiring any parameters.

In application this model exemplifies its weaknesses as Ortuzar & Willumsen (1994) show. Strong assumptions are made, such as full car ownership in future periods, meaning that results can prove to be ambiguous. For instance, the authors assume that firstly in the future, all households will own a car, then based on the assumption that income and population levels remain constant, the growth rate in trip rates is estimated directly as the growth rate in car ownership.

Example 1.

If in the current period, 74% of households own a car, the population level is 5000 households (3700 with cars and 1300 without) and the income level is ignored, the current level of trip generations by group is known to be:

- Car owning households: 5 trips per day
- Non car owning households\(^4\): 2 trips per day

Then the current trip generation rate can be defined as:

$$t_i = (3700 \times 5) + (1300 \times 2) = 21,100$$

If the assumption is made that population and real income remain stable over time (thus allowing these to be factored out of the equation), the growth factor can

\(^4\)Where households without cars in the current period sustain their travel activities using public transport, walking, cycling, taxis, car-sharing etc.
be defined;

\[ F_i = \frac{(C_i^d)}{(C_i^f)} \]

Also, for the design year it is assumed car ownership will have risen to 93%, the growth factor works out to be;

\[ F_i = \frac{(C_i^d)}{(C_i^f)} = \frac{0.93}{0.74} = 1.26 \]

From this it is possible to predict the design year’s trip generation rate;

\[ T_i = F_i t_i = 1.26 \times 21,100 = 26,586 \]

The assumptions made in this model are reasonably intuitive, as population levels and real income do not change by significant amounts over short to medium term time horizons. As shown in example 1 above, the application of this model under these assumptions is somewhat crude but reasonably effective. However, as will be shown below in an extension to example 1, the linear growth factor method presented is subject to scrutiny in terms of accuracy. This results in application of this model only being used to predict external trip generations to the analysis zone.

Example 2.

Assuming now that in the initial period, car ownership is 93%. In this case there are 4650 households with cars and 350 households without cars. Following the same assumptions from example 1 about trip generation rates by household type;

- Car owning households: 5 trips per day
- Non car owning households: 2 trips per day
i.e. recreate the end situation for example 1 as the start point for example 2.

The daily trip generation rate using these data is:

\[ t_i = (4650 \times 5) + (350 \times 2) = 23,950 \]

This is the level of trip generations that will occur using the mix of population (differentiating by car ownership status) at the end of the sample (e.g. in the design year). The difference between this value and the original base year trip generation level from example one serves to be a zone of discrepancy. Primarily because this method makes no attempt to understand the dynamics of the population distribution in terms of car ownership status. Furthermore, this difference is regarded as an overestimate and consequently an error of the growth rate in trips. This stems from the intuition that this portion of the increase in trip generation is the source of a change in the proportion of car ownership. In order to retrieve the growth in trips more directly attributable to 'actual growth', the final estimate from example 1 must be subtracted from the estimate above (the example 2 estimate).

The difference observed is a result of the imbalance in car ownership levels being followed through the entire forecast period (e.g. the ratio of trip patterns in the base year is dependent on one distribution of car ownership, which is used to estimate a final trip generation rate in the design year, which is for a different car ownership distribution), as a result errors are created and exaggerated.

In this instance nearly 52% of the estimate of growth in trip generations is subject to error due to the reasons outlined above. This is alarmingly high, especially for such a crude technique.
Example 2 exemplifies the need to consider more complex methods, in particular it highlights the potential benefits of introducing updating methods which are able to constantly review forecast parameters at each new time period within the dataset. As will be discussed below, this will help to overcome the inefficiencies of comparatively simple linear extrapolation methods.

For the reasons outlined above, this model receives as little use, as these errors will be passed on to subsequent stages in the overall transport model. In practice, it is used for modelling growth rates on external trips into the analysis zone. These values are empirically quite small, therefore reducing the level of error arising in their estimation. This is not ideal, however it suits, as the factors influencing external trips are too complex to model easily, and as mentioned, values are small so error levels are deemed acceptable.

Bayesian Updating

As mentioned previously, this approach stems from the original Bayes theorem. It serves no real purpose to formally express the original theorem however the fundamental tenet of the theorem is that information of the past is observed which is used to develop a prediction of what the future may look like.

These updating methods use information from the past in order to inform behaviour in the future, denoted respectively the prior and posterior. In terms of time-series applications, in particular growth rates, the distribution of growth rates observed at every point of time within the data set is recorded and used to form a parameter estimate for the current (or future) periods growth rates. This is done in a stepwise fashion for each time period in the dataset, sequentially growing the
dataset from very few observations (there is arguably little value in starting with just one observation), right through to the very end of the dataset. As the dataset grows, the distribution pattern in growth rates will continue to change (marginally as the dataset gets large), as a result the parameter estimate will also change. Therefore the Bayesian approach produces a non-linear time variant parameter estimate. One issue that arises in time-series application of Bayesian updating is exactly how much past information to use, and what (if any) weights should be placed on the age of the data. These are issues that should be considered more carefully in the context of each individual piece of research, though guidelines are not set in stone, and this may often resort to intuition and arbitrary (scientific) choice.

Bayesian updating can also be used for more general updating of information in a framework less consistent with time-series analysis (i.e. a cross section). For instance; building trip matrices in the presence of insufficient regional information or funds to obtain the necessary information. Bayesian updating techniques can be applied to update data from a similar region, or from historic data on the base, using only a small amount of data on the base zone in the current (e.g. a stratified sample survey, in order to keep costs down).

Presented for our needs in transport;

\[
t_2 = \frac{1}{1/\sigma_1 + 1/\sigma_s} t_1 + \frac{1}{1/\sigma_1 + 1/\sigma_s} t_s
\]  

(3.21)

with variance;

\[
\sigma_2 = \frac{1}{1/\sigma_1 + 1/\sigma_s}
\]  

(3.22)

\(t_i\) represents the mean trip rate, \(S_i\) is the trip rate variance and being the number of observations. Subscript 1 denotes the prior information (i.e. from the historic
data or other region), subscript 2 denotes the posterior information whilst subscript s denotes the new information (i.e. from the stratified sample).

From equation 3.21 future period trip rates are estimated, based on the proportionate variances of the distribution of all available data, where the term $\frac{1/\sigma_1}{1/\sigma_1 + 1/\sigma_2} t_1$ multiplies the prior level of trip generations by the proportion of total variance in all available information directly attributable to the variance in the prior information. Likewise, the term $\frac{1/\sigma_2}{1/\sigma_1 + 1/\sigma_2} t_s$ multiplies the level of trip generations in the new data by the proportion of total variance attributable to the variance in the new information. This all holds true so long as the prior and new sample distributions are normal, which so long as a suitable number of surveys are conducted (more than thirty) they should be.

Inputting the variables $S_i$ and $n_i$ into equations 3.21 and 3.22, noting that the variance in trip rates is given by $\frac{S_i}{n_i}$ yields;

$$t_2 = \frac{n_1 S_s t_1 + n_s S_1 t_s}{n_1 S_s + n_s S_1} \quad (3.23)$$

and

$$\sigma_2 = \frac{S_1 S_s}{n_1 S_s + n_s S_1} \quad (3.24)$$

This approach works on estimating means, and this is a fundamental pre-requisite due to the reliance upon rules of distributions. As a result, Bayesian updating, as a method to extend information in trip matrices (or even just in one zone), is only able to offer insights at an aggregate level and provides no information on the activities of individual households.

\*Following the central limit theorem
CHAPTER 3. MODELLING TRAFFIC GENERATION

The UK National Trip End Model (NTEM)

The previous discussion has provided a very general review of alternative trip modelling methods, the present sub-section conversely turns attention directly to the current state-of-play of trip modelling in the UK by briefly reviewing the DfT's transport model. The finer details of the model are not well publicised, although the model is known to contain over 300 equations operating in a vast number of different ways to help expand data sets to cover areas where no information is available, measure growth factors so that relationships can be updated over time, provide modal splits for trip generation forecasts etc. The NETM\textsuperscript{6} provides information that later gets fed into the National Road Traffic Forecast (NRTF) of the DfT for areas of the model lacking data from primary sources.

Consequently these are the base models which help provide guidance on the expected future levels of traffic and consequently the political decisions made. The accuracy of these models is expected to be high, as the government will use their results to set future budgets for transport expenditure and provisions for road maintenance.

The NTEM is itself highly complex and combines information from the National Travel Survey (NTS) with National Census data. The data is subject to some limitations which will be more clearly defined shortly.

The aim of the NTEM is to create multi-modal estimates for trip generations differentiated by trip purpose (e.g. shopping, work etc) for all of the major constituencies within the UK. This in deserves further attention as this is a source of

\textsuperscript{6}See www.dft.gov.uk/pgr/economics/ntm/tripendmodel for a discussion of this model on the Department for Transport website.
CHAPTER 3. MODELLING TRAFFIC GENERATION

much complexity inherent within the model. Although the national census data is almost fully representative of the population, the NTS data is conducted using approximately 300,000 surveys. Following the law of large numbers it is reasonable to assume that this is a good representative sample of the population, however it is still potentially misrepresentative of portions of the population, when it is considered that the population is in the region of sixty million. Another data issue which quickly arises is the time consistency of the Census data, which is only compiled every ten years.

These are unquestionably the best sources of data available for this exercise, however it is clear and must be recognised that they are not infallible and may be a source of error within the model.

Without focusing on the intricacies of this model, which are more fully explained in the 125 page guide to this part of the overall transport model, HETA division of the Department for Environment, Transport and the Regions (DETR) (2000), one of the main aims of the model is to be able to provide forecasts for trip generations in the future. A large number of assumptions are made (none too extreme), along with an extensive set of data manipulations in order to define trip generations by purpose type for all areas in the UK in a base year. This data is then subject to growth modelling which forecasts the changes in trip levels over time.

This growth modelling does not seem ideal as it appears to be simple combinations of linear growth factors on variables such as population and car ownership and does not account for potential social changes in trip distribution patterns such as the move from 'own vehicle' travel towards public transport travel. Furthermore the current framework does not provide the ability to factor in changes in modal
split over time, so any observable changes cannot be input by an analyst. Arguably, it is not the scope of this stage of the model to do such things, and this will all be considered in the later stage of assignment, so it may be a little harsh to criticise the model on this issue, however it does seem that the NTEM is a little inflexible to sensitivity testing.

Inevitably the current Trip-end Model will become outdated as trip behaviour patterns of the general public alter, and as policy implementations start to take effect and modal split patterns change. The existing framework will remain capable of producing the necessary results, however its nature is such that it may not be a simple matter of adding the data. This however is more of a software issue than anything else and is a question of whether the data can be updated real time or not.

The NTEM is certainly far too complex to provide a full derivation, or even a basic presentation of all the fundamental functional relationships. Furthermore, the scope of the model is vastly beyond the goals of this research and a full appreciation of the NTEM (although useful knowledge in a wider understanding of transport modelling) is not considered to provide notable benefits for this thesis.

3.2.3 Summary of Trip Modelling Approaches

This section has introduced a number of the most common approaches to predict trip generations and/or attractions. These methods all have some commonalities but also differ in a number of key dimensions and can be considered to be complements as they are alternatives. Evidence of this can be seen in the current UK trip end model NTEM, which is part of the wider national transport model and applies many of the different techniques outlined here.
CHAPTER 3. MODELLING TRAFFIC GENERATION

It is in many cases necessary to have estimates of trip generations that are compatible with later stages of the overall transport model, in particular with the distribution stage. Though with respect to the research goals of this thesis, the desire to estimate a full FSM is immediately precluded by the ability of the data to facilitate it. The same point holds true for consideration of updating techniques, as the data structure is not directly compatible with this. Further the discrete choice methods are ruled out as a preferred modelling approach for similar concerns.

The regression based approach stands out as a flexible approach to trip generation modelling which overcomes some of the flaws of some techniques, such as cross classification analysis. Further, given the decent availability of data in this instance, precludes the need to apply updating methods\(^7\). The discussion on regression based techniques suggests that diary data is needed to model trip generations accurately, therefore when estimating the models for residential developments as will be done later, there may be an issue of accuracy in the results. This is essentially due to the fact that survey data aggregates behaviour whereas diary data directly reflects individuals behaviour based on their individual characteristics. However this potential weakness aside, the regression technique is the most appropriate methodology to help robustly and scientifically scrutinise the hypothesis that land zone type has a statistically observable relationship with travel behaviour, while controlling for socio-economic information and given the data.

The following section reviews the literature surrounding car ownership and how this information relates to travel behaviour.

\(^7\)Though as will be discovered later in the thesis, alternative data considerations/limitations arise with respect to the analysis of office developments
3.3 Car Ownership Literature

Inspection of data\(^8\) from the UK between 1952-2003 (summarised in Figure (3.1)) reveals that a high degree of correlation exists between these aggregate car ownership and use (calculated to be 0.992). Thus supporting, at least at the national level, the notion that car ownership modelling is not necessarily to be thought of as completely separate to trip rate analysis.

Figure 3.1: Car ownership and total distance travelled in the UK. (Data source: UK Department for Transport and Office of National Statistics online data collections)

This section reviews the theory and literature surrounding models of car ownership.

\(^{8}\)Taken from the Department for Transport online data collection at www.dft.gov.uk/pgr/statistics.
ship. The principal aim of such models is to explore the relationships between the
decision to own a car (or several cars), with a selection of defining variables. As
Button (1974) discusses, such models are often used in practice in order to help build
the full picture of travel behaviour. In practice, as will be shown in the following
discussion, standard economic theory is drawn upon to build the relationships that
one would expect to observe in the decision to purchase any good or service, such
as the price effects observed (via own price, substitute price and complementary
good prices) and the income level. Thus such models generally operate in a quite
standard demand modelling framework.

Early car ownership models focussed exclusively upon the decision to own a car,
such as Bates, Gunn & Roberts (1978) and Quarmby & Bates (1970). Aggregate
models of car ownership can be found prior to this in de Wolff (1938) and Chow
(1957) *inter alia* who both consider the demand for automobiles in the United States.
However it was further recognised that the decision to own a car is strongly correlated
with the propensity to travel by car (Khan & Willumsen (1986), de Jong (1989)),
and certainly on an aggregate scale, an observable relationship exists between car
ownership and trip generations.

3.3.1 Microeconomic Foundations

Button, Fowkes & Pearman (1982) provide a short discussion on the theoretical
framework influencing the rationalised decision to own a car, providing information
of key importance in the derivation of *a priori* restrictions in the modelling strategy
section. Two primary variables are considered, namely $B$, the net benefits to house-
hold members of making journeys (which will consider both car and non-car travel)
and $Y^A$, which captures household income after the costs (fixed) of car ownership
are deducted.
Assumptions

1 - The total household utility $U$, is derived purely as a function of these two primary variables.

\[ U = U(Y^\Delta, B) \]  \hfill (3.25)

2 - Any household considering the purchase of a second vehicle will already have prior experience and consequently knowledge of which car and set of journeys would be best for it subject to constraints on time and money (e.g. rationality is not affected by the decision to own more than one car in a household).

\[ \frac{\delta U}{\delta Y} > 0; \frac{\delta U}{\delta B} > 0 \]  \hfill (3.26)

This relationship holds for the decision to purchase a first car as much as it does an additional car.

In this framework, utility gains are derived from one of two sources, either an increase in the 'left over' income (e.g. the income left after fixed costs of car ownership have been discounted) or from a beneficial change to journey making.

A number of other variables can now be introduced in order to help better define car ownership choice.

Income effects

Considering firstly the relatively simple variable of income defining total household income after fixed costs of motoring($Y^\Delta$), as the total household income ($Y$) less
(C(PC)), the fixed costs attributable to car ownership;

\[ Y^\Delta = Y - C(PC) \]  \hspace{1cm} (3.27)

Where the term \( C(PC) \) is a multiplicative relationship between the number of cars owned and \( C \) the fixed cost per car \( PC \) (or price per car).

Differentiating with respect to car ownership,

\[ \frac{\delta Y^\Delta}{\delta C} = -PC; \quad \frac{\delta^2 Y^\Delta}{\delta C \delta PC} = -1 \]  \hspace{1cm} (3.28)

Therefore, the change in income relative to an increase in car ownership is directly proportional to the level of fixed cost.

Journey making

Button et al. (1982) derive 7 properties (summarised in Table (3.1)) in consideration of the net benefits arising from journey making. \( b \), the net travel benefits are defined as a function of the number of cars owned \( C \), the cost of motoring \( PM \), population density \( D \) and public transport accessibility \( A \) such that,

\[ B = B(C, PM, D, A) \]  \hspace{1cm} (3.29)

More conclusive implications can be drawn if the effects of income and travel benefits are considered together. From equation (3.25) i.e. \( U = U(Y^\Delta, B) \) and recalling equations (3.27) and (3.29)), the following can be derived;

\[ \frac{dU}{dC} = \frac{dU}{dY^\Delta} \frac{dY^\Delta}{dC} + \frac{dU}{dB} \frac{dB}{dC} \]  \hspace{1cm} (3.30)
### Property 1
\[
\frac{dB}{dC} \geq 0
\]
- The change in net travel benefits will be positive with increased car ownership.

### Property 2
\[
\frac{dB}{dPM} \leq 0
\]
- An increase in the cost of motoring reduces the net benefits offered to a household as fewer destinations are accessible.

### Property 3
\[
\frac{\partial B}{\partial dPM} \leq 0
\]
- The higher the cost of motoring combined with an increase in car ownership will have a negative impact on increases in net benefits.

### Property 4
\[
\frac{dB}{dD} \geq 0
\]
- Population density fosters development. Amenities will be closer, reducing the need to travel thus increasing net benefits.

### Property 5
\[
\frac{\partial B}{\partial dD} \leq 0
\]
- Increased density leads to a reduction in the rate of increase of travel benefits as car ownership increases.

### Property 6
\[
\frac{dB}{dA} \leq 0
\]
- Net travel benefits increase with a reduction in the generalised cost of public transport.

### Property 7
\[
\frac{\partial B}{\partial dA} \geq 0
\]
- The lower the cost of public transport accessibility (measured via generalised cost), the lower will be the increase in net benefits with increases in car ownership.

---

**Table 3.1: Microeconomic properties of car ownership**

<table>
<thead>
<tr>
<th>Property</th>
<th>Formulation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property 1</td>
<td>( \frac{dB}{dC} \geq 0 )</td>
<td>The change in net travel benefits will be positive with increased car ownership</td>
</tr>
<tr>
<td>Property 2</td>
<td>( \frac{dB}{dPM} \leq 0 )</td>
<td>An increase in the cost of motoring reduces the net benefits offered to a household as fewer destinations are accessible</td>
</tr>
<tr>
<td>Property 3</td>
<td>( \frac{\partial B}{\partial dPM} \leq 0 )</td>
<td>The higher the cost of motoring combined with an increase in car ownership will have a negative impact on increases in net benefits</td>
</tr>
<tr>
<td>Property 4</td>
<td>( \frac{dB}{dD} \geq 0 )</td>
<td>Population density fosters development. Amenities will be closer, reducing the need to travel thus increasing net benefits</td>
</tr>
<tr>
<td>Property 5</td>
<td>( \frac{\partial B}{\partial dD} \leq 0 )</td>
<td>Increased density leads to a reduction in the rate of increase of travel benefits as car ownership increases</td>
</tr>
<tr>
<td>Property 6</td>
<td>( \frac{dB}{dA} \leq 0 )</td>
<td>Net travel benefits increase with a reduction in the generalised cost of public transport</td>
</tr>
<tr>
<td>Property 7</td>
<td>( \frac{\partial B}{\partial dA} \geq 0 )</td>
<td>The lower the cost of public transport accessibility (measured via generalised cost), the lower will be the increase in net benefits with increases in car ownership</td>
</tr>
</tbody>
</table>
CHAPTER 3. MODELLING TRAFFIC GENERATION

If the right hand side of equation (3.30) is positive this would indicate that there are utility gains to be made from increased car ownership.

Let $X$ stand for any of the variables $PM$, $D$ or $A$ and the second order derivative is found to be:

$$\frac{d^2U}{dCdX} = \frac{\delta U}{\delta B} \frac{d^2B}{dCdX}$$

(3.31)

So that an increase in $X$ will cause an increase in $\frac{dU}{dC}$ so long as $\frac{d^2U}{dCdX}$ positive thus increasing car ownership. As (from assumption 2) $\frac{\delta U}{\delta B}$, then it is expected that the terms $\frac{d^2U}{dCdX}$ and $\frac{d^2B}{dCdX}$ will take the same sign.

The following conclusions are then drawn (in relation to properties 3, 5 and 7):

- Increases in motoring costs are associated with reduced car ownership (an own price effect)
- Increases in population density are associated with reduced car ownership (shifting the demand)
- Increases in the generalized cost of transport are associated with increased levels of car ownership (Substitute goods effect)

This theoretical framework is not conclusive, as Button et al. (1982) emphasise. Marginal utility effects differ across various household structures, as different household compositions have fundamentally different demand patterns. Effects like social class are not included, though identified as a likely significant determinant of car ownership.

Attention for the remainder of the chapter from this point forward will be focused on the discussion of empirical car ownership models i.e. pure car ownership.
This will then be supplemented with the discussion of models that consider car use within a simultaneous setting i.e. joint models of car ownership and use.
3.4 Empirical Modelling Approaches

In practice, the approach used to analyse the car ownership choice is very much dependent on the data level at which the study is concerned. Most applications work on a disaggregated level of data, which generally applies the use of cross section or panel data. Further discussion on this is given in Chapter 7 where the specific empirical results are given more attention as is necessary. Consistent surveys of car ownership over time in set geographical regions are not entirely feasible due to the expense involved with the data collection. The national travel survey provides some scope in overcoming such issues in data collection, however is not geographically stable over time and draws upon data from different sets of individuals. Although such data can be used to define cohorts of individuals over time and examples of this being done with UK data for car ownership include Dargay (2002) or Dargay & Giuliano (2005).

The following general discussion explores some of the more common empirical methods used to model car ownership independently for different data types (in terms of aggregation and dynamic structure).

3.4.1 Pure Car Ownership; Time-Series Models

The term 'time-series' refers to data that includes observations relating to one specific variable that are measured over an extended period of time in a sequential and often equidistant order. Thus time series modelling focuses upon issues not only relating to whether or not one variable has a statistically observable relationship with another variable (e.g. whether income affects the decision to own a car), but also further considers the dynamic relationships that may be observed between groups of variables (e.g. what effect do last year's earnings have upon my decision to own
a car this year?).

It is easy to see that time series modelling provides a level of insight that is simply not attainable through cross-sectional applications, for example asserting the direction of causality (see for example Granger (1969). The introduction of dynamic relationships creates scope for not only understanding the effects observed historically (the term history being used loosely and referring primarily to the past information held within the analysis dataset), but also creating models for forecasting potential future outcomes, see for example Romilly, Song & Liu (2001), a strength that is not easily matched by the relative merits of cross sectional applications. Although in terms of econometric application, new problems are thrown up requiring a more extensive modelling process, such as Granger causality, unit root tests, cointegration, structural change etc, see for example Maddala & Kim (1998).

Time series applications however come with a brand new set of problems, particularly in relation to econometric methods, and the level of testing required to specify a good time-series model can prove to be intensive. The ability to use past periods information in models can provide an easy way out of measuring difficult to observe variables. Consequently empirical understanding can find itself tainted with half-hearted theoretical explanations that are backed up with the use of lagged dependent variables to help capture unobservable elements. However this is subjective, and is a point that can be as much a benefit as a weakness. Such methods sometimes allow the quick and easy specification of good models (in terms of high explanatory power and well specified coefficients), allowing for easy forecasting of future expected values using current data.
Pure Car Ownership; The Logistical Extrapolation Model

The first time-series approach to be considered is the logistics curve model. This is an approach which has received much attention due to the work of Tanner (1978), though applications can be found as far back as de Wolff (1938) and also as recently as Ogut (2004). The process does not rely upon the use of econometric techniques, although does introduce some regression equations to help calibrate the model. This approach is not data intensive and only requires information relating to the growth rate of car ownership, the car ownership rate and also the saturation level of car ownership.

Figure 3.2: Example of a Logistical Curve of Car Ownership
Figure 3.2 provides a graphical illustration of the logistical curve model, identifying the growth relationship offered, and the role of the saturation level. These concepts are more clearly identified and defined below.

The variables used in the analysis are:

- $C_0$: Which is the level of car ownership in the chosen base year (measured as the rate of cars per head of population)
- $g_0$: The growth rate of car ownership, given by $\frac{1}{C} \frac{dC}{dt}$ evaluated at $t = 0$
- $S$: Which is the saturation point of car ownership.

Using the above variables, a logistics curve is fitted, recognising the 's-curve' relationship offered by this type of curve, which closely follows the product life cycle hypothesis. In fact this is a fundamental reason as to why the Logistic curve is used, as it is clear that over time some level of saturation will be achieved. Although the actual level of saturation itself may never be observed, it is certainly reasonable to assume that the growth rate in the ownership of cars will decrease over time as the market moves towards saturation. Ortuzar & Willumsen (1994) provide evidence from the USA and the UK over the approximate period 1953-1988 that support the above argument, and reinforce the use of the logistic extrapolation approach.

The Logistical Extrapolation Model is formalised as follows:

$$\frac{dC}{dT} = aC_t(S - C_t)$$

(3.32)

Where $a$ is a constant, $C_t$ and $S$ are already defined and $T$ represents time. Solving this equation for $C_t$;
\[ C_t = \frac{S}{1 + b \exp(-aSt)} \]  

(3.33)

\( b \) being the integrating constant. Boundary conditions at \( t = 0 \) are used to solve for the constants \( a \) and \( b \) from the above two equations obtaining the following:

\[ g_0 = a(S - C_0) \]  

(3.34)

and

\[ C_0 = \frac{S}{1 + b} \]  

(3.35)

Which can be substituted into equation 3.33 to produce;

\[ C_t = \frac{S}{1 + \left[\frac{(S-C_0)}{C_0}\right] \exp\left[\frac{-aS}{S-C_0}\right]} \]  

(3.36)

This is the final equation used to extrapolate rates for car ownership, requiring the input of base year data for car ownership levels, and the accompanying growth rates. Further to this the only other input required make this model to function is the saturation rate \( S \).

The saturation level is not in itself easily defined, and this is where the modelling process appeals to the standard regression, specifying an equation of the following form;

\[ g = \alpha + \beta C_t \]  

(3.37)

Saturation occurs, by definition, at the point where the rate of growth in car ownership is set equal to zero \( (g = 0) \). The level of car ownership, \( C_t \), is then calculated at this saturation point, and this is the saturation level input into the above equations. So, setting \( g = 0 \);
\[ 0 = \alpha + \beta C_t \]  

(3.38)

Solving for \( C_t \) yields;

\[ C_t = -\frac{\alpha}{\beta} \equiv S \]  

(3.39)

This is the final input required to operate the Logistic Extrapolation Model.

Prior expectations in this model are that the initial level of car ownership will be above zero\(^9\) so the constant should be positive. Furthermore, the slope coefficient, in order to retain the s-shape, is assumed to be negative, so that the growth rate decreases with higher levels of national car ownership per head of population.

This approach implies that the saturation rate/level is the result of an extremely basic relationship between the growth rate in car ownership and the level of car ownership, and may be subject to some specification error, placing the accuracy of this technique in doubt. However, the saturation level is defined by theory, as opposed to a well specified econometric model.

This sets out the initial framework necessary to apply the logistic extrapolation models for time-series data. This model is the subject of some scrutiny, notably from Button et al. (1982), identifying that it is a largely simplified relationship and neglects to emphasise the full range of explanatory variables.

---

\(^9\)With the understanding that the model was primarily intended for national level data and does not use data for periods where cars were not in use (i.e. the 19th century).
Pure Car Ownership: A Discrete Choice Approach for households.

As with the majority of the approaches to modelling transport interactions, the different models of car ownership are generally suited to different types of data, and are not necessarily transferable from one scale of analysis to another (e.g. from household level to a zonal or national aggregate level).

The methods outlined previously in this chapter refer to models which consider interactions over time, and look into more aggregate levels of car ownership, allowing for a choice of time-series methods such as the logistic extrapolation and time-series econometric approaches discussed. These approaches are only of use in the presence of time-series data, and find little application when the available data is only cross-sectional in nature. This requires models that use a different viewpoint to those outlined already, as they will need to capture a fundamentally different relationship.

When looking into car ownership decisions at the household level, it is much less likely that data will be available over any notable time-period. Indeed it is far more realistic to assume the data will be purely cross-sectional.

As already pointed towards in discussing trip models, the data relating to household level decisions shows far less variation and is often better depicted through the construction of discrete variables. For instance, one household is not likely to own ten cars, let alone a hundred, a thousand or a million! It is not realistic to assume that the level of car ownership at the household level can be considered continuous, as in all reality it never could be. Especially if impositions upon individual households, such as budget constraints and physical land constraints are factored in, which when considered together will ensure a definitive cut off point for ownership.
Bearing these issues in mind, household level analysis proceeds to apply discrete choice models, though with some marginally more complex extensions to the models, than those offered in trip rate analysis.

Initial models such as the methods proposed by Quarmby & Bates (1970), work from a standard application of the Logit model, using only two independent variables then, once again, further extending the model to incorporate a saturation level of car ownership. The variables used are the residential density \( D \) (given a predefined area) and the annual family income \( I \).

The following relationships are defined, firstly;

\[
\frac{P_0}{1 - P_0} = a_0 I^{-b_0} D^{c_0} \quad (3.40)
\]

where equation 3.40 is the standard log-odds ratio measured as a function of \( I \) and \( D \). Secondly;

\[
\frac{P_2}{P_1} = a_1 \exp(b_1 I) D^{c_1} \quad (3.41)
\]

which defines a relationship for the log-odds ratio when there are more than two outcomes.

\[
P_0 + P_1 + P_2 = 0 \quad (3.42)
\]

Stating that the sum of the probabilities of the three outcomes defined is equal to one. This model is a multinomial extension of the standard logit, allowing for more than two outcomes in the dependent variable (only three in this example). The parameters to be estimated in this model are \( a_i, b_i \) and \( c_i \).
CHAPTER 3. MODELLING TRAFFIC GENERATION

Rearranging 3.42 for $P_1$, then substituting into 3.41 and taking natural logarithms gives:

$$\log\left[\frac{P_2}{1 - P_0 + P_2}\right] = \log(a_1) + b_1 I - c_1 \log(D)$$  \hspace{1cm} (3.43)

$D$, the resident population per acre, is regarded as discrete in the original model, and consequently further regarded as constant for any given segment. Reducing 3.43 to:

$$\log\left[\frac{P_2}{1 - P_0 + P_2}\right] = b_1 I + \alpha$$  \hspace{1cm} (3.44)

Where $\alpha$ is a constant. This relationship implies that the level of income is directly related to the size of the left hand side of the equation. Assuming the relationship between the two to be positive, as Quarmby & Bates (1970) do, it is possible to express the following:

$$\lim \inf (1 - P_0 - P_2) = 0$$  \hspace{1cm} (3.45)

Which indicates that as the level of income moves to infinity, $(1 - P_0 - P_2)$ tends towards zero. At the limit, this implies that $P_2 \to (1 - P_0)$ i.e. the probability of owning 2 or more cars tends towards 1 minus the probability of not owning a car. This introduces problems when considering the implied relationship at various income levels. At higher levels of income one would assume that the probability of not owning a car is almost zero. This being the case would suggest at high income levels, that everybody owns at least two cars.

Such an assumption does not accurately capture preferences over alternative goods, nor does it accurately capture the notion that some high income households may be lived in by an individual (no partner, children or any other dwellers), in
which case the propensity to own more than one car is greatly reduced.

Due to these reasons the model framework above is ‘tweaked’ to include a saturation level, similar to that applied in the logistic extrapolation method 3.2, thus constraining growth to a maximum value. The value of the saturation value is predetermined in the model, as opposed to calculated from it and fits into the model in the following fashion;

\[
\log \left[ \frac{P_2}{S(1-P_0+P_3)} \right] = \log (a_1) + b_1 I - c_1 \log (D) \tag{3.46}
\]

The value that should be taken for S, as mentioned in the previous paragraph, is predefined, which can be done arbitrarily, or more appropriately empirically. With the advances in computing speed and power over recent decades, arbitrary choice proves to be feasible, allowing the modeller to let the saturation level choose a value that provides the closest fit to the data. This however introduces to it an element of data mining, which is always less preferred to the use of empirical values. Quarmby & Bates (1970) suggest a value of \( S = 0.78 \) in their analysis.

The use of a saturation level in the model is analytically straightforward, as the probability values for \( P_1, P_2 \) and \( P_3 \) are easily defined from the dataset, thus the dependent variable can be easily computed both with and without the imposed saturation levels included (providing for useful comparison).

It is recognised that the final model (equation 3.46), in practice, only requires the level of household income (once the saturation level is known). Although household density is introduced in the model, it is quickly assumed out of the practical application. As a result this model is not data intensive, and may appeal to modellers, particularly in instances where costs are a factor. Ortuzar & Willumsen (1994)
3.4.2 Models of Car Ownership and Use

de Jong (1989) takes initial steps in defining a household level approach to analysing this problem, introducing a simultaneous system, solving jointly for car ownership and vehicle mileage using discrete choice models. In this research, de Jong comes to the conclusion that the key variable in defining vehicle use simultaneously with car ownership is indeed vehicle mileage, and not an alternative measure for car use, such as frequency.

These models of joint car ownership and use are generally of a structural (or simultaneous equations) nature, where disturbances across the various equations (generally only two in this situation) are allowed to share some correlation with each other. This imbues a statistical relationship that requires a more complex analytical approach, leading to complex interpretation of coefficients within the system. Results arising from this approach may vary quite substantially when compared to more simple single equation estimations.

The inclusion of vehicle use is a natural extension of the car ownership model requiring little explanation. The relationship between ownership and use may help to provide some insights into the reliance upon cars within households, especially if access to alternative modes of transport can be factored into analysis. Furthermore, as Romilly et al. (2001) detail, national models of car ownership and use are of great importance to bodies such as the Department of Environment Transport and the Regions (DETR) (now under the Department for Transport (DfT)) who wish to forecast future levels of car ownership and use to help provide some bearing on
the level of maintenance and investment required by the government on the road network over forthcoming years.

Household models of joint car ownership and use

de Jong (1989) has produced some of the most comprehensive literature on household levels of car ownership. In this research de Jong outlines the reasons behind a more comprehensive analysis of car ownership and use as being a fundamental belief that they are simultaneous in nature at the individual level, and also that there is a need to better understand how one feeds information into the other. Two approaches are broadly identified in this research (with what is regarded as one of the more comprehensive reviews of general car ownership models (Ortuzar and Willumsen 1994, pp345)), which are the censored regression model and the indirect utility model. Each of these models is presented below, almost in parrot fashion to de Jong’s original dissertation, due to the high difficulty involved with gaining access to the information.

The censored regression model

This model, which is an extension of the standard Tobit model, assumes inter alia that the decision to own a car depends on the mileage of a household exceeding a certain threshold level. It takes the form:

\[ y^*_i = x^i \beta + v_i \]

\[ y_i = x^i \beta + v_i \text{ if } y^*_i > \gamma \]
de Jong (1989) makes a change to the model of Tobin (1958) and sets the threshold level (gamma) above zero such that;

\[ \gamma > 0 \]  \hspace{1cm} (3.49)

Equation (3.49) states that household mileage must exceed a threshold level in order to induce car ownership, a level of use below this threshold is not otherwise observed.

The Tobit model seeks not only to understand the observable information, but also to use assumptions surrounding the distribution of the whole data (observed and unobserved) to attempt to conduct inference on the unobservable sections of the data. This assumption is generally accepted in the application of this model, and techniques to extend the information provided by this model using a standard normal distribution are common.

de Jong (1989) extends the Tobit model (assuming a normal distribution) for the application of modelling car ownership and use. \( y^*_i \) (from above) is defined as permanent car use and is subject to a disturbance term with the following properties;

\[ \nu_i \text{ i.i.d. } N(0, \sigma^2_v) \]  \hspace{1cm} (3.50)

Which captures the unobservable effects not captured by the other variables in the model, such as the random nature of human behaviour, tastes etc.

The goal now is to define actual car use \( y_i \) and rewrite it as;

\[ y_i = y^*_i + w_i \]  \hspace{1cm} (3.51)

Also with a normally distributed disturbance term;
\[ w_i \ i.i.d. \ N(0, \sigma_w^2) \quad (3.52) \]

Which by definition is independent of \( v_i \). The relationship in equation 3.51 therefore states that the actual car use within the population is a combination of the observed car use (which incorporates its own error term) and some white noise process. From 3.47, 3.48 and 3.52 the following can be written;

\[ u_i = v'_i + w_i \quad (3.53) \]

Where \( v'_i \) is the truncated version of \( u_i \).

\[ v'_i \ i.i.d. \ N(0, \sigma_v^2) \text{ if } v_i > \gamma - x' \beta_i \quad (3.54) \]

With \( u_i \) not observed otherwise.

The term \( u_i \) is subject to a distribution formed using a combination of a truncated normal and a standard normal distribution. de Jong (1989) identifies the appropriate estimate of density as that offered by Stevenson (1984), which subsequently allows the models to be estimated using maximum likelihood methods.

The indirect utility model;

de Jong (1989) also discusses an alternative approach based upon the tenets of an indirect utility model which is expressed in general terms as thus;

\[ \max U = U(A, X) \quad (3.55) \]

With the budget constraints;

\[ y \geq vA + C + X \text{ if } A \geq 0 \quad (3.56) \]
\[ y \geq X \quad \text{if} \quad A = 0 \quad (3.57) \]

Where \( Y \) is the household's income, \( X \) is the expenditure on all other goods and \( A \) is automobile use, \( v \) is the variable cost of car use and \( C \) is the fixed cost of car ownership. It is assumed that the budget set is non-convex due to the presence of fixed costs.

de Jong (1989) draws upon the Tobit model specified above as a basis for a demand framework which with the aid of Roy's Identity (in a flexible form) is then used to derive the demand for automobile use using fixed costs and variable costs. The shortened derivation is as follows;

\[ \ln A =\alpha \ln(Y - C) - \beta v + \delta \quad (3.58) \]

Where \( \delta \) captures household socio-economic and demographic characteristics. Alpha and beta are both expected to be greater than zero. In order to apply Roy's Identity, income is aggregated up as not to take individuals separately. The indirect utility function associated with equation 3.58 is;

\[ \psi(v, Y - C) = \frac{1}{\beta}(\delta - \beta v) + \frac{1}{1 - \alpha}(Y - C)^{1-\alpha} \quad (3.59) \]

As the (variable) cost of motoring increases, the indirect utility tends towards the utility of not owning a car.

\[ \lim \psi(v, \psi) = U0, Y \quad (3.60) \]

Which then suggests that the associated direct utility function when there is no car use (\( A = 0 \)) is (recalling the a priori expectation that beta is greater than zero);
\[ U(0, Y) = \frac{1}{1 - \alpha} \gamma^{1 - \alpha} \] (3.61)

Car ownership occurs when \( \psi(v, Y - C) > U(0, Y) \). Expressions for \( v_{\text{max}} \) and \( A_{\text{min}} \) are found to be:

\[ v_{\text{max}} = \frac{-1}{\beta} \ln \beta + \frac{\delta}{\beta} + \frac{1}{\beta} \ln(1 - \alpha) - \frac{1}{\beta} [Y^{1 - \alpha} - (Y - C)^{1 - \alpha}] \] (3.62)

\[ \ln A_{\text{min}} = \alpha \ln(Y - C) - \ln \beta - \ln(1 - \alpha) + \ln(Y^{1 - \alpha} - (Y - C)^{1 - \alpha}) \] (3.63)

In order to make this a stochastic specification the following definitions must be made:

\[ \delta_i = \gamma S_i + e_i \] (3.64)

Where subscript \( 'i' \) denotes the individual household. \( e_i \) is the disturbance term and is \( i.i.d. \) normal with zero mean and \( \sigma_e^2 \) standard error. \( S_i \) is a vector of observed socio-economic and demographic characteristics. To add some intuition to the definition, the distinction between 'intended car use' \( K_i^* \), and 'actual car use' \( K_i \), is made:

\[ K_i^* = \ln A_i \] (3.65)

\[ K_i = K_i + w_i \] (3.66)

\[ w_i \ i.i.d. \ N(0, \sigma_w^2) \] (3.67)

This specifies car use in a stochastic framework, allowing for actual car use to be a function of intended car use and a random disturbance term \( w_i \). This disturbance term accounts for random human behaviour and unexpected 'transitory
disturbances that influence car use after the car ownership decision has been made' de Jong (1989). The model can be estimated using maximum likelihood methods.

When comparing the two approaches offered by de Jong (1989) the indirect utility model is preferred as it offers more information and is based upon a firm theoretical utility based model.

**Aggregate models of Car Ownership and Use**

Romilly et al. (2001) define a series of equations estimating car ownership and use jointly using time series data from the UK over the period 1953-1996, producing forecasts as far forward as 2031 (the same length as DfT forecasts). The basic relationship proposed is;

\[
\begin{align*}
CO & = f(Y, P, B, AGE, CGN, HOU, RL, TBR, U) \\
CU & = f(Y, P, B, AGE, CGN, HOU, RL, TBR, U)
\end{align*}
\]

(3.68)

Where the independent variables are real disposable income \(Y\), real motoring costs \(P\), Bus fares \(B\), age structure \(AGE\), Congestion \(CGN\), number of households \(HOU\), road length \(RL\), the real interest rate \(TBR\) and unemployment \(U\). The dependent variables are car ownership \(CO\) and car use \(CU\). All information is measured in per capita terms derived from national aggregate data. Stochastic multiplicative demand functions of the following form are specified and estimated;

\[
\begin{align*}
CO & = AY^{\alpha_1} P^{\alpha_2} B^{\alpha_3} HOU^{\alpha_4} RL^{\alpha_5} U^{\alpha_6} \exp(\alpha_7 TBR)u_1 \\
CU & = BY^{\beta_1} P^{\beta_2} B^{\beta_3} HOU^{\beta_4} RL^{\beta_5} U^{\beta_6} \exp(\beta_7 TBR)u_2
\end{align*}
\]

(3.69)

Though in the empirical implementation some of the variables were dropped (mainly due to data problems of non-stationarity), 'A' and 'B' are constants on \(Y\). This framework is operationalised by taking logarithms and hence making the model linear in parameters. A number of approaches are considered in the estimation of
this system including a cointegration approach and general to specific modelling on Engel Granger 2 Stage Least Squares (EG2SLS), ARDL models and Johansen Maximum likelihood estimators among others.

The main results of the work by (Romilly et al. 2001) are that firstly, congestion plays a significant role in the car ownership and use decisions. A time trend is introduced into the model to proxy for congestion (on the assumption that the level of congestion is increasing over time at a steady rate), and is found to have a significant and negative impact on car ownership and use, which follows intuition. The assumptions surrounding congestion should however be considered restrictive for two reasons. Firstly it is assumed in this study that growth in congestion is linear, which may not necessarily be true, further this possibility is not tested for. Secondly and complementary to the first point, congestion is constrained by spatial limits, congestion can only increase up to the point where all road space is used. Although in practice this is unlikely to ever occur, it does suggest that the assumptions underlying this model are not entirely realistic.

As would be expected a-priori, the variables for income, motoring costs and bus fares are all found to be significant and take the expected signs.

3.4.3 Summary of Car Ownership Modelling Approaches

This section provides empirically grounded support that the decision to own and use cars falls neatly into a standard demand framework, where determinants affecting personal income levels, own price costs (of the primary good under analysis), substitute good costs and dynamic utility effects (via the change in congestion) are all factored in.
With respect to some form of appraisal of the techniques discussed, the structure of the analytical database must be borne in mind. Given this precursor, the logistical extrapolation model can be quickly rejected, as the desire to forecast future traffic levels provides no useful bearing on either the research remit, nor the ability to model car ownership or trip generations within a static framework. Furthermore, the relatively limited information requirements mean that rather little information on socioeconomic or many other indicators can really be extracted. Moreover, the data precludes the ability to usefully specify any time series models.

The discrete choice approach establishes a more formally grounded theoretical model prior to estimation stages, and thus can be advantageous over other approaches. Although the relatively strong assumptions it imposes on distributional form is a less desirable property, and could be potentially very restrictive. As will be outlined in the following chapter, due to reasons of data structure inter alia, allows for an above average level of flexibility in the estimation process. Analogous techniques for discrete choice models are far less well understood, and the subsequent imposition of a discrete choice form for one part of a car ownership and use model but not for the other would detract from the benefits of the econometric method. It is therefore decided that a standard cross section regression technique, similar to the time series approaches, but with the removal of the element of time, will be adopted. This will allow for consistency of modelling technology across the two stages of the model and also for the full benefits of the estimation technique to be realised.
3.5 Overall Summary of Modelling Chapter

This section brings together both strands of modelling discussed within this chapter, namely trip generation and car ownership. Figure 3.3 shows the unique contribution offered by the TRICS dataset in terms of modelling strategy and this summary section should be mindful of exactly where the data does fit in.\(^\text{10}\) In general, transport applications take one of two approaches: Problems are either analysed from the point of view of individuals on the understanding that information can then be aggregated up to provide inference at regional or national levels. Alternatively aggregate models are offered which explain fundamental relationships with no attempt to offer insights into the more complex behaviour of individuals, merely the average individual.

Models using multi-level data are also discussed in the literature (see for example Ettema & Timmermans (1997)), although empirical research using data of the form presented in the TRICS database is not encountered. The TRICS data offers traffic count and Site specific characteristics from around the country. Individual sites are considered (a form of micro level data), supplemented with situational characteristics of the surrounding area (data at a higher level of aggregation). The final analysis datasets will consider the characteristics of individuals in an area, but is unable to differentiate between individual households. On the other hand, the dataset considers only individual sites and their interactions with the wider transport network (with unknown origins, or unknown destinations in the analysis of households), thus not considering all similar sites within a region, or across the nation, consequently this is not an aggregate level analysis.

This research will produce empirical results at a level of analysis which sits in-

\(^{10}\)However the specific details of the data itself are present in the following Chapter.
between individual level models and aggregate level models, helping to guide policy at the regional level, based upon data inputs from around the entire UK. These results will be based upon multi-level data.

Given the unique form of the TRICS data, and the information collected together within this chapter linear regression techniques as summarised in section 3.2.3 are the only feasible approach to model the trip generations. As is summarised in section 3.4.3 the time series approaches are not feasible, and the discrete choice methods are not really suitable given the data structure. Further given the desire to estimate a joint model of car ownership and use, the specification of a standard linear regression equation relating absolute car ownership at a site to its characteristics seems the most logical approach.
Chapter 4

Data and Methodology

This chapter presents the data used for the three analytical Chapters within this thesis, then based upon the data and the preceding discussion of trip generation modelling, outlines the general model by which the three empirical chapters are broadly based. The chapter therefore discusses the TRICS database as the general source of data, then provides a more focussed analysis upon the data for each individual chapter i.e. Offices, Food Superstores then Residential developments.

4.1 The Data

This section of the chapter will provide both general information regarding the database, as well as specific information on the data used within this thesis and will therefore cover the following;

1. The TRICS database (as a whole)

2. The Office developments sub-sample of data

3. The Food Superstores sub-sample of data

4. The Residential developments sub-sample of data
In so doing it will offer general comments on how the TRICS consortium evolved, and the work they have done over the years. For the site specific data sections, a brief descriptive analysis of the data will be presented analysing the average flows of trips across an average day for each site type.

4.1.1 The TRICS database

The TRICS database collates a large selection of trip rate surveys for a wide range of different developments, in fact several thousand surveys across over 100 different land use classifications. As set out in the introduction, the database was initiated by a consortium of six county councils from the Southern regions of England in the early 1980’s. The purpose of the database was to record baseline information on traffic impact levels of alternative site types that may be used to help guide developers in assessing the transport requirements of any new development.\(^1\) Given the policy discussion offered in Chapter 2, the data collection process was constrained to car borne traffic only. In line with the growth in demand for alternative transport modes and gains in knowledge by the transport profession over recent years, the data collection has evolved to become multi-modal. However it is prudent to identify that this is only a relatively recent change to the data collection process, and as such has not yet provided enough meaningful data to conduct a formal, activity specific, trip rate analysis.

The data collection process involves cordon counts of traffic emanating from every entry and exit point at a site, on the understanding that these can be clearly defined. As such, each site survey captures all motion that can be easily defined.

\(^1\)Furthermore, the TRICS database has the advantage of being nationally recognised at government levels, and is also the Department or Transport’s standard data collection tool for ‘travel plan’ assessment.
as owing to that particular site. There are issues that these types of surveys do not accurately reflect travel behaviour, partly because they can give no accurate indication of journey lengths, or if trips are random ‘pass-by’ trips. These are valid concerns, however given the analysis process and aggregation over multiple surveys, it is considered that such conjectures, although difficult to comment upon, will be systematically featured into the stochastic error term. Thus the bias imparted by ‘out-of-sequence’ behaviour will be minimised. Another issue which arises is that of parking, where it is frequently commented that surveys do not accurately record parking, as off site parking is hard to measure but may be a significant feature in the decision to travel to a site by car. The empirical phase will provide some bearing of how prominent such effects could potentially be, by revealing the amount of variation that is attributable to exogenous factors through the error. This will be returned to in the conclusions in Chapter 8. One clear strength of the data source is the level consistency ensured in the collection process over the life of the database, in part due to the stable oversight of a corporate entity, namely JMP consultants.

The traffic count data is typically taken from established sites (i.e. not new developments), it is therefore assumed that the general customer base has leveled out after any initial opening ‘boom’, where the customer levels are essentially stable. Furthermore, given the model features no dynamic elements\(^2\), there is no distinction between long-run and short-run effects. It is not difficult to formulate hypotheses to suggest that initial opening levels could be (and are likely to be) higher OR lower than the estimates produced by the static model. Although initial opening levels should, on average, converge towards the model estimate over time. The estimated model does not differentiate between trips to the store from linked (or chained) trips

\(^2\)Effects of the day of week are captured in the model, though the form of the data means that this is not dynamic. These ‘weekday’ variables are strictly qualitative.
or single purpose journeys. The framework therefore focuses on the fundamentals of the determinants of trip 'attractions' to a particular site.

In their initial conception the consortium recognised the potential effect that different land-uses, sub land-uses and location types may have upon trip rates. However, prior to this thesis, the information has been little used to identify if there are in fact any significant differences in terms of traffic impact. As it has been recorded from the initial days of the database, every survey contains such information and therefore provides substantial data for estimation purposes. Supplementary to this, the survey process also includes the gathering of site specific characteristic information, relative to the site in question (e.g. beds at a hotel, or holes on a golf course) as well as a certain amount of information pertaining to the socio-economic characteristics of the surrounding area. Indeed the survey process provides a wealth of information that is simply not matched by any other data source in such a consistent fashion.

The information extracted from the raw form of the TRICS database are as shown in the following list. There is a broad range of data capture incorporating data on the site, the survey details, local area characteristics, as well as a wealth of qualitative information. This (extensive) provides a broad indication of the depth of data which is being collated through the TRICS database. A great many of the variables identified in the following list however, have at present (TRICS build 2004b) insufficient, or in certain cases no data available.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDAREACODE</td>
<td>Two letter area code identifier</td>
</tr>
<tr>
<td>DDLANDUSE</td>
<td>Main land use type</td>
</tr>
</tbody>
</table>
### CHAPTER 4. DATA AND METHODOLOGY

<table>
<thead>
<tr>
<th><strong>Code</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DDSUBUSE</td>
<td>Sub use type within DDLANDUSE</td>
</tr>
<tr>
<td>DDSITENO</td>
<td>Site number</td>
</tr>
<tr>
<td>DDDAYNO</td>
<td>Day of week</td>
</tr>
<tr>
<td>DDSURVDAY</td>
<td>Survey date (of month)</td>
</tr>
<tr>
<td>DDSURVMONT</td>
<td>Month of survey</td>
</tr>
<tr>
<td>DDSURVYEAR</td>
<td>Year of survey</td>
</tr>
<tr>
<td>DDSURVTYPE</td>
<td>Type of survey; unknown, manual or ATC</td>
</tr>
<tr>
<td>DDINCOCCUP</td>
<td>occupancy of the car park at the start of the survey</td>
</tr>
<tr>
<td>DDFINOCCUP</td>
<td>occupancy of the car park at the end of the survey</td>
</tr>
<tr>
<td>ddarr01-ddarr48</td>
<td>Trip arrivals per time period, broken down</td>
</tr>
<tr>
<td>dddep01-dddep48</td>
<td>Trip departures per time period, broken down</td>
</tr>
<tr>
<td>DDPROPMCS</td>
<td>Proportion of trips which are motorcycles</td>
</tr>
<tr>
<td>DDPROPCAR</td>
<td>Proportion of trips which are cars</td>
</tr>
<tr>
<td>DDPROPLGV</td>
<td>proportion of trips which are light goods vehicles</td>
</tr>
<tr>
<td>DDPROPGV1</td>
<td>proportion of trips which are other goods vehicles classification 1</td>
</tr>
<tr>
<td>DDPROPGV2</td>
<td>proportion of trips which are other goods vehicles classification 2</td>
</tr>
<tr>
<td>DDPROPPSV</td>
<td>proportion of trips which are public service vehicles</td>
</tr>
<tr>
<td>DDPROPTOT</td>
<td>Total proportion of trips</td>
</tr>
<tr>
<td>DDATC</td>
<td>identifies if count was automated or not.</td>
</tr>
<tr>
<td>DDCMTFLAG</td>
<td>Comments for survey day</td>
</tr>
<tr>
<td>DDIRECTYPE</td>
<td>Type of count; multi - modes</td>
</tr>
<tr>
<td>DDAMWEATH</td>
<td>Morning weather</td>
</tr>
<tr>
<td>DDPMWEATH</td>
<td>Afternoon weather</td>
</tr>
</tbody>
</table>
CHAPTER 4. DATA AND METHODOLOGY

DCcomment Comments relating to trip observations
Eccomment General comments on site facilities
EDTRADENAM Trade Name
EDSITEAREA Site area (hectares)
EDGFA Gross Floor area (sqm)
EDRFA Retail Floor area (sqm)
EDEMPLOYEE number of employees at the site
EDYEAROPEN Year the site opened
EDUNITS number of units the site has, e.g. beds in a hotel.
EDMTOPEN Mon-Thurs opening time
EDMTCLOSE Mon-Thurs closing time
EDPARKSPAC Total Parking provision
EDPARKCHAR Charge for parking
EDSURFPARK Surface parking provision
EDDISTTONE Distance to nearest similar competitor
EDPKVISCUS Parking provision for Visitors/customers
EDPKEMPLOY Parking provision for employees
EDPKDISABL Parking provision for disabled
EDPKCYCLE Parking provision for cycles
EDPKHGVLOA Parking provision for HGV’s loading
EDPKHGVPAR Parking provision for HGV’s parking
EDPKMATERN Parking provision for ‘parent and child’
EDOFFPARK Off-site parking provision
SCcomment General site comments
SDREGION Region code (see TRREGION)
SDDESCRIP Brief site description
SDSTREET Street which site is situated on
CHAPTER 4. DATA AND METHODOLOGY

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDDISTRICT</td>
<td>District</td>
</tr>
<tr>
<td>SDTOWN</td>
<td>Town</td>
</tr>
<tr>
<td>SDPOSTCODE</td>
<td>Post code</td>
</tr>
<tr>
<td>SDGRIDREF</td>
<td>Grid reference</td>
</tr>
<tr>
<td>SDLOCATION</td>
<td>Location category</td>
</tr>
<tr>
<td>SDPOP1M</td>
<td>Population within 1 mile (coded)</td>
</tr>
<tr>
<td>SDPOP5M</td>
<td>Population within 5 miles (coded)</td>
</tr>
<tr>
<td>SDCAR5M</td>
<td>Car ownership within 5 miles (coded)</td>
</tr>
<tr>
<td>SDPUBTRAN</td>
<td>Public transport provision to site (coded)</td>
</tr>
<tr>
<td>SDUSECLASS</td>
<td>Use class category</td>
</tr>
<tr>
<td>SDDEVCOUNT</td>
<td>Number of developments</td>
</tr>
<tr>
<td>SDDAYCOUNT</td>
<td>Number of survey days</td>
</tr>
<tr>
<td>SDPTEMPL</td>
<td>Part time employment at site</td>
</tr>
<tr>
<td>SDWALKCATC</td>
<td>Walk in catchment (500m)</td>
</tr>
<tr>
<td>SDRECTYPE</td>
<td>Multi-modal survey flag</td>
</tr>
<tr>
<td>SDSURVOLD</td>
<td>Resurvey history</td>
</tr>
<tr>
<td>SDSURVNEW</td>
<td>Resurvey history</td>
</tr>
</tbody>
</table>
The land uses definitions used in TRICS are in accordance with PPG13 (1992) and are defined as follows:

**Figure 4.1: TRICS location definitions**

**Town Centre** Within the central core area of the heart of the town e.g. the primary shopping area, as defined in the local development plan.

**Edge of Town Centre** For retail, a location within easy walking distance (i.e. up to 300 metres) of the primary shopping area, often providing parking facilities.
that serve the centre as well as the site, thus enabling one trip to serve several purposes. For other uses, edge-of-centre may be more extensive, but within 300m of the town centre boundary, based on how far people would be prepared to walk. For offices this may be outside the town centre but in the urban area within 500m of a public transport interchange. Local topography and barriers will affect pedestrians perceptions of easy walking distance. Examples of barriers include crossing major roads and car parks. The perceived safety of the route and strength of the attraction of the town centre are also relevant.

**Neighbourhood Centre (Local Centre)** Residential area, similar to "Suburban Area", but with additional amenities like local shops, schools, etc. Could be described as a small "district" or "village" within the town itself. Would also apply to actual villages. The Local centres include a range of small shops of a local nature, serving a small catchment. These may include a general grocery store, a newsagent, a sub-post office and a pharmacy. These centres provide accessible shopping for peoples day-to-day needs.

**Suburban Area (Out of Centre)** A Residential area that is outside the town centre, but not at the physical edge of the town itself. Villages are included as Neighbourhood Centre.

**Edge of Town** At the physical edge of the town/city, where the town meets the countryside.

**Free Standing (Out of Town)** Out of town, beyond the physical edge of the nearest town/city, in the countryside.

**Commercial Zone** An area of significant business activity within a town.

**Industrial Zone** An area of significant industrial activity within a town.
Development Zone An area of redevelopment or regeneration, for example London Docklands (or on a smaller scale for other towns and cities).

To enhance the TRICS database, including this information on land zone placement, it is augmented with data from the government sources, namely the National Online Manpower Information System (NOMIS) archive and the 2001 Census of Population, thus culminating in a rich database. The NOMIS archive provides information pertaining to labour market statistics for local and national areas, combining data from the Labour Force Survey (LFS), Claimant Count, Annual Business Inquiry (ABI), New Earnings Survey (NES) and the 1991 UK Census of Population with information from 1970 through to the present date, although only data from 1986 (onwards) is extracted for the purpose of this study.

The 2001 UK Census of Population provides further data on socio-demographic and economic characteristics for households, which is consistent with the data extracted from the NOMIS database. The data collection processes used in the 1991 and 2001 Censuses of Population are not perfectly consistent, resulting in a need to aggregate data together to create consistent variables, however no discernable difficulties are encountered in doing so. Variables from the Census (and NOMIS) data sources are considered as strong proxies (as opposed to more precise or exact values) due to the irregularity at which the data is collected.

Three subsequent datasets have been constructed (i.e. for Offices, Food superstores and Residential developments) so as to maximise the information used in analysis. A number of the variables have either only recently entered into the database, consequently containing limited information on only the most recent surveys, or have simply been poorly recorded over time. The result is that a number of variables must be dropped from statistical assessment. Furthermore, a great num-
ber of variables contain information which is either too rich or too inconsistent to facilitate a large scale statistical analysis, for instance some comments fields provide in depth information on bus linking services in the town. The impact of such information on an aggregated statistical assessment is likely to be small, further such information is offered for the distinct minority of surveys. The remainder of the first section of this chapter describes each of these three datasets in more detail prior to describing the general modelling approach.

Variable names description

The following list denotes the variable names used for the data section and empirical analyses that follow in the subsequent chapters;

\[ \text{ln} = \text{Natural logarithm.} \]

\[ \text{ARR} = \text{The average hourly flow of passenger vehicle arrivals to a site.} \]

\[ \text{T} = \text{The average hourly departures from a site.} \]

\[ \text{GFA} = \text{The Gross Floor Area of a site.} \]

\[ \text{RFA} = \text{Retail Floor Area of a site.} \]

\[ \text{SEMP} = \text{Employment at the site.} \]

\[ \text{PARK} = \text{On site parking.} \]

\[ \text{CAR} = \text{Car ownership in the area of the site.} \]

\[ \eta = \text{Car ownership of the site (residential developments only).} \]

\[ \text{PSV} = \text{Public service provision at the site.} \]

\[ \text{AVHS} = \text{Average household size in the area surrounding the site.} \]
AVLH = Proportion of 'large' households in the area surrounding the site.

AVHC = Proportion of households with children in the area of the site.

AEMP = Employment levels in the area of the site.

APOP = Population in the surrounding area of the site.

HHEMP = Average household employment in the area of the site.

MSA = Metropolitan area size.

YEAR = Year of the survey for the site.

LU1 = Town centre

LU2 = Neighbourhood centre

LU3 = Suburban area

LU4 = Edge of town

LU5 = Free standing

LU6 = Commercial zone

LU7 = Industrial zone

LU8 = Development zone

LU9 = Edge of town centre

PFS = Indicates the presence of a petrol filling station at the site.

HTA = Houses privately owned

HTB = Houses for rent
CHAPTER 4. DATA AND METHODOLOGY

HTC = Flats privately owned
HTD = Flats for rent
HTE = Institutional hostels
HTF = Sheltered accommodation
HTG = Student accommodation
HTH = Nurses homes
HTI = Caravan parks
HTJ = Holiday accommodation
HTK = Mixed private housing
HTL = Mixed non-private housing
HTM = Mixed private/non-private housing
HTN = Retirement flats

RESISTANCE = Proxy for competition.

ECONOMY = Indicates if a food superstore is 'less mainstream'.

MON-THURS = Day identifier.
SAT = Day identifier.
SUN = Day identifier.

Not all of the variables feature within each empirical model as some are specific to land use types and/or capture behavioural information specific to a certain site type. The rationale behind the inclusion or omission of certain variables from particular
empirical models is outlined in the following. The following three subsections will therefore outline the key characteristics of the datasets incorporated within the empirical chapters. In order to avoid confusion, only those variables which feature prominently within the statistical analysis are included in the following descriptions. For instance, day of week characteristics were only found to be significant in the analysis of food superstores, and hence for ease of clarity, day of week details are not recorded for offices or residential developments.
4.1.2 Office developments data

The composite dataset provides a cross-section of information for 50 office block developments over the period 1987-2002. This sample size is smaller than the other two site types considered due to the commercial nature of the database and the needs of developers and local authorities. The overall national population of office block developments will contain more developments than this. Given the method in which data is collated for the TRICS database, the dataset is confined to include only 1 observation per site, where complete information is available for all independent variables considered. Descriptive statistics of the full dataset (Prior to imputation) are provided in Table 4.1.

Figure 4.2 provides the mean average of the total trip arrivals and departures over the course of the average 24 hour period. The majority of the surveys are conducted around the normal working hours periods (i.e. based around 9am-5pm), however a number of the sites are known to be in operation 24 hours a day. This figure reveals a number of points, firstly there is evidence of bi-modal peaking for both the arrivals and departures, which are situated around the typical start and

---

3Given the depth of the data and spread of observations over time, the possibility of applying time series and/or panel methods can not be considered appropriate at present. This is largely because the individual data series will be heavily unbalanced, and therefore would require a significant amount of augmentation/manipulation in order to be overcome. More data, as will be collected over coming years, will however make these estimation technologies feasible in the future.

4Thus the dataset has been partially imputed already, so as to constrain analysis to missingness in the dependent variable only. As such the policy implications of these results should not be considered complete, though certainly indicative of the overall population of office developments over the period of analysis.

5All observations are recorded 50 times with the exception of \( \text{lnArr} \), which only enters 35 times in the unimputed dataset.

6Using the raw data and not the natural logarithm transformations.
Table 4.1: Descriptive Statistics for Office Data

<table>
<thead>
<tr>
<th>Continuous Variables</th>
<th>Mean</th>
<th>Median</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnArr</td>
<td>5.550</td>
<td>5.727</td>
<td>1.055</td>
<td>2.734</td>
<td>6.993</td>
<td>35</td>
</tr>
<tr>
<td>lnGFA</td>
<td>8.638</td>
<td>8.627</td>
<td>1.087</td>
<td>6.011</td>
<td>11.054</td>
<td>50</td>
</tr>
<tr>
<td>lnSEMP</td>
<td>5.746</td>
<td>5.851</td>
<td>1.050</td>
<td>3.611</td>
<td>7.937</td>
<td>50</td>
</tr>
<tr>
<td>lnPARK</td>
<td>5.135</td>
<td>5.244</td>
<td>1.220</td>
<td>2.303</td>
<td>7.601</td>
<td>50</td>
</tr>
<tr>
<td>lnCAR</td>
<td>-0.176</td>
<td>-0.223</td>
<td>0.372</td>
<td>-1.386</td>
<td>0.262</td>
<td>50</td>
</tr>
<tr>
<td>lnPSV</td>
<td>3.738</td>
<td>4.248</td>
<td>1.214</td>
<td>0</td>
<td>4.248</td>
<td>50</td>
</tr>
<tr>
<td>lAVHS</td>
<td>0.904</td>
<td>0.904</td>
<td>0.055</td>
<td>0.792</td>
<td>0.958</td>
<td>50</td>
</tr>
<tr>
<td>lnAEMP</td>
<td>11.113</td>
<td>10.998</td>
<td>0.603</td>
<td>9.977</td>
<td>12.282</td>
<td>50</td>
</tr>
<tr>
<td>lnAPOP</td>
<td>11.927</td>
<td>11.800</td>
<td>0.578</td>
<td>10.999</td>
<td>12.987</td>
<td>50</td>
</tr>
<tr>
<td>lnHHEMP</td>
<td>0.091</td>
<td>0.088</td>
<td>0.130</td>
<td>-0.154</td>
<td>0.431</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummy Variables</th>
<th>Mean</th>
<th>Median</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>LU1(base)</td>
<td>0.18</td>
<td>0</td>
<td>0.388</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU2</td>
<td>0.04</td>
<td>0</td>
<td>0.198</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU3</td>
<td>0.12</td>
<td>0</td>
<td>0.328</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU4</td>
<td>0.26</td>
<td>0</td>
<td>0.443</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU5</td>
<td>0.1</td>
<td>0</td>
<td>0.303</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU6</td>
<td>0.08</td>
<td>0</td>
<td>0.274</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU7</td>
<td>0.06</td>
<td>0</td>
<td>0.240</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU8</td>
<td>0.06</td>
<td>0</td>
<td>0.240</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LU9</td>
<td>0.1</td>
<td>0</td>
<td>0.303</td>
<td>0</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>
end of the work and lunch periods. The start of work peak in average arrivals occurring at 09:00, though spread from 06:30-10:00. Similarly the end of work peak in average departures is at 17:30, with an observably more concentrated spread of trip generations starting from around 16:30 and finishing at 18:30. The lunchtime network traffic loading occurs between 12:00 and 14:00, with the peak sizes suggesting that there is a significant proportion of employees who do not return to work after lunch.

![Mean trip generations for sample of office developments](image)

Figure 4.2: Total car traffic arrivals by time of day - Offices

Reasons for this more diffused peak after the lunchtime period may include the end of shift, the need to visit another office in the afternoon and/or the decision to
take a half day or continue working the rest of the day from home (teleworking). The implications of these different reasons in terms of their contribution to non-point pollution concerns and other negative externalities are diverse. For instance, taking a half day and/or teleworking for half a day still necessitates at least two journeys (from home to work and the associated return journey). In the instance of a half day, this means almost the same amount of non-point pollution is generated as per the employees working whole days, with only half the productivity for the firm. A slight reduction in externalities might arise as a result of one end of the return journey being made during the off peak period, thus providing less network friction (congestion etc.). Teleworking overcomes the productivity issue by allowing employee's to continue working from home, and is arguably beneficial as the non-point pollution will be marginally lower relative to those who work a full-day in the office and subsequently travel almost exclusively during peak periods.

The observed total number of arrivals exceeds the total number of departures. This phenomenon is a result of the 24 hour nature of some of the sites, where the shifts of some individuals will begin halfway through a traffic survey, and the associated departure will not occur until after the survey is finished. Given then, that most of the surveys are conducted around normal working hours and hence, include peak travel hours. The implication is that the inception of 24 hour operation at office developments would appear to be associated with lower peak traffic loads. Thus contributing less to non-point pollution effects by avoiding congesting the network further and not causing increased waiting times for all individuals using the roads.
### 4.1.3 Food Superstore Data

#### Table 4.2: Descriptive statistics for Food superstore data

<table>
<thead>
<tr>
<th>Continuous variables</th>
<th>Number</th>
<th>mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnCAR</td>
<td>201</td>
<td>-0.136</td>
<td>0.32</td>
<td>-1.386</td>
<td>0.262</td>
</tr>
<tr>
<td>lnPSV</td>
<td>201</td>
<td>3.882</td>
<td>0.633</td>
<td>2.303</td>
<td>4.5</td>
</tr>
<tr>
<td>lnGFA</td>
<td>201</td>
<td>8.565</td>
<td>0.4</td>
<td>7.097</td>
<td>9.218</td>
</tr>
<tr>
<td>lnRFA</td>
<td>199</td>
<td>8.006</td>
<td>0.385</td>
<td>6.783</td>
<td>8.817</td>
</tr>
<tr>
<td>lnRESISTANCE</td>
<td>201</td>
<td>1.12</td>
<td>0.866</td>
<td>-1.386</td>
<td>3.367</td>
</tr>
<tr>
<td>lnAVHS</td>
<td>201</td>
<td>0.861</td>
<td>0.527</td>
<td>0.759</td>
<td>0.977</td>
</tr>
<tr>
<td>lnAVLH</td>
<td>201</td>
<td>-1.109</td>
<td>0.137</td>
<td>-1.406</td>
<td>0.846</td>
</tr>
<tr>
<td>lnAVHC</td>
<td>201</td>
<td>-1.322</td>
<td>0.142</td>
<td>-1.639</td>
<td>-1.003</td>
</tr>
<tr>
<td>lnHHHEMP</td>
<td>201</td>
<td>0.19</td>
<td>0.205</td>
<td>-0.185</td>
<td>0.595</td>
</tr>
<tr>
<td>lnPARK</td>
<td>201</td>
<td>6.107</td>
<td>0.504</td>
<td>-4.489</td>
<td>6.908</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummy Variables</th>
<th>Number</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFS</td>
<td>201</td>
<td>0.507</td>
<td>0.501</td>
<td>1</td>
</tr>
<tr>
<td>LU2</td>
<td>201</td>
<td>0.129</td>
<td>0.336</td>
<td>0</td>
</tr>
<tr>
<td>LU3</td>
<td>201</td>
<td>0.194</td>
<td>0.396</td>
<td>0</td>
</tr>
<tr>
<td>LU4</td>
<td>201</td>
<td>0.358</td>
<td>0.481</td>
<td>0</td>
</tr>
<tr>
<td>LU5</td>
<td>201</td>
<td>0.054</td>
<td>0.228</td>
<td>0</td>
</tr>
<tr>
<td>LU6</td>
<td>201</td>
<td>0.099</td>
<td>0.099</td>
<td>0</td>
</tr>
<tr>
<td>LU7</td>
<td>201</td>
<td>0.035</td>
<td>0.184</td>
<td>0</td>
</tr>
<tr>
<td>LU9</td>
<td>201</td>
<td>0.02</td>
<td>0.14</td>
<td>0</td>
</tr>
<tr>
<td>ECONOMY</td>
<td>201</td>
<td>0.094</td>
<td>0.293</td>
<td>0</td>
</tr>
<tr>
<td>SATURDAY</td>
<td>201</td>
<td>0.313</td>
<td>0.465</td>
<td>0</td>
</tr>
<tr>
<td>SUNDAY</td>
<td>201</td>
<td>0.144</td>
<td>0.352</td>
<td>0</td>
</tr>
<tr>
<td>MON-THURS</td>
<td>201</td>
<td>0.104</td>
<td>0.306</td>
<td>0</td>
</tr>
</tbody>
</table>

The variables considered in the empirical phase of this analysis are described and defined in Table 4.2 and shows that the dataset covers surveys between 1986-2003. 390 cases of food superstore developments were originally extracted from the TRICS database but depending on the model specification, approximately 190 cases were
removed (varying by model specification), as many of these sites featured multiple instances of missing values for some important explanatory variables. There are no a-priori expectations that the distribution of missing values were associated with any systematic biases, furthermore Shapiro-Wilk normality tests (see (Shapiro & Wilk 1965)) on the final analysis dataset reveals that no bias was imposed.

The dummy (ECONOMY) is used to separate the sites into the more 'mainstream' food superstores (e.g. Tescos or Sainsburys), and the less 'mainstream' food superstores which feature in much smaller chains and/or have a discount or low cost market orientation; taking the value 1 if the site is a 'less mainstream' store.

AVHS, AVLH and AVHC as discussed above, are socio-economic variables reflecting population composition characteristics derived from Census and NOMIS data. AVHS measures the average household size, AVLH is a measure of the proportion of large households in the area (defined as households with three or more people) and AVHC is the proportion of households in the area with children.

RESISTANCE - measures the distance to the nearest competitor (in kilometres). This variable proxies the generalised cost of travelling between the surveyed site and the nearest similar site, consequently acting as an indirect attraction factor.

The other variables included into the analysis are; PFS - which is a dummy indicating whether a site has a petrol station or not (i.e. an attraction factor\(^7\) for those who do drive cars, as they can buy their fuel also). CAR - which is a measure

\(^7\)There are a range of other factors/features at food superstores that could also be considered as attraction factors such as photo processing facilities, a pharmacy, cafe or some other service. However it was not possible to feature such information into the analytical dataset as they have not been consistently recorded within TRICS to date.
CHAPTER 4. DATA AND METHODOLOGY

of the level of average household car ownership within the site area. PARK, measures the number of parking spaces at a site, reflecting the ease at which a potential vehicle traveller can stop and shop.

Figure 4.3 shows the flow of car traffic both arriving at and departing from the sample of food superstores. A number of interesting characteristics appear from this figure, such as the implied time spent shopping, which can be crudely observed by looking at the vertical distance between the two lines. This distance may also reflect other issues such as the number of staff on tills at different times of day and/or the different purposes being achieved by each visit. For instance the distance between the two lines becomes discernable around the lunchtime period. This is likely a product of both (i) an increased number visits being made by people working in the surrounding area to grab some groceries during their lunch break and (ii) increased provision of open checkouts to account for this rush.

From this figure it can be seen that traffic flows are largely relatively constant through the period 10am-8pm. Although it is widely accepted that there are peaks in travel behaviour at food superstores, the general implication from the data held in TRICS is that these peaks are not so pronounced as one might expect. Though only conjecture, as the data sources used here provide no means to prove this, it is likely that this is due to the different societal groups visiting at different times of the day. Though it should be added that this is guided by 7 years part time work within a large supermarket. For instance, the morning period customers will most likely be elderly customers, people on holiday and people who run their own businesses and can keep their hours flexible. As the day progresses, the lunchtime peak sees a younger customer who is in a hurry to get lunch and some necessary groceries. The afternoon sees mothers and fathers who come in after collecting their children from
nursery and school, followed by the evening rush on the return home from work. Those that do not wish to shop at these times would perhaps come after an early dinner, which is why the peak period extends so late into the evening.

![Mean trip generations for sample of Food Superstores](image)

Figure 4.3: Total car traffic arrivals by time of day - Food Superstores
4.1.4 Residential developments data

Table 4.3: Descriptive Statistics for Residential Data

<table>
<thead>
<tr>
<th>Continuous Variables</th>
<th>Mean</th>
<th>Median</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \ln T )</td>
<td>2.823</td>
<td>2.798</td>
<td>1.467</td>
<td>-1.043</td>
<td>6.945</td>
<td>146</td>
</tr>
<tr>
<td>( \ln MSA )</td>
<td>11.876</td>
<td>11.791</td>
<td>0.478</td>
<td>10.857</td>
<td>13.801</td>
<td>146</td>
</tr>
<tr>
<td>( \ln V H H S )</td>
<td>0.906</td>
<td>0.908</td>
<td>0.050</td>
<td>0.786</td>
<td>1.027</td>
<td>146</td>
</tr>
<tr>
<td>( \ln H H E M P )</td>
<td>0.086</td>
<td>0.074</td>
<td>0.098</td>
<td>-0.118</td>
<td>0.413</td>
<td>146</td>
</tr>
<tr>
<td>( \ln P S V )</td>
<td>3.931</td>
<td>4.248</td>
<td>0.953</td>
<td>0</td>
<td>4.500</td>
<td>146</td>
</tr>
<tr>
<td>( \ln A )</td>
<td>4.892</td>
<td>4.682</td>
<td>1.333</td>
<td>2.416</td>
<td>8.957</td>
<td>146</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummy Variables</th>
<th>Mean</th>
<th>Median</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>( LU_1 ) (base)</td>
<td>0.014</td>
<td>0</td>
<td>0.117</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>( LU_2 )</td>
<td>0.096</td>
<td>0</td>
<td>0.295</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>( LU_3 )</td>
<td>0.295</td>
<td>0</td>
<td>0.457</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>( LU_4 )</td>
<td>0.425</td>
<td>0</td>
<td>0.496</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>( LU_5 )</td>
<td>0.089</td>
<td>0</td>
<td>0.286</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>( LU_9 )</td>
<td>0.082</td>
<td>0</td>
<td>0.276</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
</tbody>
</table>

Table 4.3 shows the descriptive statistics for the residential developments data, there are 146 separate site surveys used in the empirical analysis coming from 14 uniquely defined household types, thus making the information in this particular data substantially richer than the office and food superstore samples.

Table 4.3 identifies the different household classifications which are recorded within the TRICS database as well as the different land zone types observed in the residential site specific dataset. From this table it can be seen that the dwelling types in places can relate to very different situations, such as the ‘institutional hostels’ and the ‘caravan parks’. The rationale for considering such diverse accommodation types is defended on the grounds that each site type represents a different lifestyle choice, but with the same underlying demand inducing characteristics. The char-
Table 4.4: Dichotomous variable classification schema

<table>
<thead>
<tr>
<th>Household Classification Variables</th>
<th>Mean</th>
<th>Median</th>
<th>s.d.</th>
<th>Min</th>
<th>Max</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - Houses Privately Owned</td>
<td>0.281</td>
<td>0</td>
<td>0.451</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>B - Houses for Rent</td>
<td>0.021</td>
<td>0</td>
<td>0.142</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>C - Flats Privately owned</td>
<td>0.034</td>
<td>0</td>
<td>0.182</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>D - Flats For Rent</td>
<td>0.075</td>
<td>0</td>
<td>0.265</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>E - Institutional Hostels</td>
<td>0.027</td>
<td>0</td>
<td>0.164</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>F - Sheltered Accommodation</td>
<td>0.075</td>
<td>0</td>
<td>0.265</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>G - Student Accommodation</td>
<td>0.021</td>
<td>0</td>
<td>0.142</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>H - Nurses Homes</td>
<td>0.007</td>
<td>0</td>
<td>0.083</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>I - Caravan Parks (Non-Holiday)</td>
<td>0.007</td>
<td>0</td>
<td>0.083</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>J - Holiday Accommodation</td>
<td>0.082</td>
<td>0</td>
<td>0.276</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>K - Mixed Private Housing</td>
<td>0.240</td>
<td>0</td>
<td>0.428</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>L - Mixed Non-Private Housing</td>
<td>0.034</td>
<td>0</td>
<td>0.182</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>M - Mixed Private/Non-Private</td>
<td>0.075</td>
<td>0</td>
<td>0.265</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
<tr>
<td>N - Retirement Flats</td>
<td>0.021</td>
<td>0</td>
<td>0.142</td>
<td>0</td>
<td>1</td>
<td>146</td>
</tr>
</tbody>
</table>

Characteristics of the dwelling type arguably reflect, on average, initial endowments of travel behaviour in the sense that they cluster together groups who will generate more (or less) trips on average when compared with other dwelling types. However, given this initial endowment, the marginal propensity to generate trips is due to the same demand influencing relationships across all household types. More on these assumptions are offered in Chapter 7.

Table 4.4 indicates the different household types considered in the empirical phase of the analysis. From the median values column it can be seen that no single household type dominates the dataset e.g. no single household type constitutes more than 50% of the observations. The mean values indicate that privately owned houses and mixed private housing are the most common housing types.

The flows of traffic identified from residential developments, not controlling for
dwelling type\textsuperscript{8} follow patterns that one might expect \textit{a-priori}. It is seen that the morning peak is largely concentrated around standard work-departure and school-run times, with a small peak in arrivals where those not working return from the school-run. There is a second small peak around the lunchtime period where some people return home perhaps after half days, or just for their lunch break. The evening period sees the largest peak in arrivals as the majority of residents return home from work or from school etc. This evening peak is more widely diffused than the morning period peak due most likely to people whom combine the journey home from work with a trip to the supermarket or to see a friend for instance. There is a clear peak in departures between 19:00-20:00 which is likely to be indicative of the level of social activity i.e. journeys purely for recreational purposes.

\textsuperscript{8}When controlling for dwelling type, the general pattern of the relationship remains the same, albeit that the magnitude of the peaks in some households are not so extreme when compared with others. As such dwelling type generally reflects the magnitude of trips generated, rather than generating trips at different times of the day.
4.2 Methodology

Given the theoretical framework discussed in Chapter 3, an Activity-Based Trip Generation Model is specified under the same auspices as a standard (derived) demand relationship for consumer behaviour. Transport is a necessary component in the production, delivery and consequent consumption of any good or service, thus its demand is nearly always derived from the demand for these other goods (Ortuzar & Willumsen (1994), Hensher & Button (2000) and Hibbs (2003)). Following funda-

9 It is widely understood that the demand for transport primarily stems from the desire to participate in activities or purchase goods (see for example Ettema & Timmermans (1997), for a more comprehensive discussion on the role of activity in transport analysis).
mental micro-economic concepts, the demand for a good or service is determined by the desire and ability to purchase the good (e.g. characteristics of the good, and individual capacity to consume), as well as interactions with substitute markets/goods.

Given the small sample problems associated with the sample of data for office developments, as well as some issues of recurring heteroskedasticity in the food superstore data, a bootstrap algorithm is used (discussed in more detail in Section 4.2.2). This 'semi-parametric' approach provides accurate inference, irrespective of the sample size and is also not constrained to the assumption of a normal distribution of the errors. The bootstrap methodology is outlined below in more detail, along with the associated hypothesis testing procedure.

4.2.1 The general semi-parametric model

The general model which underpins the three analytical chapters is as follows:

Model I

\[
\ln(T_{m,o,d}) = \mu_d + \sum B^{-1} \sum (\beta_1^*, \beta_2^*, \beta_3^*)' \ln(\delta_d, \ln \gamma_{o,d}, \lambda_d)
\] (4.1)

where

'ln' denotes the natural logarithm of a variable,

'T' is the total daily traffic flow to a site,

subscript 'm' denotes the mode of travel where \( m \in M \) and 'M' is the set of travel modes,
CHAPTER 4. DATA AND METHODOLOGY

subscript ‘o’ denotes the trip origin where o ∈ O and ‘O’ is the set of origins,

subscript ‘d’ denotes the destination (i.e. the office development) where d ∈ D and ‘D’ is the set of destinations,

‘μ’ is the stochastic error term,

‘B’ is the number of bootstrap replications,

‘(β_δ, β_γ, β_λ)’ are parameter vectors relating to the vectors of independent variables (δ, γ, λ),

‘*’ indicates the sum of the coefficients from each of the B bootstrap replications,

‘δ’ is the Gross Floor Area (GFA) of each site,

‘γ’ is a vector of more general site specific characteristics and surrounding area socio-economic characteristics.

The vector ‘λ’ is formed of (up to) 9 dummy variables relating to the land zones identified below, hence logs are not taken.

B = 100, 000) and is the number of bootstrap replications and (*) denotes the sum of the estimated coefficients from each of the bootstrap replications. Each replication creates a new data set with the same dimensions as the original dataset (i.e. the same number of observations and variables), however each cell is uniquely drawn with replacement from the original dataset, with each individual observation in the
original dataset having an equal probability of being drawn. For further explanation of the tenets of bootstrapping processes, see for example Efron & Tibshirani (1993). or for a brief overview of bootstrapping in econometrics see the survey article of MacKinnon (2002).

The stochastic error term ($\mu$) is likely to be influenced by random factors but possibly also by some potential direct-demand (which is not measurable) for travel (Mokhtarian et al. (2001)). This is not likely to be a source of (noticeable) bias, as people who make journeys for the direct pleasure derived from driving, are far less likely to couple that journey with other activities. Therefore such trips are not likely to feature largely within the present data.

Three simplifying assumptions are made to the model;

- **1**: $M$ is constrained to passenger vehicle traffic only.
- **2**: $O$ is not known with certainty, and is therefore assumed to be a function of the surrounding areas characteristics.
- **3**: $D$ is constrained to one type of destination, in this instance food superstores. i.e. the model is estimating the levels of trips for only one individual type of activity.

The dependent variable is defined as the average hourly car trip arrivals for food superstores and trip departures for residential developments\(^1\). For office developments the dependent variable was expressed as total daily arrivals (rather than hourly average) though given the data and model structure this bears no implication on the interpretation of the results other than some slight caution in the discussion

\(^1\)The average and total trip generations at residential developments are almost identical for departures and arrivals
CHAPTER 4. DATA AND METHODOLOGY

of dummy variables.

The most general model (Model I) is conducive with the activity-based theory of travel demand through its constrained analysis of a single activity type. The inclusion of socioeconomic characteristics in the model reflects constraints on time and income within households and furthermore the demand to participate in work. i.e. larger households have more extensive and diverse demand requirements than smaller households thereby requiring a larger income for their general maintenance. It is thus conjectured that the demand to travel to a given activity by a specific mode of transport (to work by car) is therefore likely to be subject to a set of estimated parameters which are heterogenous to alternative activity purposes.

Equation (4.1) implies that the desire to travel to a site (by any chosen mode of travel, for instance cycling, walking or taking public transport), is determined by factors that influence the ability and/or wish to partake in (or consume) the activity (or service) which that site offers. In particular, ‘E’ represents the economic characteristics of the local population, thus capturing the ability of individuals within that area to consume. ‘S’ is site specific attributes, which may be considered as features of that site which may serve to attract more trips. The variable ‘G’ suggests that wider geographic information may have some bearing on trip levels including site accessibility, and potential resistance offered by the existing geography. Finally ‘SE’ are socio-economic characteristics, which reflect (in part) lifestyles, and consequently, consumption choices within a given area. Previous multi-equation trip generation models adopting similar relationships have been advocated for instance in Brady & Betz (1971), Hensher & Dalvi (1978) and Washington (2000). Although the inclusion of physical characteristics of a site is not often introduced, probably due to the nature of the datasets used in previous works.
4.2.2 The Bootstrap procedure.

Misspecification problems arising from the variation and distribution of variables in the composite dataset used in this study are overcome using a bootstrap algorithm, for the creation of 'synthetic' standard error values. In this way it is possible to undertake more-meaningful statistical inference in the presence of such problems (see for instance Mooney & Duval (1993)). The bootstrap algorithm for estimating standard errors (Efron & Tibshirani (1993), pp47)) implements the following three steps;

1. B independent bootstrap samples \( x_1^*, x_2^*, ..., x_B^* \), each consisting of \( n \) data values drawn with replacement from \( x \) are selected. Where \( x \) is the initial dataset and the bootstrap samples are bound by the characteristics of the initial dataset.

2. Evaluate the bootstrap replication corresponding to each bootstrap sample,

\[
\hat{\theta}^*(b) = s(x^{*b}) \quad b = 1, 2, ..., B.
\]

3. The standard error is estimated by the sample standard deviation of the B replications;

\[
\hat{s}_B = \left\{ \frac{\Sigma [\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2}{(B - 1)} \right\}^{\frac{1}{2}}
\]

Where,

\[
\hat{\theta}^*(.) = \frac{\Sigma \hat{\theta}^*(b)}{B}
\]

This algorithm produces an empirical distribution and consequently an empirical standard deviation that can be shown to approach the population standard deviation.
as the number of bootstrap replications tends to infinity,

\[ \lim_{b \to \infty} s_{EB} = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*) \]

Note that \( \hat{F} \) is the bootstrap samples \((x_1^*, x_2^*, ..., x_n^*)\). This is based around the assumption that the data sample available to the analyst is reflective of the population sample\(^{11}\). Thus the bootstrap method produces the appropriate empirical distributions specific to the data in the original sample.

Efron & Tibshirani (1993) stress that although the bootstrap method has a number of uses it was primarily designed as a tool to estimate standard errors, thus it is well suited for this application. Furthermore, Mooney & Duval (1993) highlight, with supporting application, the benefits of the bootstrap method for overcoming non-normality in a standard OLS regression model (which has the effect of making traditional/frequentist inference inaccurate). This is because the bootstrap algorithm produces a normally distributed error structure (using the empirical distribution which is based directly on the characteristics of the original dataset), making standard statistical inference possible.

4.2.3 The Achieved Significance Level (ASL)

As this 'semi-parametric' modelling approach produces the exact distribution of the coefficients in the model it thus allows for exact inference to be conducted, rather than approximations based on a normality assumption. This is done using an achieved significance level (ASL), see Efron & Tibshirani (1993), which are analogous to standard p-values for mainstream significance tests, but for the exception that

\(^{11}\)Or alternatively merely treated as an independent population
they are exact with respect to the data sample and functional form in question. The following defines the test procedure;

$$t(z^*) = \frac{\hat{z}^* - \hat{z}}{\sigma^*/\sqrt{n}}$$

where \(n\) is the number of observations in the original data sample, and \(z^*\) are the bootstrap replications of the value for \(z\). \(\hat{z}\) is the value of \(z\) that is being hypothesised/tested against, which is normally set to zero, and \(\sigma^*\) is the standard deviation of the observed bootstrap coefficients \(z^*\). Applying the conventional null hypothesis for standard two-tailed significance tests the following null hypothesis is offered;

$$H_0 : \hat{z} = 0$$ (4.2)

thus reducing the hypothesis test to;

$$t(z^*) = \frac{\hat{z}^* - 0}{\sigma^*/\sqrt{n}} = \frac{\hat{z}^*}{\sigma^*/\sqrt{n}}$$ (4.3)

This is the ‘critical’ t-value\(^{12}\) for testing the assumption that the estimated coefficient is equal to zero. Once this has been computed, the empirical distribution of the coefficients must then be translated about zero (i.e. the mean is forced to be zero), which is the hypothesised value of the observed coefficient. Following this, each bootstrap replication is then tested against this critical value using the following decision criteria;

$$H_0 : t_{calc} < t_{crit}$$ (4.4)

thus indicating that there is no evidence that the parameter of interest (i.e. the individual bootstrap replication) is significantly different from zero. The ‘transla-

\(^{12}\) Analogous to the pre-defined table values given in many statistics or econometrics texts, see for example Gujarati (2003)
tion' of the empirical distribution of $z$ (so as to create a new 'null distribution' with mean equal to the null hypothesis) is done using the following formula;

$$
\tilde{z}_i = z_i - \bar{z} + z_{H_0} = z_i - \bar{z} + 0
$$  \hspace{1cm} (4.5)

thereby centering the distribution first about zero (by subtracting the observed mean $\bar{z}$), and then redistributing it about the hypothesised mean $z_{H_0}$. These values $\tilde{z}_i$ are subsequently used in calculating the t-values for hypothesis tests on the now known empirical null distribution, noting that the null distribution does not need to be normal, using the formula;

$$
t(\tilde{z}^*) = \frac{\tilde{z}^* - 0}{\sigma^*/\sqrt{n}}
$$  \hspace{1cm} (4.6)

thus t-values are created for each bootstrap replication, where $\tilde{z}^* = t_{\text{calc}}(z^*)$. These values are then compared to the previously calculated critical value in order to reveal the number of bootstrap replications which violate the null hypothesis, i.e. When $t(\tilde{z}) > t_{\text{critical}}$ it is not possible to reject the null hypothesis that the (empirical or untranslated) distribution of $z$ is centered around zero.

Defining;

$$
\gamma = \# t(\tilde{z}) > t_{\text{critical}}
$$  \hspace{1cm} (4.7)

i.e. the number of times that the null hypothesis cannot be rejected. Then the achieved significance level (ASL) is found to be

$$
ASL = \frac{\gamma}{B}
$$  \hspace{1cm} (4.8)

where $B$ is the number of bootstrap calculations. This is then interpreted as the probability that the (untranslated) empirical distribution of the bootstrap coef-
ficients is centered around a zero mean, and can consequently be considered statistically insignificant.

The key difference between this test and standard significance test used in mainstream econometric applications is in the assumption that the estimated coefficient is derived from a standard normal distribution is now flexible. Hypothesis testing becomes feasible irrespective of the actual distribution that the data follows, where this evidently includes a normal distribution. Therefore the bootstrap framework coupled with the ASL approach to interpreting the significance of the coefficients provides a theoretical advantage over the standard OLS t-test procedure. A further advantage is that the ASL's allow for the generation of asymmetric confidence intervals, thus providing more information than can be achieved using standard inferential assumptions.

Given the estimation and inference methodology discussed within this chapter, the remainder of the thesis presents and discusses the empirical results. The three specific datasets (also discussed within this chapter) namely Offices, Food Superstores and Households are considered in turn in the next three chapters.

\footnote{Data handling, descriptive analysis and model estimation are implemented with the software package STATA (2004).}
Chapter 5

Office Block Trip Generation Models

5.1 Introduction

This chapter\textsuperscript{1} reviews the determinants of work-based trip generations (attractions). Through econometric estimation of elasticities of policy-relevant variables, this study contributes much needed empirical evidence and debate to help support policy development in the transport planning community. The results aid the development of work-based travel plans and further add depth to the understanding and effectiveness of local level policy interventions. Of particular interest are the way in which negative externalities (both uniformly and non-uniformly mixing pollutants) can be reduced. Short conceptual discussion of Local Travel Plan implementation would

\footnotesize{\textsuperscript{1}Some of the results contained within this chapter may be found in;

lead to the conclusion that policy instruments introduced at this scale may be less effective than previously hoped for. The resulting implication is that individual sites subject to Section 106 agreements may fail to meet the necessary pass rate criteria through no fault of their own. Moreover, the Section 106 agreement, coupled with LTP's may result in incorrect upstream policy interventions trying to cater for inefficiencies in the travel planning process, which quite simply are not a true feature.

The data used in this analysis is a composite dataset assembled to investigate the determinants of car borne UK office development trip rates. Planners and traffic engineers on the ground have to make defensible cases for decisions on the acceptance/rejection or scale of land use developments at particular sites. Typically they rely on comparators drawn from 'similar sites' elsewhere. In an attempt to systematise and expand the range of available comparator sites, the Trip Rate Information Computer System (TRICS) has evolved. This is essentially a UK based mirror (though with more detailed site specific information) of the US Institute of Transport Engineers (ITE) trip rate generation. Another element of the composite dataset comprises socio-economic data from publicly available sources. As such it represents the data that in principle could be readily used by the most informed transport planners/engineers. In short, it represents the best available real data without commissioning new and costly travel/traffic surveys. Nevertheless, this data features a considerable number of missing elements. This chapter focusses on eight alternative approaches applied to address missingness in the dependent variable.²

²Cases where there is missingness in the right hand side variable(s) were addressed by simple casewise deletion. Although this may feasibly impart some systematic bias onto the results produced, for the purpose of this exercise this is not a detrimental feature of the study.
From the perspective of policy considerations, the desire to analyse the determinants of trip generations at office developments is borne of a desire to manage the externalities associated with a site's business operations. The trip to work (and accompanying return trip) serves to be an area of extended debate in which there are many accomplished works looking into multi-faceted aspects of travel behaviour, see for example Jara-Diaz (1998).

Shortle & Horan (2002), describe non-point pollution as pollution that can be specifically attributed to a specific entity, albeit not necessarily generated at the entity's site of residence. In this instance this can be described as the pollution generated by an office block for the purpose of fulfilling its business activities, which are not borne at the office block. i.e. the negative externalities caused by their employees travelling to the office, which is a necessary trip to ensure that the office block's business activities are achieved. Hence the pollution generated in the trip to work is a burden which the office block should in some way bear as its own, albeit the decision over which mode employees use to travel to work is often beyond the scope of the employer to control (at least in the short term). The rate at which a specific site contributes to non-point pollution concerns will be determined by the employee's place of abode, and the spatial proximity this has with the place of work. There are essentially two situations which may occur:

1. The employee lives within the local vicinity of the workplace ('localised employee).

2. The employee lives without the local vicinity of the workplace

The proportions of each will then give rise to three potential situations, outlined in Figure 5.1, namely (i) where all employees live nearby, (ii) all employees live far away or, most likely, (iii) some live nearby and some live far away. Though unable
to define the precise proportions of employees that live within a certain radius of the firm given the data available, the results of the estimated models will still provide some bearing on the way in which these effects can be managed. Further the results will provide an indication of the best ways in which these reductions can be achieved.

This chapter proceeds by defining the specific modelling framework for analysing the office developments data. Following this and given the data considerations the chapter proceeds to discuss the analysis of missing data, describing the 8 specific modelling methods used to circumvent the problem. Section 5.3 then applies the methods discussed to the data with the remainder of the chapter focussing attention upon the implications of the results with particular respect to policy interventions.

### 5.2 Modelling Strategy and Estimation

The following section outlines the approach taken to estimate the models in this and the subsequent substantive chapters of the thesis. As discussed in section 7.5, the regressors are grouped according to their type, e.g. land zone indicator or socio-
econometric indicator etc., and then alternative nesting structures are built up around these groups of variables. Based upon a general to specific modelling approach, groups of indicators are then sequentially removed so that the alternative model structures can be compared. For the purpose of the present chapter it is pertinent to present the results from each of the alternative nesting structures, though for the remaining two substantive analyses only the results of the preferred specifications are formally presented.

Given this discussion and the structure of the general model, this gives rise to the following model structures;

Model II

\[
\ln(T_{m,o,d}) = \mu_d + \sum [B^{-1} \sum_{b=1}^{B} (\beta_b^*, \beta_h^*)]'(\ln \delta_d, \lambda_d) \quad (5.1)
\]

Model III

\[
\ln(T_{m,o,d}) = \mu_d + \sum [B^{-1} \sum_{b=1}^{B} (\beta_b^*, \beta_h^*)]'(\ln \delta_d, \ln \gamma_{o,d}) \quad (5.2)
\]

Model IV

\[
\ln(T_{m,o,d}) = \mu_d + \sum [B^{-1} \sum_{b=1}^{B} (\beta_b^*)]'(\ln \delta_d) \quad (5.3)
\]

Subject to the same simplifying assumptions as Model I (the general model) as discussed in Chapter 4.

The most general model (Model I, where II, III and IV are nested versions of I) is thus conducive with the activity-based theory of travel demand, through its constrained analysis of a single activity type. Such that any parameter estimates resulting from this model are not biased by alternative activity trip types. The inclusion of socioeconomic characteristics in the modelling framework reflects constraints on time and income within households and furthermore the demand to participate
in work. i.e. larger households have more extensive and diverse demand requirements than smaller households thereby requiring a larger income for their general maintenance. It is thus conjectured that the demand to travel to a given activity by a specific mode of transport (to \textit{work} by \textit{car}) is therefore likely to be subject to a set of estimated parameters which are heterogenous to alternative activity purposes. This hypothesis will be returned to later.

Although expected \textit{a-priori} to be the least well specified model, equation (5.3) features commonly in the trip generation analysis literature. This specification has some redeeming characteristics, in that it simplicity reflects in part the simplicity of the travel demand behaviour for an employment site. i.e. trips are a direct result of the level of business activity, which is directly proportional to the size of the development.\footnote{This is deserving of a number of caveats, as site size is arguably reflective of other characteristics and must therefore be combined with the other factors hence the specification of models I and II} In developing the models within this general to specific analysis, the focus will initially remain upon the effects of imputation on the observed coefficient values. Preferred model specification will be based on a number of criteria including the $R^2$ and the user cost of implementation.

\subsection*{5.2.1 Missing Data and Imputation Methods}

The data available for estimation in the TRICS database is unfortunately subject to problems of missing data, as alluded to in the data section. A number of approaches exist in the standard toolkit of statisticians to address the problem, differing enormously in their level of sophistication. Missing data problems may detract from the integrity of model estimates used in prediction and forecasting and affect both left hand side and/or right hand side variables. Inevitably where missing data features on both sides of the model used to define their relationship, the problem becomes
additively more complex.

Hence this specific analysis considers and applies a number of alternative methods used to deal with this missingness, in this case of the left hand side (or dependent) variable only. Missing data is classified by empty cells within the observed data matrix $D$, where the data matrix contains all available terms (both dependent and independent). Consider the following $(5 \times 3)$ example;

$$
D = \begin{pmatrix}
Y_1 & X_{11} & X_{12} \\
Y_2 & X_{21} & X_{22} \\
Y_3 & X_{31} & X_{32} \\
Y_{IC} & X_{41} & X_{42} \\
Y_5 & X_{51} & X_{52}
\end{pmatrix}
$$

Before imputing values for the missing data elements one key consideration must be borne in mind, which is the nature of the missingness. The nature of the missing data must be carefully considered.

There are three types of missing data (sometimes known as item non-response) broadly considered in the literature surrounding imputation methods, following the definitions used by Kofman & Sharpe (2000) these are;

**Missing Completely At Random (MCAR)** when the missingness in $y$ is independent of both $x$ and $y$. The missing data are then missing-at-random and the observed data are observed-at-random.

**Missing At Random (MAR)** when the missingness in $y$ depends on $x$ but not on $y$. Missing data are still missing-at-random but observed data are no longer observed-at-random.
Non-Ignorable (NI) when the missingness in $y$ depends on $y$ and possibly also on $x$

Most crucially of importance is to understand whether the missing data is type 'NI'. In this instance as Kofman & Sharpe (2000) point out, it is highly likely that the distribution of the missing data is dependent on information/variables which are not present within the available dataset. When a variable is considered to be 'non-ignorable', then there is no evidence that the relationship between the predictor variables and the explained variables (e.g. $X$ and $Y$) is of use in the extrapolation of unobserved values of the dependent variable. In such an instance, imputation should be avoided, although as Little & Rubin (2002) identify there are methods available to cope with this.

Furthermore it should be noted that for similar reasons, if the data is MCAR then there is still no evidence that the observed dataset holds any information which will be of value in imputing missing data values [this is not correct MCAR means its missingness is not explained by X, however the actual value of Y may still be a function of X]. The ignorability condition subsequently concerns itself with the relationship between the missing in $Y$ in relation to the observed $X$, and provides no indication of the relationship between the actual (absolute) values of $Y$ as a function of $X$.

Kofman & Sharpe (2000) state that "..., it is impossible to test the ignorability assumption against the NI assumption.", though that is not to say that it should be disregarded. The implication is that one must go through a logical decision making sequence with well justified and rationalisable decision making criteria in defining the nature of the missing data. Data which is defined as Missing Completely at Random is data with a probability of being missing which is entirely independent
of any information contained within the data matrix (where the data matrix is the complete matrix containing both the dependent and independent variables with the complete and the incomplete observations), e.g. independent of $Y$ or $X$. 
The following variance/covariance matrix defines the relationship between the observed missingness (defined as $M$) in $Y$ and a selection of the key explanatory variables (as defined earlier in Table 4.1).

<table>
<thead>
<tr>
<th></th>
<th>ingfa</th>
<th>Inemp</th>
<th>Inpark</th>
<th>Incar</th>
<th>Inpsv</th>
<th>lavhs</th>
<th>lnaemp</th>
<th>napop</th>
<th>lnhepp</th>
<th>yearopen</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lngfa$</td>
<td>$-0.05$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnemp$</td>
<td>$-0.02$</td>
<td>0.95</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnpark$</td>
<td>$-0.05$</td>
<td>0.89</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnincar$</td>
<td>0.27</td>
<td>0.26</td>
<td>0.30</td>
<td>0.16</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnlnpsv$</td>
<td>$-0.24$</td>
<td>$-0.01$</td>
<td>0.01</td>
<td>$-0.11$</td>
<td>0.04</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnlavhs$</td>
<td>0.09</td>
<td>0.13</td>
<td>0.09</td>
<td>0.07</td>
<td>0.30</td>
<td>$-0.26$</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnlnlnaemp$</td>
<td>$-0.08$</td>
<td>$-0.13$</td>
<td>$-0.19$</td>
<td>$-0.16$</td>
<td>$-0.55$</td>
<td>0.20</td>
<td>$-0.15$</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$lnlnnapop$</td>
<td>$-0.08$</td>
<td>$-0.06$</td>
<td>$-0.14$</td>
<td>$-0.05$</td>
<td>$-0.56$</td>
<td>0.15</td>
<td>$-0.14$</td>
<td>0.98</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$lnlnlnhepp$</td>
<td>0.01</td>
<td>$-0.27$</td>
<td>$-0.21$</td>
<td>$-0.47$</td>
<td>0.07</td>
<td>0.13</td>
<td>0.38</td>
<td>0.22</td>
<td>0.05</td>
<td>1.00</td>
</tr>
<tr>
<td>$lnlnyearopen$</td>
<td>0.33</td>
<td>$-0.23$</td>
<td>$-0.30$</td>
<td>$-0.14$</td>
<td>$-0.27$</td>
<td>0.02</td>
<td>$-0.17$</td>
<td>0.29</td>
<td>0.29</td>
<td>$-0.05$</td>
</tr>
</tbody>
</table>

The matrix reveals that the missingness in $Y$ can be seen to be correlated to some degree with car ownership ($0.27$), public services provision ($-0.24$) and the year of a sites opening ($0.33$). This suggests some evidence of a potential relationship, though the strength of the correlation is not enough to make this a certainty. These findings are further elaborated upon in the data section.
In the development of the trip generation model, with the desire to use as much information as possible, eight alternative methods will be applied for imputing incomplete cells within the data matrix. This will provide the added benefit of being able to assess how robust the imputed values are to imputation method, which is chosen by the analyst, the methods user are as follows:

**Specification (a)** Listwise deletion

**Specification (b)** Mean observation replacement

**Specification (c)** Group mean observation replacement

**Specification (d)** Worst case scenario

**Specification (e)** Simple random imputation

**Specification (f)** Regression based forecast

**Specification (g)** Bartlett’s ANCOVA regression

**Specification (h)** Approximate Bayesian bootstrap

All specifications, with the exception of specification (h), can be defined as 'single' imputation methods. The distinction between a single and a multiple imputation (MI) method is characterised by MI methods using more than one choice of imputed value for any given empty data cell. For instance, making a number of draws from a known distribution for the missing data, then observing the changes in model output when using these alternative draws. Given just one empty cell in the data matrix, there is no restricting the number of random draws that could be made, though as Efron & Tibshirani (1993) identify, from a practical point of view in resampling

---

4This approach will only be used for Model I, where the land zone indicators are used to define the groups.
experiments it is possible to restrict analysis to a number of repetitions that can be more easily handled.

In the subsequent discussion on imputation methods, the following notation is used: \( \hat{y} \) denotes the estimated/imputed value for the missing observation by the method denoted by a superscript. The subscript \( ic \) is used to identify the incomplete case, or missing observation, while \( cc \) denotes complete case or non-missing observations. \( N \) is the number of observations and \( n \in N \).

**Specification (a) Casewise deletion (CW)**

\[
\hat{y}_{ic}^{CW} \ni D \quad (5.5)
\]

\( Acw \) : *Out-of-sample unobservable behaviour is distributed identically to in-sample behaviour*

Casewise deletion methods encapsulate two main categories, Listwise and Pairwise deletion. Each method handles the relationship between multiple variables (with multiple instances of missing data) in a slightly different way. Listwise deletion involves the removal of all observations with ‘missing data’ in a purely indiscriminate fashion i.e. any observations that have data missing are simply removed from the analytic dataset. Therefore, in the present example, any surveys without explicit information on the modal split of trip arrivals are simply disregarded.

It is therefore implied when using this method that the distribution of the missing data is exactly the same as that of the observed data, such that their removal will not significantly change the results of any analysis. This is a strong assumption which may be unfounded however it can be overcome with careful interpretation.
A related method of imputation is pairwise deletion, which is commonly associated with the development of correlation matrices. In practice this method is identical to listwise deletion when comparing only two variables, however, in the case where there are more than two variables, this method correlates data across non-comparative groups by including certain observations for calculating correlations in some parts of the matrix whilst not in others. This may result in a 'biased' correlation matrix which would be unsuitable for multiple regression extensions and is therefore omitted from further consideration/discussion.

**Specification (b) Unconditional mean observation replacement (MV)**

\[
\hat{y}_{ic}^{MV} = \frac{1}{n} \sum_{n=1}^{N_{cc}} y_{cc} \tag{5.6a}
\]

\(A_{MV}:\) *Out-of-sample unobservable behaviour is distributed as the mean of in-sample behaviour, without exception*

\[
\hat{y}_{ic}^{MV} = \frac{1}{n} \sum_{n=1}^{N_{cc}} y_{cc} + \varepsilon_i^{MV} \tag{5.6b}
\]

\(A_{MV}:\) *Out-of-sample unobservable behaviour is distributed as the mean of in-sample behaviour, subject to some in sample random deviation*\(^5\)

The mean observation replacement approach, also known as unconditional mean imputation, does as it suggests. That is to say that the imputed value for the value of \(y\) in the incomplete (or empty) cell is equal to the arithmetic mean of the observed values of \(y\) in the complete observations.

\(^5\) in sample random deviation for the MV and GMV approach is defined as a random draw from the observed errors from the estimation of models I-IV using casewise deletion. As such shocks are not constrained to be normally distributed and are derived from realistic behaviour patterns which are known with certainty, albeit due to a restriction of a specific functional form.
Little & Rubin (2002) suggest that this approach cannot be recommended due to problems which may manifest in the variance-covariance matrices which may subsequently impair the efficiency/accuracy of least squares results. One clear drawback with the mean value approach is that at the limit i.e. as the number of missing observations (in relation to the total number of observations) tends to infinity, then it can be shown that estimation via model based methods will tend towards a constant, with the effects of explanatory variables tending toward zero. Therefore use of the mean value approach intrinsically biases estimated slope coefficients towards zero, making the dependent variable explainable by a constant value, which will ironically be the imputed mean value. i.e.

\[
\lim_{M \to \infty} \frac{\theta}{N} = \alpha
\]

Where \( M \) is the number of missing observations in the complete dataset and \( N \) is the number of complete observations in the full dataset, \( \hat{\theta} \) is in this instance the estimated least squares coefficient. Hence the approach is supplemented with some stochastic variance as in equation 5.6b, based upon a bootstrap draw on the residuals from the casewise cleansed dataset. This provides the added advantage that the shocks imposed upon the data are based upon an observed distribution of shocks rather than an imposed one. However it should be noted that the shocks are in turn determined by the choice of model, and hence the analyst will still play a role in determining the shocks imposed.

\footnote{6} or alternatively as \( \frac{1}{n} \) tends towards zero
\footnote{7} It is not enough to say that

\[
\lim_{N \to \infty} \hat{\theta} = \alpha
\]

This expression is not incorrect, however by weighting \( n \) by \( m \), the expression becomes ultimately more informative by reflecting not only the characteristics of the missing data, but the whole analytic dataset.
Its inclusion in this study is therefore merited on its parsimony and relatively low user cost.

**Specification (c) Conditional mean observation replacement (GMV)**

\[
\hat{y}_{ic}^{GMV} = \frac{1}{n} \sum_{j=1}^{J} \left( \sum_{i=1}^{r_j} y_{ij} + \sum_{i=r_j+1}^{n_j} \hat{y}_{ij} \right) = \frac{1}{n} \sum_{j=1}^{J} n_j \hat{y}_{jR} \quad (5.7a)
\]

\(A_{GMV}\): Out-of-sample unobservable behaviour is distributed as ‘within-groups’ mean of in-sample behaviour, without exception

\[
\hat{y}_{ic}^{GMV} = \left( \frac{1}{n} \sum_{j=1}^{J} n_j \hat{y}_{jR} \right) + \varepsilon_i^{GMV} \quad (5.7b)
\]

\(A_{GMV}\): Out-of-sample unobservable behaviour is distributed as ‘within-groups’ mean of in-sample behaviour, subject to some in-sample random deviation

With regards to the mean value (MV) method, one could potentially improve the efficiency of the method if the data can be segmented into observable groups, such that the mean values of groups (GMV) are used. This introduces some extra degree of variation into the values imputed for the missing data fields. Also known as ‘conditional mean imputation’, see p62 Little & Rubin (2002). This method offers an advantage over unconditional mean value imputation by recognising that data may naturally fall into uniquely observable clusters/groups. In the instance of the present study, these will be defined by the previously outlined land-zone categories, in conjunction with the hypothesis that land-zone placement influences travel characteristics at a site.

As with the MV approach, the efficiency of this method in providing an accurate representation of real life behaviour is heavily impaired by asserting a constant

---

8This approach will only be used for Models I and III, where the land zone indicators are used to define the groups.
value. In order that this deficiency is overcome to some degree, the results of the listwise imputed models are called upon to provide residuals from the largest set of available complete case data. From these residuals, bootstrap samples are drawn (with replacement), and added to the conditional and unconditional mean values. Little & Rubin (2002), discuss the use of similar methods on multi-equation settings, the extension is made here to the single-equation case on the understanding that the inclusion of shocks, derived from an observed distribution, adds further realism.

**Specification (d) Worst case scenario (WC)**

\[
y_{ic}^{WC} = y_{cx}^{max} \tag{5.8}
\]

\( \text{A}_W^C : \text{Out-of-sample unobservable behaviour is distributed identically to in-sample behaviour} \)

The worst case scenario approach involves filling in the missing data with the worst possible observation in the available data (or worst possible case if there exists a natural candidate for this). In some cases this will have no natural interpretation, and even when it appears sensible to use, it is a method which must be treated with some caution. Modal splits derived using this approach will be 100% in favour of cars and subsequently, if used for development purposes will suggest a higher trip rate than will likely be observed in reality. Although this may potentially encourage developers to seek greater parking provision, it will mean that local authorities can be more certain that the car-traffic impact of a new development will be reducible, and likely not even as large, as a development proposal might suggest.

**Specification (e) Simple random imputation (SRI)**

\[
y_{ic}^{SRI} = \frac{1}{n} Y_j \tag{5.9}
\]

\(^9\) The bootstrap is chosen over a monte carlo resampling technology so as to avoid imposing normality (or any other known form of distribution) on the residuals where it might not be appropriate.
ASRI: Out-of-sample unobservable behaviour is a random function of observed in-sample behaviour

Rubin & Schenker (1986), in arguably the most important paper contributing to the mass inception of imputation methods in applied statistics (Donald B. Rubin being a particularly active researcher concentrating in this field), apply a number of imputation methods starting with simple random imputation. This process (known also as the Hotdeck method) is quite straightforward, whereby missing observations for any given variable are simply drawn with replacement from the observed values for that variable.

This method, along with the others hitherto discussed, is considered to be inferior to some of the alternative methods presented below primarily because the imputed data do not make use of any available knowledge on the observed prior distribution of that variable (see Rubin & Schenker (1986) and Mander & Clayton (1998). These methods have subsequently sometimes been referred to as ‘ad-hoc’ methods. The following three methods to be outlined and then applied take imputation into a slightly different realm by incorporating other information into the process. These methods allow for the information of external data and/or known functional forms to help fill any empty observations. Thus to some extent there is almost a learning process, or level of artificial intelligence deriving the imputed values.

**Specification (f) Regression based forecast (RF)**

\[
\hat{y}_{ic}^{RF} = \hat{y}_i = \mathbf{x}_i' \mathbf{b}_{CW} \tag{5.10}
\]

ARF: Out-of-sample unobservable behaviour is distributed identically to in-sample behaviour

A fitted equation is estimated for the data using the observations with non-missing elements, i.e. the ‘casewise’ cleaned dataset. The estimated equation is
then used to ‘fit’ values for only the missing observations so as to fill all empty elements of the data matrix. Once the data matrix is full, the fitted equation is then re-estimated in order to observe how sensitive the variables are, with the difference between the estimated coefficients being representative of the bias incurred by not using all available data.

**Specification (g) Bartlett’s ANCOVA regression (ANC)**

\[
\hat{y}_{im}^* = \left( \frac{1}{N} \sum_{n=1}^{N_{cc}} y_{icn} \right) - \hat{\gamma}_m
\]  

\[\text{A}_{\text{ANC}} : \text{Out-of-sample unobservable behaviour is distributed identically to in-sample behaviour}\]

Following Little & Rubin (2002), Bartlett’s ANCOVA procedure can be summarised by the following steps;

1. Fill in the missing elements of the Data matrix with some initial guess (the grand mean \( \hat{y}_{cc} \) is as reasonable a guess as any other, though the choice is entirely arbitrary).

2. Define a matrix \( Z \) of missing value covariates, which has a number of columns equal to the number of missing elements, identifying each missing data value using a dummy variable. i.e. for the present data, with 15 missing values, there will be 15 separate dummy variables.

3. The regression model can now be written;

\[
Y = X\beta + Z\gamma + \varepsilon
\]  

where \( Y \) is column vector of dependent variable observations, \( X \) is the matrix of exogenous variables and \( Z \) is the matrix of missing value covariates. \( \beta \) and \( \gamma \) are the
column vectors of coefficients relating to matrices $X$ and $Z$ respectively, whilst $\varepsilon$ is a stochastic term, assumed for inferential purposes to follow a normal distribution.\(^{10}\)

(4) The correct least squares estimates of the missing values can then be expressed as:

$$\hat{y}_{im}^* = \left( \frac{1}{n} \sum_{n=1}^{N_{cc}} y_{cc} \right) - \hat{\gamma}_m$$  \hspace{1cm} (5.13)

where the first term on the right hand side of equation 5.13 is the initial guess of the missing observation, i.e. the grand mean. Thus the imputed values equal the initial guess minus the coefficient of the missing values covariate.\(^{11}\)

The estimates from this approach will generally be equivalent to those from the use of a regression based forecast approach. However the ANCOVA approach provides a unique benefit in the presence of an estimation using multiple groups delineated by dummy variables. As will be seen in the later estimations, specifications $f$ and $g$ do not provide identical results for models I and III. This is because some land zone dummies are not present in the regression based forecast approach as all the left hand side observations are missing. However the ANCOVA approach still features all land zones into the imputation process, providing an initial guess and then correcting it based upon the estimated coefficients.

This raises questions as to the validity of the information generated by the simple regression based forecast, as it will sometimes neglect potentially important information. However, even given this criticism, the method offers a more theoretically

\(^{10}\)This assumption is not a prerequisite of the ANCOVA method

\(^{11}\)\(\gamma_m\) is subtracted in all instances, even when not statistically significant, as its inclusion within the model is to correct for known data discrepancies.
CHAPTER 5. OFFICE BLOCK TRIP GENERATION MODELS

grounded approach to imputing missing values based upon a pre-defined relationship between the dependent variable and the regressors. An interesting question would be to attempt to understand if any systematic bias is imposed by using the regression based approach, compared to the ANCOVA method, under these circumstances. However in the absence of setting up a controlled experiment, this objective is beyond the scope of this thesis and cannot be answered with the data at hand.

**Specification (h) Approximate Bayesian Bootstrap (ABB)**

\[
\hat{y}_{k cm}^{ABB} = B^{-1} \sum_{b=1}^{B} (\hat{y}_{bc}^{B})
\]  

(AABB) Out-of-sample unobservable behaviour is distributed within a region of data with similar behavioural patterns to the missing data

Davison & Hinkley (1997) identify the use of the Bayesian bootstrap primarily as a tool for imputation in incomplete datasets, whereby the process uses information about the known distribution of the missing values as a 'Bayesian prior'. Then the expectation of the unknown values is derived, given the information on the prior distribution, to create a 'posterior' distribution for the unobserved data, Rubin (1981) and Rubin & Schenker (1986) for instance apply an improper Dirichlet distribution to define the prior. This approach is however quite expensive in terms of human capital (although modern computer packages make this problem less pronounced than they may have been in the past), which had resulted in the inception of the Approximate Bayesian Bootstrap (ABB) as a kind of 'short-cut' method to arriving at (generally) the same result. The ABB works on very much the same fundamental concepts of the Bayesian Bootstrap in that it acknowledges a relationship of the missing data with the observable prior distribution and so forth. In practice the Approximate method is computationally far more simple, making it easier to both understand and apply and more accessible to practitioners with a non-statistical
background. The following representation of the ABB is originally due to Lavori, Dawson & Shera (1995) and summarised by Allison (2000);

1. Do a logistic regression\(^{12}\) in which the dependent variable is whether or not the observation under consideration is missing from the sample. The independent variables are chosen by the analyst

2. Use the estimated logistic model to evaluate a predicted probability of a value being missing. This value is known as the "propensity score", see Rosenbaum & Rubin (1983).

3. Sort the observations by their calculated propensity scores and group them into quintiles (i.e. five groups)

4. From each of the five quintiles there will be a number of cases with observed data and a number of cases with unobserved data. For each quintile, the unobserved cases are then imputed by drawing with replacement from the observed cases

This process thus produces a complete (albeit partially imputed) analytical dataset which can subsequently be used for estimation. In order that the potential bias is reduced, which may be present in a single run of this process, step 4 is repeated (in an analogous fashion to normal bootstrapping procedures) on the premise that as the number of replications tends towards infinity, the accuracy of the estimated coefficients will tend toward the true population parameter (see Efron & Tibshirani (1993) for further discussion of resampling methods).

\(^{12}\)The logistic regression model can be specified quite generally as;

\[ \theta = \frac{e^{X\beta}}{1 + e^{X\beta}} \]

note that for each specific application in this chapter, the regressors in the logistic regression are the same as those used in the linear regression.
CHAPTER 5. OFFICE BLOCK TRIP GENERATION MODELS

5.3 Results

In determining the factors which influence the level of trip generations at office developments this analysis has applied missing data techniques to allow the conclusions to be found upon the most available data. The following section therefore proceeds by presenting the results for model specifications I, II, III and IV sequentially. Each model specification is initially analysed so as to observe the effects of the imputation methods (denoted (a)-(h)) on the qualitative results, looking specifically for changes in the magnitude and/or sign of the estimated elasticities. The coefficients reported in each of these tables are the bias corrected least squares estimates. As such the first part of this section is not entirely consistent with a general to specific modeling approach, this is dealt with in the second part of this section. The first part of this section alternatively provides a thorough account of the effects on the coefficient values of using the eight alternative imputation methods considered.

Table 5.4 shows the coefficients are extremely stable across all imputation methods and strongly suggest that 0.6-0.7% of car trips generated at office developments are a result of the physical scale of the business operations at that site, i.e. a 1% increase in the floor space of an office development leads to about a 0.6-0.7% increase in trips by car. Explanatory power of the models is generally high, with in excess of 50% of variation being explained, regardless of imputation method chosen.

Table 5.1 provides the estimates the most general model considered in this study,

\[ \text{Note that for the ASL's, the unimputed (or listwise cleansed data) uses } n = 35, \text{ while for the imputed data } n = 50 \text{ for determination of the critical value of the test statistic } (t(z^*)). \]

\[ \text{The bias correction takes the mean value of the bootstrap replications to be the true population estimate, as suggested by Efron & Tibshirani (1993). Alternatives exist such as maximum-density bias correction however these are not explored within this thesis.} \]
implying that trip rates are a function of site characteristics, socio-economic and demographic information and also spatial placement. The results are clearly not particularly robust across the alternative imputation methods used, with the majority of variables producing elasticities that differ not only in magnitude, but also in sign. For instance, the elasticity estimates upon gross floor area lie within the range \([-0.427,0.689]\), although it must be recognised that none of the estimates satisfy standard hypothesis tests.

An interesting feature of the results is the observed difference between the generated results of specifications (f) and (g) for models I and III. The general formulation of these imputation methods means that they should generate identical
### Table 5.2: Spec II results: $T_{m,o,d} = f(\delta, \gamma)$

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnGFA</td>
<td>-0.198</td>
<td>-0.323</td>
<td>n/a</td>
<td>-0.248</td>
<td>-0.289</td>
<td>-0.250</td>
<td>-0.250</td>
</tr>
<tr>
<td>lnSEMP</td>
<td>0.321</td>
<td>0.457**</td>
<td>n/a</td>
<td>0.336</td>
<td>0.372</td>
<td>0.349**</td>
<td>0.349**</td>
</tr>
<tr>
<td>lnPARK</td>
<td>0.703***</td>
<td>0.492***</td>
<td>n/a</td>
<td>0.580*</td>
<td>0.571**</td>
<td>0.722***</td>
<td>0.722***</td>
</tr>
<tr>
<td>lnCAR</td>
<td>0.092</td>
<td>-0.438</td>
<td>n/a</td>
<td>-1.124**</td>
<td>-0.482</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td>lnPSV</td>
<td>0.063</td>
<td>0.096*</td>
<td>n/a</td>
<td>0.245***</td>
<td>0.063</td>
<td>0.065**</td>
<td>0.065**</td>
</tr>
<tr>
<td>lnAVHS</td>
<td>2.372*</td>
<td>3.862***</td>
<td>n/a</td>
<td>4.177*</td>
<td>2.466</td>
<td>2.325***</td>
<td>2.325***</td>
</tr>
<tr>
<td>lnAEMP</td>
<td>-0.044</td>
<td>-0.055</td>
<td>n/a</td>
<td>-0.144</td>
<td>0.036</td>
<td>-0.024</td>
<td>-0.024</td>
</tr>
<tr>
<td>lnHHEMP</td>
<td>-0.556</td>
<td>-1.456*</td>
<td>n/a</td>
<td>-0.573</td>
<td>-0.673</td>
<td>-0.505</td>
<td>-0.505</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>0.30</td>
<td>0.90</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td>50</td>
<td>0.44</td>
<td>0.72</td>
<td>0.26</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Notes: *** = 1%, ** = 5%, * = 10%

### Table 5.3: Spec III results: $T_{m,o,d} = f(\delta, \lambda)$

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnGPA</td>
<td>0.138***</td>
<td>0.594***</td>
<td>0.444***</td>
<td>0.532***</td>
<td>0.497***</td>
<td>0.779***</td>
<td>0.258</td>
</tr>
<tr>
<td>LU1</td>
<td>1.202</td>
<td>0.194</td>
<td>1.164</td>
<td>1.266</td>
<td>0.761</td>
<td>-1.567***</td>
<td>3.046**</td>
</tr>
<tr>
<td>LU2</td>
<td>0.738</td>
<td>0.631</td>
<td>0.924</td>
<td>1.652</td>
<td>0.739</td>
<td>-0.558</td>
<td>2.775**</td>
</tr>
<tr>
<td>LU3</td>
<td>1.298</td>
<td>0.750</td>
<td>2.122**</td>
<td>1.315</td>
<td>1.643</td>
<td>-0.936*</td>
<td>3.824**</td>
</tr>
<tr>
<td>LU4</td>
<td>1.303</td>
<td>0.813</td>
<td>2.462***</td>
<td>2.038</td>
<td>1.488</td>
<td>-0.880*</td>
<td>3.861**</td>
</tr>
<tr>
<td>LU5</td>
<td>1.230</td>
<td>0.405</td>
<td>1.870*</td>
<td>1.499</td>
<td>1.139</td>
<td>-0.886*</td>
<td>2.998*</td>
</tr>
<tr>
<td>LU6</td>
<td>1.322</td>
<td>0.328</td>
<td>1.958*</td>
<td>1.183</td>
<td>1.463</td>
<td>-1.110*</td>
<td>3.067*</td>
</tr>
<tr>
<td>LU7</td>
<td>0.976</td>
<td>0.417</td>
<td>1.275</td>
<td>1.232</td>
<td>1.120</td>
<td>-0.774*</td>
<td>2.577*</td>
</tr>
<tr>
<td>LU8</td>
<td>1.115</td>
<td>0.382</td>
<td>1.521*</td>
<td>0.855</td>
<td>1.125</td>
<td>-1.009**</td>
<td>2.928*</td>
</tr>
<tr>
<td>LU9</td>
<td>1.158</td>
<td>-0.370</td>
<td>0.831</td>
<td>0.126</td>
<td>0.414</td>
<td>-1.841***</td>
<td>2.316</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>0.44</td>
<td>0.76</td>
<td>0.34</td>
<td>0.53</td>
</tr>
<tr>
<td>50</td>
<td>0.49</td>
<td>0.63</td>
<td>0.36</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Notes: *** = 1%, ** = 5%, * = 10%

### Table 5.4: Spec IV results: $T_{m,o,d} = f(\delta)$

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
<th>(f)</th>
<th>(g)</th>
<th>(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lnGFA</td>
<td>0.648***</td>
<td>0.637***</td>
<td>0.691***</td>
<td>0.632***</td>
<td>0.648***</td>
<td>0.648***</td>
<td>0.642***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>RMSE</th>
<th>$R^2$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>0.60</td>
<td>0.66</td>
<td>0.38</td>
<td>0.4</td>
</tr>
<tr>
<td>50</td>
<td>0.72</td>
<td>0.39</td>
<td>0.55</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Notes: *** = 1%, ** = 5%, * = 10%
results, however since some land zone types don’t have any observations that are not missing, then they drop out of the regression forecast models completely. However such observations remain present when applying the ANCOVA method and as such, when controlling for land zone type, different values are observed in the variance-covariance matrices and as a result the estimated coefficients from each specification is subject to some difference.

The RF approach is more parsimonious having 15 fewer regressors (i.e. the missing value covariates), whereas the ANCOVA approach will not induce systematic bias by omitting any regressors. From a practical perspective the choice of imputation method will be dictated by the proportion of missing data non-missing data, relative also to the number of variables in the model. For example, in the present application, using specification (g) on model I results in 17 degrees of freedom being used for the standard covariates plus a further 15 being consumed for the missing-value covariates. Given there are only 50 observations, this results in only 18 available degrees of freedom for inferential purposes, hence validating the application of the bootstrap\textsuperscript{15} approach to inference. In more severe situations it may come about that the ANCOVA approach consumes all available degrees of freedom and more besides and thus cannot be applied at all.

Prior to expanding upon the policy implications of the results, a model selection process is defined so as to help define the most appropriate model for subsequent discussion.

\textsuperscript{15}The bootstrap process was designed specifically to cope with small samples of data and provide accurate analysis and inference, sometimes with much smaller samples. Hence, the estimation methodology ensures that small sample sizes do not adversely impact upon the results and subsequent conclusions.
5.3.1 Model Cross Comparison: Preferred imputation method and overall model specification.

Model cross-comparison can be done in two dimensions, one of which is easily dealt with using conventional model specification test approaches, and the other not so easily dealt with. The first dimension relates to testing which model (I, II, III or IV) best explains vehicle trip rates given the data available? There exist multiple tests for evaluating this type of question, four of which will be applied here, namely the Root Mean Squared Error (RMSE), adjusted R-squared ($\bar{R}^2$), the Akaike information criterion and Schwartz-Bayes information criterion tests ($AIC$ and $BIC$). These tests are well expounded in the literature, see for example Greene (2003) or Diebold (2004) and respectively take the following forms\(^{16}\):

\begin{align}
RMSE &= \sqrt{\frac{1}{n} \sum_{i} (y_i - \hat{y}_i)^2} \\
\bar{R}^2 &= 1 - \frac{n}{n-k} \left( \frac{\mathbf{e}'\mathbf{e}}{\sum_{i=1}^{n} (y_i)^2} \right) \\
AIC(K) &= s_y^2 (1 - R^2) e^{2K/n} \\
BIC(K) &= s_y^2 (1 - R^2) n^{K/n}
\end{align}

The second dimension by which these models can be compared is with respect to the appropriateness of the alternative imputation methods ((a),..., (h)), however

\(^{16}\)With the exception to an amendment of the definition of $R^2$ which can be expressed as

$$R^2 = 1 - \frac{\hat{e}'\hat{e}}{y'y}$$

for a model with no constant. see for example Judge, Griffiths, Hill & Lee (1980). This ammended definition is carried through to the definitions of $\bar{R}^2$, $AIC$ and $BIC$. 


this is somewhat less straightforward. There exist tests for comparing ‘non-nested’
model structures against each other, such as encompassing tests or Davidson and
Mackinnon’s J-test, see for example Davidson & Mackinnon (2004). These tests
however seek to reveal the preferred model specification given a consistent mod-
elling technology but with mutually exclusive sets of explanatory variables. Neither
these non-nested tests nor the model selection criteria discussed above offer a suf-
ficiently acceptable framework to compare the model specifications in this second
dimension. This arises due to the reliance of these methods to select the model which
best fits the real data, sometimes controlling for parsimony on the model structure.
However in the instance of the present application, the imputation methods intro-
duce substantial ambiguity which makes the relevance of these tests questionable.
When applying the alternative imputation methods it is not possible to be certain
how close the predicted value for a missing variable is to the actual value.

The results of the tests are summarised in Figure 5.2, the cells highlighted in
grey denote preferred model given each specification, though given the discussion
above provide no indication on the preferred specification. These results suggest
that there is an overall efficiency in model specification made by including site spe-
cific attributes and land-zone identifiers, i.e. model I. This conclusion is qualified
through the generally lower RMSE values and higher $R^2$ values, however this model
is not strictly preferred based on these criteria across the alternative specifications.
The AIC and BIC tests provide conflicting results in general, though the pattern
points towards a preference of model II over III. Particularly, the BIC tests strictly
prefer model II over III, suggesting that the lack of available degrees of freedom is
significantly inhibiting the ability to test the hypothesis that land-zone placement
affects vehicle trip rates. For the majority of alternative imputation methods, it
is seen that model I is outperformed by model II. Though the overriding emergent
story is that no model can be considered strictly preferred based on these results.

![Figure 5.2: MaxiMin tables of results](image)

Constraining analysis to only the models estimated using specification (a), i.e. only estimates based entirely upon observed data, a consistent result appears. Given the four model selection criteria considered, model I is preferred in every situation to the other models, hence it is considered to be the preferred specification given the data available.

The absence of a validation process precludes the ability to directly compare the relevance of the alternative imputation approaches. It is therefore, arguably, more
pertinent to consider the sensitivity of the results across the alternative specifications ((a),..., (h)). In this particular empirical application it is easy to observe a high degree of diversity in the results generated. Though it will be shown that even in instances where (at least under visual inspection) the results are robust to the choice of imputation method, the subtle variation in the estimated model parameters can have potentially significant policy implications.

By way of illustration a sensitivity analysis is presented using the results arising from models III and IV\textsuperscript{17}, illustrated in Figure 5.3. The two most extreme estimates are taken from each of these specifications, i.e. the two (significant) estimates of \( \delta \) with the greatest distance between them. This figure therefore indicates the range of predicted trip rates when GFA is adjusted from the minimum value observed in the dataset, through to the maximum using a consistent estimation and inference methodology, but competing imputation methodologies. It thus exemplifies the situation where, what may at first appear as an almost negligible difference in parameter estimates, can in fact have more severe consequences when applied. The associated parameters are identified in the Figure and the \textit{ceteris paribus} impact upon trip rates at the uppermost extreme point of observed site sizes are as follows for the specifications considered; (i) III(a)=4.60, (ii) III(f)=5492.10 [specification III range\textsuperscript{max}=5487.50 trips, indicated by dashed region], (iii) IV(d)=2076.24, (iv) IV(e)=1081.53 [specification I range\textsuperscript{max}=994.71 trips, indicated by dotted region].

This type of figure provides useful insights for development control officers and regional planners in that it instructively reveals the relative weight in the trip attraction force of mean sized developments as well as developments at the extreme

\textsuperscript{17}These specifications are chosen as they provide consistently significant representations of the variables in question.
points of the observed distribution. Initial consideration of the results would imply that for model IV they are generally quite robust, particularly with respect to all specifications other than the WC. However given the lack of knowledge of the exact nature of the unknown observation(s) and hence which imputation method generates the most realistic results, it is pertinent to observe the range of predicted results generated by the two most extreme methods. Unsurprisingly, the WC method generates the most extreme results with the highest recorded coefficient on the GFA variable of $\beta = 0.691$. With respect to the other model specifications, the story becomes a little less encouraging with substantially larger disparity being observed in the parameter estimates arising from the application of alternative imputation methods.

**5.4 Concluding Remarks.**

From the results presented within this chapter it can be seen that the most effective policy intervention would involve the reduction of parking provision. This must be supported with some caveats, as to reduce parking provision indiscriminately may create more problems than it remedies. Such an approach to travel demand management, suitably constrains the ability to use car and hence directly reduce demand at the site. However, if there are no alternative modes of transport available to reach the office in a reasonable fashion, then the substitution away from car use may come at a high cost to the individual employee. If this cost is too high it may subsequently encourage the employee to seek employment elsewhere and may subsequently impact upon the productivity of the firm. The presence of parking facilities near the workplace, but not managed by the workplace might make policy instruments based on parking provision blunt and potentially completely ineffective.
Figures 5.3: Sensitivity of trip rate estimates to changes in imputation method used

This is particularly important if one considers that the individuals decision to use the car for a specific trip can be considered a rational, utility-maximizing choice over a set of alternatives (see for instance Jara-Diaz (1998)). Any travel demand management policy must be tailored to suit the needs of the individual site and its employees, therefore resulting in an infinite number of policy responses. One solution might be a wider scale implementation of teleworking schemes, which have been seen to be successful in Lyons (1998) and the various evidence collected by Cairns et al. (2004b). On a more general note, wider scale expansion of mass transit systems and encouragement of incentive driven car share schemes may also provide significant shifts in travel behaviour as is alluded to in the Eddington (2006).
evidence of the models presented in this chapter suggest that there is significant scope to foster public service as a means of commuting to work, as they are not found to be significant. Although this phenomenon is specific to the sample, given that the sample contains 50 individual sites, this suggests potentially quite damning evidence on the existence of a credible opportunity for employees to shift transport mode away from the car for the journey to work at present.

From a technical perspective, this chapter reiterates the findings of Brownstone (1998), in that the ideal response to attrition in datasets is to enforce a more rigorous data collection and recording process which will negate the presence of erroneous and/or missing data. In the absence of an appropriate validation process, which will most likely be the case otherwise imputation would not be necessary, further exploration of the power of alternative approaches on alternative data structures is necessary.

As discussed, model cross comparison is problematic for several reasons. There is no natural framework by which to compare the alternative imputation methods, as in the absence of real data, it is not possible to know how close imputed values are to the 'true' missing values. It is feasible that the most simple methods applied here provides final estimated coefficients that are closer to the true population parameters than the most complex methods can achieve, and vice versa.

With the exception of casewise deletion, application of any imputation method

\(^\text{18}\)i.e. theory would imply that different modes of transport should be substitute goods as they fulfil the same general service. However the range of characteristics offered by public transport services result in them not being effectively used. If the undesirable characteristics are attended to, then the theory would lead to the conclusion that real (and significant) substitution can be realised.
generally means making up numbers. There is no ability to validate the accuracy of this data which raises concern over the ability to pick the most accurate/appropriate model when applying such methods. As can be seen from the results table, the choice of method can result in different model structures being preferred and subsequently different policy implications being advocated. Therefore, in order to retain practical relevance, the remainder of this these applies casewise deletion methods to its estimation datasets in order to ensure that policy discussion relates to reality, rather than to a chosen assumption of reality.

The observed diversity in the estimated parameters across the different specifications reveal the lack of robustness in the results. It is also observed that the majority of regressors are insignificant within the preferred model. The approach to estimation confirms the validity of their place within the model, and can be considered accurate even with the limited sample sizes. Their individual insignificance, but joint significance, may be hinting towards a more complex land-zone/transport relationship for office blocks than can be realistically explored with the available sample. There could for instance be interactive (non-linear) relationships, such as land-zone type interacted with public transport services, or public transport services interacted with socio-demographic characteristics. However to implement such features within a general specification will consume more degrees of freedom than available.

Although not the preferred model in this application, model II seems as if it could have some potential. It would be an interesting exercise to re-test this model formulation with more data to see whether model I still outperforms the alternatives. This would be useful from a policy perspective, as if model II proves better specified given new data the the implication would be that there exist activity based economies of scale with respect to transport, at the household level. This would potentially instigate a very different set of policy initiatives as opposed to focussing
purely on parking provision. However given that this model is outperformed in the present application, discussion of scale economies is saved for the next chapter in which they feature significantly within the model.
Chapter 6

Trip Generation Models at Food Superstores

6.1 Introduction

The growth and spatial placement of food superstores, alongside society’s increasing reliance upon them, are the subject of considerable controversy and debate in the UK and elsewhere (see, for example, Yim (1992), Clarke, Horita & Mackaness (2000), or Smith & Sanchez (2003)). This issue has highlighted concerns relating to the encouragement of traffic growth on local road networks (with all its attendant negative externalities of emissions, congestion, higher accident rates etc.) and the deleterious effects food superstores may have for the trading vitality, and the continuing investment in physical fabric and vibrancy of traditional inner urban shopping centres. The food superstore phenomenon has also been accused of being a central element in the development of ‘food deserts’ Wrigley (2002) in some parts of the UK, i.e., areas where access to food shops is difficult for low income households and where there is a lack of small retail shops to meet the demand for healthy, affordable food. Thus, the problems identified with food superstores relate both to their role
in traffic generation and to their role in the development of a less sustainable urban spatial structure. Arguably a key element of such a structure is marked by increasing concentration of food provision towards food superstores, since smaller retailers may often not be able to compete on cost grounds for mobile consumers (a significant fraction of which are needed in order to retain economically viable retail customer bases). Moreover, such concentration may perpetuate less sustainable lifestyles and also militate against transport-poor, lower income households, who become much more reliant on higher cost, less healthy food from nearby 'convenience' stores (see Wrigley (2002) and Wrigley, Warm, Margetts & Whelan (2002) for a detailed consideration of the 'food deserts' issue and some policy responses).

This analysis\(^1\) contributes to the body of evidence supporting subsequent policy decisions in the controversies and debates by focussing on the estimation of a trip attraction model for food superstores in the United Kingdom, utilizing, for the first time in this context, the data held within the Trip Rate Information Computer System (TRICS) database. TRICS data from the period 1986-2003 is augmented

\(^1\)Some of the results contained within this chapter may be found in;


with the UK Census and other data sources, providing a rich dataset from which the model estimates are generated. To date (as far as is known) the TRICS data has not been previously combined with these other data sources for analysis beyond the arena of local planning and development control.

The analysis in this chapter operates in a general transport and development framework that is broadly consistent with a 'predict and manage' strategy as opposed to the 'predict and provide' approach of historic notoriety (see Berry (1960)) or the more contemporary focus on 'Transport Demand Management' (TDM) solutions (see for example, Meyer (1997), Meyer (1999)). Given that much of the literature generally neglects the potential synergies between these schools of thought, an eclectic approach is taken here. Trip attraction models have been developed for various land-uses, including food superstores and similar enterprises elsewhere (see, for example, Vickerman & Barmby (1984), Goldner & Potugal (2002), Tan & Fan (2003)) but being conscious of the fact that the growth of such retail entities may well, in part, be a product of the diffusion of a more car-orientated culture and that the relationship between land use, spatial structure and mobility is undoubtedly complex and varied (see, for example Mackett (1993), Badoe & Miller (2000) or Meurs & Haaijer (2001)).

The chapter is organised as follows. In the next section the modelling strategy developed for this application is set out and the data is described in the following section. Section 4 presents and discusses the results, with the final section offering a summary and concluding remarks.
6.2 Modelling Strategy and Estimation

This chapter applies the general modelling framework (detailed in Chapter 4 in its most literal sense. A general to specific approach is adopted, though only the results of the preferred model are presented. However en route to achieving the preferred model, the analysis attempts to define with trip generation models at found superstores is best founded by retail floor space or total floor space. Further to this alternative socio-economic classifications are considered in order to try and represent different household classifications in the area surrounding the site. Due to the way the variables are constructed, the resulting multicollinearity precludes the useful inclusion of more than one of these population classification variables.

Previous literature offers mixed views as to which is the appropriate floor-space measure to use (gross or retail), mainly surrounding the relevance of the use of Gross floor area which includes warehouse space. Retail floor area, on the other hand, arguably captures more directly the area of business operations (in terms of retail sites) that customers come into contact with. Ortuzar & Willumsen (1994)(pp97-98) provide arguments in support of both approaches. Dasgupta, Raha & Sharman (1996) and Goldner & Potugal (2002) however, restrict their analysis to gross floor area. Tan & Fan (2003) in a study of peak hour trip rates to office and retail developments find mixed evidence mostly in favour of gross floor space. Given both the gross and retail are available, both are used in the estimation, denoted by models I and II.

The negative coefficient on the SOCCOCON* variables may not be so intuitively clear without some clarification. The expectation is that household levels of economies of scale and scope exist such that “the cost per person of maintaining a given material standard of living may fall as household size rises” (Nelson 1988). Following Lazear & Michael (1980), these economies arise due to the nature of cer-
tain goods used within the physical confines of the house, as certain goods cannot avoid being ‘public goods’, i.e. goods that provide benefits to everybody (positive externalities), not just those who purchase them. Lazear & Michael (1980), term these goods ‘family goods’. Examples include lighting in hallways and locks on shared entrances. It is not feasible to exclude people from the use of such goods.

Such economies of scale in a non-nuclear household could be considered as economies of scope. These arise when two independent parties (i.e. two separate individuals living under one roof, as in shared accommodation for professionals/students etc.) are able to pool together their resources and reduce the marginal cost faced by each in achieving the same level of utility. Lazear & Michael (1980) present an example-using door locks, though characterise this purely as a scale economy. Yet this scale economy is only realised through the recognition of scope. Thus in the case of trip making behaviour, two individuals within a shared house could car share and thus utilise space that may previously have gone un-used (assuming individual shopping load constraints are not biting). This arises at a cost which is lower to each individual (presuming costs are shared). Such arrangements could still provide each individual with the same amount of car-borne shopping resources that they would have enjoyed had they not pooled together.

Nelson (1988) places an empirical value on the household economies of scale that were achieved in the US for food shopping. The results revealed with a strong degree of significance, that for households choosing to pool their resources 2 people can essentially live for the price of 1.19 people. The concept and existence of scale and scope economies would arise automatically with growing nuclear households, and would be a conscious decision within non-nuclear multiple person households. Following this reasoning, it is contended that the number of trips to a food store
are negatively related to household size\textsuperscript{2}. Further, larger households will also likely exhibit more diverse characteristics in their modal choices, i.e. as household size increases, the probability that one of the household members will prefer a non-car mode of travel to a food store also increases.

In relation to the variable \textit{SOCIOECON*} from equation (2), three alternative measures (average household size (AVHS), average large households (AVLH) and average households with children (AVHC)) are considered in the empirical phase, that give differing estimates for policy variables based on representations of the surrounding area's demographic decomposition. These are represented by a, b and c respectively in the result section.

6.3 Results

Given the two floor space variables and the three representations of the area's socio-economic characteristics, six separate models were estimated. The specification tests for these initial estimates revealed non-normality and heteroskedasticity, thus making it problematic to make accurate statistical inference from the results, given the possible biased standard errors, rendering standard t-tests ambiguous. As a result, a bootstrap algorithm (see the Appendix) was applied, essentially rebuilding the error structure and enabling accurate statistical inference. The bootstrap algorithm for standard errors is a non-parametric framework, since it is derived from the non-parametric estimation of the sample distribution. Thus, the combination of the

\textsuperscript{2}Furthermore, regardless of whether or not parking restrictions exist in a residential area (which in many, particularly urban, cases they do), there is typically finitely constrained space to store/park a car. As household size increases such spatial constraints may ultimately generate a stronger barrier to increasing car trips that could otherwise have been expected to emerge as a direct consequence of multiple car ownership.
standard OLS parametric inference using the non-parametric based error distributions, results in a semi-parametric regression model. Based on this semi-parametric approach, the coefficient of the variable \textit{YEAR} was always not significantly different from zero and hence omitted.

The results given in Tables 6.1 and 6.2 are in accordance with a-priori expectations, though the coefficient on the public transport accessibility variable warrants further discussion. In terms of the subsequent discussion and analysis, focus remains upon model ‘I(a)’, as this model reflects the ‘average’ household in the UK, and the floor-space measure which results in a better specified model (in terms of R-squared and Root Mean Squared Error (RMSE)). Interpreting the results in the presence of dummies requires a degree of care, base interpretation should initially be done where all dummy variables are set equal to zero. Following this, base interpretation then identifies the level of average hourly passenger-vehicle trips on a Friday, attracted to a ‘standard’ food superstore without a petrol filling station, positioned in a town centre zone in the UK.

The public transport variable provides what may at first be considered a counterintuitive result. A positive coefficient on this variable is observed, implying that as public service provision increases, so do trips to that site by car. Of course, the positive value may simply identify a correlation between bus provision and large business centres. Large superstores may attract higher rates of public service provision as the providers realise that the large superstores are likely to offer a larger (and more profitable) customer base than the smaller stores. If bus services were indeed being sucked towards more profitable food superstore sites and away from serving other more traditional shopping areas, then this might be considered as contributing to the development of ‘food deserts’.
The rest of the coefficients follow standard micro-economic demand theory, and it is observed that floor space, distance to the nearest food superstore competition, household type and size, parking provision and the inclusion of a petrol filling station at a site all return positive coefficients. These variables generally reflect the ability to substitute the store for alternative shopping centres, and the costs involved with doing so.

As stated above, a date variable was employed in initial specifications in an attempt to capture any natural growth effects not represented by the other variables in the model. However they were insignificant in each run, suggesting that, although over recent years there has been increasing growth in the number of superstores being used by food retailers (as opposed to smaller store sizes), this growth in floor size has not created any shifts in the fundamental behaviour of shoppers. This is intuitive if one considers the inelastic nature of food products; it is unlikely that extra income will encourage an individual to buy two packs of turkey meat. Rather, it is more likely that it will cause the individual to buy chicken instead of turkey. i.e. increased wealth alters your consumption bundle away from inferior products, though not necessarily in favour of more goods. Therefore although increased wealth may be coupled with a more opulent lifestyle, no extra travel is required to facilitate this. Increased wealth only means increased consumption in terms of utility, not necessarily physical quantity, at least that is, in buying the weekly food shop.

The results are observed to be extremely robust across the different specifications, with the only real notable change in the estimated parameters being those for the alternative measures of site area's demography. These observed changes are intuitively consistent given the qualitative explanation of each; the measure AVHS
Table 6.1: GFA Model Estimates

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.539*</td>
<td>-0.633</td>
<td>-0.949</td>
</tr>
<tr>
<td>lnCAR</td>
<td>0.202***</td>
<td>0.199***</td>
<td>0.166**</td>
</tr>
<tr>
<td>PSV</td>
<td>0.141***</td>
<td>0.129**</td>
<td>0.144***</td>
</tr>
<tr>
<td>lnGFA</td>
<td>0.303**</td>
<td>0.320**</td>
<td>0.355***</td>
</tr>
<tr>
<td>lnRESISTANCE</td>
<td>0.085***</td>
<td>0.083***</td>
<td>0.088***</td>
</tr>
<tr>
<td>lnAVHS</td>
<td>-1.724***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnAVLH</td>
<td></td>
<td>-0.613***</td>
<td></td>
</tr>
<tr>
<td>lnAVHC</td>
<td></td>
<td></td>
<td>-0.555***</td>
</tr>
<tr>
<td>lnHHEMP</td>
<td>0.494***</td>
<td>0.402***</td>
<td>0.454***</td>
</tr>
<tr>
<td>lnPARK</td>
<td>0.338***</td>
<td>0.327***</td>
<td>0.306***</td>
</tr>
<tr>
<td>PFS</td>
<td>0.136***</td>
<td>0.130***</td>
<td>0.141***</td>
</tr>
<tr>
<td>LU2</td>
<td>0.153*</td>
<td>0.158*</td>
<td>0.160*</td>
</tr>
<tr>
<td>LU3</td>
<td>0.214***</td>
<td>0.207***</td>
<td>0.222***</td>
</tr>
<tr>
<td>LU4</td>
<td>0.102</td>
<td>0.109</td>
<td>0.105</td>
</tr>
<tr>
<td>LU5</td>
<td>0.280***</td>
<td>0.266***</td>
<td>0.260***</td>
</tr>
<tr>
<td>LU6</td>
<td>0.334***</td>
<td>0.317***</td>
<td>0.352***</td>
</tr>
<tr>
<td>LU7</td>
<td>0.109</td>
<td>0.080</td>
<td>0.124</td>
</tr>
<tr>
<td>LU9</td>
<td>-0.067</td>
<td>-0.050</td>
<td>-0.052</td>
</tr>
<tr>
<td>ECONOMY</td>
<td>-0.341***</td>
<td>-0.329***</td>
<td>-0.318***</td>
</tr>
<tr>
<td>SAT</td>
<td>0.029</td>
<td>0.032</td>
<td>0.030</td>
</tr>
<tr>
<td>SUN</td>
<td>-0.658***</td>
<td>-0.651***</td>
<td>-0.652***</td>
</tr>
<tr>
<td>MON-THURS</td>
<td>-0.258***</td>
<td>-0.256***</td>
<td>-0.257***</td>
</tr>
<tr>
<td>Obs</td>
<td>201</td>
<td>201</td>
<td>201</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.276</td>
<td>0.276</td>
<td>0.277</td>
</tr>
<tr>
<td>$\bar{R}$²</td>
<td>0.694</td>
<td>0.693</td>
<td>0.691</td>
</tr>
</tbody>
</table>

Notes: *** = 1%, ** = 5%, * = 10%
Table 6.2: RFA Model Estimates

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.230</td>
<td>-0.665</td>
<td>-0.883</td>
</tr>
<tr>
<td>lnCAR</td>
<td>0.214***</td>
<td>0.208***</td>
<td>0.179***</td>
</tr>
<tr>
<td>lnPSV</td>
<td>0.143***</td>
<td>0.135***</td>
<td>0.149***</td>
</tr>
<tr>
<td>lnRFA</td>
<td>0.339**</td>
<td>0.345**</td>
<td>0.373***</td>
</tr>
<tr>
<td>lnRESISTANCE</td>
<td>0.102***</td>
<td>0.101***</td>
<td>0.107***</td>
</tr>
<tr>
<td>lnAVHS</td>
<td>-1.557***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnAVLH</td>
<td></td>
<td>-0.534***</td>
<td></td>
</tr>
<tr>
<td>lnAVHC</td>
<td></td>
<td></td>
<td>-0.458***</td>
</tr>
<tr>
<td>lnHHEMP</td>
<td>0.432***</td>
<td>0.347***</td>
<td>0.384***</td>
</tr>
<tr>
<td>lnPARK</td>
<td>0.343***</td>
<td>0.337***</td>
<td>0.321***</td>
</tr>
<tr>
<td>PFS</td>
<td>0.150***</td>
<td>0.146***</td>
<td>0.155***</td>
</tr>
<tr>
<td>LU2</td>
<td>0.161*</td>
<td>0.168**</td>
<td>0.172**</td>
</tr>
<tr>
<td>LU3</td>
<td>0.172**</td>
<td>0.167**</td>
<td>0.178**</td>
</tr>
<tr>
<td>LU4</td>
<td>0.096</td>
<td>0.105</td>
<td>0.103</td>
</tr>
<tr>
<td>LU5</td>
<td>0.241**</td>
<td>0.230**</td>
<td>0.225**</td>
</tr>
<tr>
<td>LU6</td>
<td>0.375***</td>
<td>0.366***</td>
<td>0.427***</td>
</tr>
<tr>
<td>LU7</td>
<td>0.096</td>
<td>0.077</td>
<td>0.120</td>
</tr>
<tr>
<td>LU9</td>
<td>-0.019</td>
<td>-0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>ECONOMY</td>
<td>-0.315***</td>
<td>-0.307***</td>
<td>-0.298**</td>
</tr>
<tr>
<td>SAT</td>
<td>0.032</td>
<td>0.034</td>
<td>0.032</td>
</tr>
<tr>
<td>SUN</td>
<td>-0.656***</td>
<td>-0.650***</td>
<td>-0.650***</td>
</tr>
<tr>
<td>MON-THURS</td>
<td>-0.251***</td>
<td>-0.247***</td>
<td>-0.246***</td>
</tr>
</tbody>
</table>

|       |       |       |
| Obs    | 199   | 199   |
| RMSE   | 0.276 | 0.277 | 0.278 |
| $\bar{R}^2$ | 0.689 | 0.687 | 0.684 |

Notes: *** = 1%, ** = 5%, * = 10%
is seen to have a coefficient of a greater magnitude than both AVLH and AVHC in absolute terms. The measures of AVLH and AVHC essentially both capture households which have greater restrictions imposed upon their time, where the lifestyles may be more typified by fewer trips to a food superstore for larger grocery needs (i.e., the family monthly food run). This applies as much to large households as it does to households with children, as the family routines may be continued into young adult life within the family home, alternatively shared houses may choose to car-share, so as to reduce their costs (and perhaps make shopping a semi-social event\(^\text{3}\)).

The stability of the coefficients under changes in (qualitative) model specifications is encouraging and the robustness suggests that the general theoretical framework is a good reflection of the observed relationships for vehicle trip-making behaviour at food retail sites.

Bonsall, Champernowne, Mason & Wilson (1977), undertake a sensitivity analysis on key significant variables in order to help identify the most effective policies, based on the reality of their application. This is based on the understanding that as the variables were found to be significant, the results of policy changes can be expected to have significant effects too. Two types of parameter are considered; firstly are those normally considered fixed in nature (e.g., average car ownership within an area), the second type of parameter represents more feasible (in the short to medium term) policy alternatives that can be enacted upon in a reasonable time frame, for example changes in the level of public transport provision. It is hard to conceive that any of the variables analysed in this study must be fixed in the

\(^{3}\)It has indeed been suggested that the supermarket is one of the best places to meet potential partners (see for example Cameron & Collins (2000).
long-run, though it is still prudent to consider that some of the variables included in this analysis will be 'more fixed' than others.

The following sensitivity analysis is focused primarily on model 'I(a)' only, given the floor-space measure GFA produces marginally better R-squared and lower RMSE values than the RFA counterpart specifications, thus model 'I' is preferred over model 'II' as the measure AVHS most accurately reflects the average individual in the area surrounding each site, AVLH and AVHC reflect more specific segments of the community. Hence, model I(a) is more appropriate in a 'general' discussion.

It should be noted that the results provide some response to the debate over which is the appropriate floor-space measure to use in the trip attraction models. The selection criteria discussed in the previous paragraph reveal that for food supermarkets, GFA is favoured. This suggests that, at least in terms of food shops, the attractiveness of a site is determined by their ability to meet demand, not just the goods they sell. Using RFA as a determinant reduces the explanatory power of the model by approximately 0.5 percentage points (i.e. R-squared for I(a) = 0.6943, R-squared for II(a) = 0.6895) in all cases. This change is not large enough to provide irrevocable evidence, but does imply that consumers consider the warehouse and administrative space in their trip making decision. This finding is consistent with that of (Tan & Fan 2003) who find statistical evidence favouring the use of GFA at retail sites.

The sensitivity analysis in figure 2 is based upon the fitted equation:

\[ \ln FLOW_i = \hat{\beta}_0 + \hat{\beta}_\lambda X_{i\lambda} + \hat{\beta}_\eta \bar{X}_{i\eta} \] (6.1)

With all independent variables fixed at the mean (i.e. all \( X_{\eta} \), where \( \lambda \neq \eta \)),
other than the policy variable under consideration (i.e. $X_A$). From this the predicted average hourly trip value ($\ln FLOW_i$) is computed (for model I(a)). This is evaluated at regular intervals on the difference between the highest and lowest observed values for $X_n$ within the dataset. i.e.

$$z(X_{n,\text{max}} - X_{n,\text{min}})$$

Where $z$ is the evaluating parameter and ranges between zero and one at regular intervals (100 in this case, or 1% intervals of the difference).

Figure 6.1 identifies several important features with respect to the implementation of policies (for the average household) based on model I(a). All relationships in this figure are non-linear and convex (as a result of the log-linear model) implying that a) as the actual value of the policy variables depicted in this graph decrease, so do the level of vehicle trips, b) the marginal effects of these policies decreases as actual traffic flows move towards the origin. This is true for all policy variables given in Figure 2, but most prominent for the PARKING variable. Analysing the results in this manner adds a new light to the estimated model, suggesting that in terms of feasible application, the most effective way of reducing vehicle trips to a supermarket is to decrease parking provision. The estimated model would suggest (via initial inspection of the coefficients) that the variable GFA is almost equally an affective policy variable as PARKING is. However the variability in the dataset (assuming this represents acceptable bounds from which policy makers can affect changes), implies that regardless of its equally large elasticity, imposing restrictions on floor-space may not be the most realistic route for the policy maker to take in

---

4 RESISTANCE is the hashed line following a similar path to EMP.

5 Even though the focus is on model I(a), it is worth noting that the qualitative results are the same for I(b), I(c), II(a), II(b) and II(c).
reducing vehicle trip levels.

Figure 6.1: Effects of Key Variable Changes upon Food Superstore Traffic Flow

Figure 6.2 extends the sensitivity analysis to include the effects of policy variables on each of the qualitatively different models I(a), I(b) and I(c) (though subsequently only considering the constant elasticities). Consistent relationship directions are observed, with only very subtle changes in magnitude relative to the ‘base’ model I(a). Changes in elasticities across the three models provide insights into the relative
importance of policy measures on large households and households with children (compared to the average household). Policy makers and setters could potentially use this information to focus policy implementations at specific demographic groups. This graph exemplifies the robustness of the results (refer back to Tables 6.1 and 6.2), and reveals that in terms of policy based discussion, there is no justifiable rationale to prefer models I(b) or I(C) over I(a) or to think (for food superstores) that policy decisions can effectively be focussed on particular socio-economic groups. The horizontal axis on this graph is the natural log of the average hourly flow, whilst the vertical axis identifies the policy measure considered.

The three variables with the strongest impact upon traffic flow were identified as household size/composition, parking provision and floor-space. The results indicate that decreasing the number of parking spaces will reduce the trips to a food superstore by car. The graph also highlights the relationship observed that smaller households seem to make more trips to 'smaller' food superstores (or at least a 1% change in GFA attracts more trips from large households and households with children than it does from the average household).

Turning to the public transport provision variable, discussing policy with respect to this variable is overlaid by a desire to encourage increased levels of public transport use. The results suggest on a naive face value that less provision for such services might be preferred as it is associated with lower car traffic levels. Perhaps a more reasonable explanation might simply be that public transport service providers are not providing adequate service levels at the origin end of the food superstore

6However policies surrounding parking management need to be treated with care as, in the absence of viable alternative modes of transport, this could simply redirect trips to the next nearest alternative. i.e. site specific traffic will reduce, but network total traffic will remain constant or may even increase.
journey. If they were providing the level of service at the origin that people really wished to have, then one may find that the coefficient on this variable decreases and maybe could change sign altogether.

A further key advantage of the TRICS database is the ability to explore in greater depth the role of land-zone type, that is not commonly presented in the literature. Banister et al. (1990) identify, among other things, that further empirical research is required on this issue, in short, this study addresses that identified concern to some extent. Figure 6.3 shows the effects of land-zone placement upon trip rates to
a mainstream food superstore (on a Friday) without a petrol filling station, when all parameters are held at their means.

Figure 6.3 therefore identifies the perceived level of accessibility for each of the demographic groups considered in this study. As with the policy variables, the elasticities on accessibility remain stable across each of the models (in terms of direction and magnitude), suggesting that fundamental relationships hold true across all representations of the population considered. Households with children have a preference towards food superstores in industrial and commercial zones, whilst large households in general (including shared houses) prefer them less, at least in relation to trips by
car, when compared to the average household. It should be noted that land uses 2 and 3 were found to be statistically insignificant and consequently take values of zero.

6.4 Food Superstores and 'Food Deserts'

The definition of a Food Desert is highly subjective albeit the fundamental tenets are broadly agreed upon; see for example Cummins & Macintyre (2002), Wrigley (2002) and Shaw (2006). The results given in Table 4 have a number of implications relating to the way in which the average food superstore in the UK implicitly contributes to the Food Desert problem (though it is perhaps more pertinent to suggest that the activity purpose of grocery shopping is simply not consistent with current idealisms of sustainable transport). Dietary health and food access remain predominant political concerns within a changing society (see Dibb (2005), O’Neill (2005) and Wilson, Alexander & Lumbers (2004)), and the desire to shift economic activity towards an environmentally and economically sustainable growth path remains testament to this.

The results of the trip attraction model may be construed as implying that discount/economy food superstores are more conducive (in relative, not absolute terms) to sustainable development, as they are associated with 0.3% fewer trips by car relative to a non-economy food superstore. However, such stores may also be associated with lower quality foods, therefore leaving urban planners with a trade-off between the negative externalities caused by road traffic and the dietary health effects upon the general community of providing access to a lower quality of food.

The negative coefficient on the ACCESSIBILITY variable indicates that pub-
lic service provisions at the sites considered in the study are not able to satisfy consumer’s requirements. This result coupled with the positive and highly significant coefficient on the SOCIOECON* variables coincide with the carrying capacity conundrum identified by Shaw (2006) who points out that “'Disposable weight-carrying capacity' may be as important as disposable income for some consumers”, particularly those living in 'less-mobile' communities (p.232). Furthermore as Clarke & Guy (2004) point out, design of food superstores since the 1970’s has been car orientated, where alternative modes of transport for this trip purpose may prove to be disadvantageous. The car satisfies the user’s requirements and its use is rationalised in large households through economies of scale. Concerns relating to social exclusion and access to amenities, as well as developers preferences to achieve development approval make the provision of public transport services at food superstores a desirable feature. However the burden of evidence indicates that provision is not conducive to shoppers requirements, a view which is supported by the findings of Shaw (2006) and that perceived provision may be as influential as actual provision as Kirkup, de Kervenoael, Hallsworth, Clarke, Jackson & del Aguila (2004) show. This result implies that public transport provision is a cost inefficient trip reduction policy intervention for the activity of grocery shopping.

Responses to these highlighted issues and more general Food Deserts concerns are neither straightforward nor obvious. It is easy to suggest that accessibility needs to be made more prevalent in low mobility areas, however as Shaw (2006) identifies, affluent areas (which are by inference more mobile), are not immune to becoming Food Deserts. Firstly it is evident that the transport community must take efforts to create more-sustainable modes of transport which can offer shoppers the specific bundle of characteristics they need to complete a satisfying and successful shopping journey. Therefore such transport must be direct (or almost direct) between origin
and destination ends of the journey, must provide the carrying capacity to satisfy the travellers' requirements and produce fewer negative externalities than existing means of transportation. Achievement of these goals is made most difficult by the prerequisite for the journey to be direct, as chilled and frozen foods should not be allowed to warm, thereby restricting trip times ideally to less than 30 minutes. This in itself makes mass transit a poor choice, as mass transit must primarily serve the aggregate needs of the majority, extending services to more secluded communities as a marginal objective (assuming firms are profit maximisers). As a caveat to this is the recognition that if, for instance, hydrogen fuel cell technology was instantaneously able to replace existing petrol/diesel car technology, thus satisfying the criteria given above (i.e. direct, meets carrying capacity requirements and produce fewer externalities), no impacts would be made upon congestion or accident rates. Therefore mixture of supply-side policies (i.e. store placement policies such as Urban-Regeneration schemes which have been seen to be successful, see Wrigley, Guy & Lowe (2002), Wrigley, Warm & Margetts (2003) or Wrigley, Lowe & Wood (2006)) must be combined with travel-mode specific demand-reduction policies, though the precise interventions must be tailored to the needs of the individual community. Alternative solutions already exist that provide some alleviation of these concerns, such as home delivery (see for example Cairns (1996), though far more consideration is still required. Further, Clarke, Eyre & Guy (2002) and Kirkup et al. (2004) highlight that introduction of an increased number of smaller stores is an obvious solution, though this introduces only one amenity into a previously deprived area and as such may only solve a part of a much larger problem.
6.5 Concluding Remarks

In the light of increasing concern over the external effects arising from traffic growth emanating from food superstore developments and the controversy over the role of such superstores in creating ‘food deserts’ in the UK, this chapter presents a new empirical framework and offers fresh insights. Using traffic count and site specific data derived from the TRICS database, not previously used for econometric estimation and research, a composite dataset was constructed in order to develop and estimate a trip attraction model for food superstores in the UK, applying a bootstrap algorithm to ensure that the inferences drawn in this research are unbiased. Furthermore some sensitivity analysis and basic policy modelling was undertaken on the model to show the effects of changes to key parameters.

It is found that traffic to a given food superstore, ceteris paribus, increases with car ownership, parking provision, retail floor space, distance to the nearest food superstore competitor and, perhaps surprisingly, increased public transport provision. The latter effect is discussed in the light of a possible explanation linked to the ‘food deserts’ debate. Trips by car to a food superstore are seen to decrease as household size is increases.

Shopping in a large food superstore is on the whole a time consuming experience, therefore the decision to engage in such an activity will be consciously influenced and weighed against the size of the shopping baskets that are being filled (i.e. the extent of the grocery needs). The utility derived from the ‘bundle of goods’ purchased at each visit will be weighed against the costs involved in obtaining those goods. Even in the face of current changes to the convenience store market, as a general rule, goods are cheaper in the food superstores than in other outlets for food (due to market power being exerted along the vertical supply chain, economies of
scale and specialisation). Households with large shopping requirements (e.g. large households or households with children) are naturally attracted towards the food superstore. Furthermore, due to their larger shopping load requirements, the relative ease in which the trip can be conducted may be enhanced by the use of a car (due to increased comfort and security, door-to-door service etc).

Hence, food superstores seek to justify their extensive requirements for land in terms of customer parking provision. As city space is at a premium, it may not be feasible for superstores to locate at inner urban locations in a way that meets such business requirements (i.e. to maximise their profits they wish to facilitate customer accessibility which, as the empirical results reveal, are greatly influenced by the level of parking provision). Developers may then be rationally showing a preference towards outer urban areas, where it is easier to satisfy their parking ratio requirements and thereby contributing to the genesis and maintenance of food deserts.

The revealed existence of scale and scope household economies in connection with food superstore trips suggests that communities with larger household sizes will naturally have a preference towards large food superstores. Given the argument above, these stores are increasingly less likely to be placed in inner urban locations. These household scale and scope economies essentially help perpetuate and accentuate food desert concerns, as well as concerns with urban sprawl and poor progress in the development of sustainable communities. Such phenomena are, however, clearly not in the range of realistic direct policy variables that lie within the grasp of urban and transport planners in advanced urban economies.
Chapter 7

Residential Site Car Ownership and Use

7.1 Introduction

As Button (1974) highlights, transport practitioners often rely upon knowledge of car ownership levels and their interaction with travel demand in order to help define future transport behaviour. Through this knowledge planners are subsequently able to manage and develop the transport system as is most appropriate, this is as much true now as it was in 1974 as for instance can be seen from Romilly et al. (2001). Figure 7.1 shows the aggregate level of car ownership and car use (in billions of passenger kilometers) in the UK between 1952-2005. It is clear that there has been an ongoing upward trend in the demand, though with some evidence of reduction in the rate of growth for car use towards the end of the period. Eskeland & Feyzioglu (n.d.) explore in some detail the interactions between car use, car ownership and pollution concerns and the fiscal costs associated with achieving pollution targets. Hence in a similar vein to the work by Eskeland & Feyzioglu (n.d.), this chapter provides an alternative approach to modelling car use and ownership using data.
from the UK in order to consider the feasibility to manage the demand for car travel.

Figure 7.1: UK Car Ownership and Use 1952-2005. (Data source: UK Department for Transport online data collection)

Many previous studies have considered the empirical relationship between car ownership and car use, such as de Jong (1989), Dargay & Giuliano (2005), Boarnet et al. (2003), Boarnet & Sermiento (1998) or Button, Ngoe & Hine (1993). Though equally as many treat car ownership as a separate entity, see for example Tam & Lam (2000), Hess & Ong (2001), Dargay (2002), Romilly et al. (2001) or Ogut (2004). Although many of these studies operate at different levels of aggregation and further apply a range of different methodologies, however they stand testament to the conflicting opinions over whether they should be modelled individually or together.
Joint models of car ownership and car use can help extend the information provided by pure car ownership models (namely the change in the levels of ownership dependent on the input factors) to include information on the change in usage levels which can be compared directly to the level of ownership. Although it is not feasible to consider such interactions directly as a function of activity purpose or trip length, it is possible to consider the total number of trips produced, and differentiate by different household structures. This provides some bearing on the lifestyles of different household types and areas.

A further issue arising in the literature is the relationship between travel behaviour and urban form. Though some of the studies previously mentioned take into account some aspects of urban form in their analysis, more considered examples can be found in Kiss (2001), Brownstone & Golob (2005) or Buliung & Kanaroglou (2006) for instance. A common feature of these particular studies seems to be a stronger emphasis on the health effects associated with the varying transport levels at different location types. In particular it is interesting to see whether or not certain location or dwelling types are empirically associated with increased car use, given that cars are the source of so many air pollutants.

The present chapter therefore seeks to identify whether land zone and residential dwelling type can be statistically related to car ownership levels and also car use. In so doing it will employ a sequential, general to specific modelling strategy as outlined in the following section. This approach is broadly consistent with that employed in the previous two chapters, although is unique in that the most general model is multi-equation, unlike in the previous two chapters. Section 7.3 will subsequently provide empirical response to the debate as to whether or not car ownership should
be modelled independently of car use, or whether it should feature in a system based approach. This and the subsequent conclusion section will draw together the implications of the results and provide some comparison to the previous literature.

7.2 Modelling Strategy and Estimation

This chapter estimates two alternative model structures, namely a single-equation specification (one dependent variable) and a multi-equation approach (two dependent variables). The single-equation approach is adopted for two reasons, firstly for consistency with the other estimations contained within this thesis, and thus to allow for direct comparison. Secondly, by estimating the single-equation models and comparing them to the multi-equation specifications, allowing for scope to determine whether car ownership choices are independent of car use decisions given the data available. For consistency in comparison with the multi-equation specification, models of car ownership and car use are estimated independently of each other. The parametric model specifications are thus as follows;

Car Use (CU)

\[
\ln(T_{m,0,d}) = \mu_d + (\beta_0^*, \beta_\gamma^*, \beta_{\lambda}^*)' (\ln \delta_d, \ln \gamma_{o,d}, \lambda_o, \eta_o) \quad (7.1)
\]

Car Ownership (Co)

\[
\ln(\eta_o) = \mu_d + (\beta_0^*, \beta_\gamma^*, \beta_{\lambda}^*)' (\ln \delta_o, \ln \gamma_{o,d}, \lambda_o) \quad (7.2)
\]

where \(\eta\) is the car ownership variable delineated from the other socio-economic characteristics.
For the multi-equation modelling approach a two-stage least squares (2SLS) estimator is used in order to allow for over-identification in the system. This is borne of a desire not to impose any particular model structure in order to exactly identify the parameters in the system, and also to allow full flexibility in the first stage regression, such that it can reflect the data as is most appropriate. The general multi-equation model is specified as:

**Car Ownership (Stage 1) and Car Use (Stage 2)**

\[
\begin{align*}
\ln(T_{m,o,d}) &= (\beta_5, \beta_7, \beta_8, \beta_9)'(\ln \delta_d, \ln \gamma_{o,d}, \lambda_d, \eta_o) \\
\ln(\eta_o) &= (\beta_5, \beta_7, \beta_8, \beta_9)'(\ln \delta_d, \ln \gamma_{o,d}, \lambda_d, \ln T_o)
\end{align*}
\] (7.3)

The preferred models coming out of this analysis are then estimated via a semi-parametric regressions for consistency with the previous estimations in the thesis, and also to test the robustness of the policy results under a more flexible estimation framework. The semi-parametric counterparts of the above specifications are thus defined as:

**Car Use (CU)**

\[
\ln(T_{m,o,d}) = \mu_d + \sum_{b=1}^{B} [B^{-1} \sum_{b=1}^{B} (\beta_5^*, \beta_7^*, \beta_8^*, \beta_9^*)]'(\ln \delta_o, \ln \gamma_{o,d}, \lambda_o, \eta_o) 
\] (7.4)

**Car Ownership (Co)**

\[
\ln(\eta_o) = \mu_d + \sum_{b=1}^{B} [B^{-1} \sum_{b=1}^{B} (\beta_5^*, \beta_7^*, \beta_8^*)]'(\ln \delta_o, \ln \gamma_{o,d}, \lambda_o) 
\] (7.5)

**Car Ownership (Stage 1) and Car Use (Stage 2)**
CHAPTER 7. RESIDENTIAL SITE CAR OWNERSHIP AND USE

Inference on the parameter in both the one stage and two stage regression models is conducted using the ASL test framework as outlined in Section 4.2.

The testing procedure

The following section outlines the core testing procedure applied to the parametric model results to facilitate the general to specific testing procedure. For the single equation models, Ramsey’s general reset test and the Breusch-Pagan test for heteroskedasticity are applied. See Gujarati (2003) or Greene (2003) for further discussion of these tests. Following this, a standard F-test for linear restrictions is then applied to the remaining model specifications to see if any models nested within the most general specification are statistically more efficient at explaining the data.

Given the estimation methodology outlined, the testing procedure must be flexible enough to choose among nested and non-nested models. This becomes problematic with the IV estimations, as the definition of the error term in the second stage equation, subsequently used for model selection purposes is defined by Gujarati (2003), as:

\[
\hat{u}_{2i} = Y_{2i} - \beta_{20} - \tilde{\beta}_{21}Y_{1i}
\]  

\[(7.7)\]
i.e. founded upon the true value of car ownership as opposed to its predicted value from the first stage regression. Gujarati (2003) explains that standard t-test and F-test procedures can be directly applied to the results of the second stage regression. However, given this formulation for the residuals in the second stage regression, it was found during empirical implementation that it is subsequently possible for nested restricted models to achieve a lower residual sum of squares (RSS) than a general model, suggesting that the nested restricted model outperforms the general model, which is not consistent. Therefore, it was not possible to use an F-test in these circumstances.

A further consideration for model selection is the desire to cross compare the alternative nesting structure on the first stage regression against each other. Different assumptions on the form of the first stage regression result in different fitted values for the first stage regression, which then generate differing parameter estimates in the second stage even if the second stage functional form remains unchanged. This is a feature particular to overidentified systems, and hence not a common feature in the empirical literature. The implication of this is that essentially the regressors in the second stage equations are not necessarily directly comparable as the variables are not equivalent. Hence a standard F-test would not be applicable in this instance.

Therefore, in order to select the preferred multi-equation model, a process of elimination is adopted whereby all models that fail to satisfactorily pass the following tests are instantly rejected; First are the tests for overidentifying restrictions, namely the Sargan and Basmann tests as outlined by Baum, Schaffer & Stillman (2003). Then the Pagan-Hall general test for heteroskedasticity, also outlined in the paper by Baum et al. (2003), is applied to the IV estimates. Once these tests are passed, then the Hausman test of exogeneity is applied to ensure that the model
CHAPTER 7. RESIDENTIAL SITE CAR OWNERSHIP AND USE

outperforms its single-equation counterpart. The RMSE is used as an indiscriminate measure (of model structure or variables chosen) to select the model which specifies the data best.

7.3 Results

Following implementation of the estimation and testing procedures outlined above, this section outlines the results from the semi-parametric estimation of both the single-equation and multi-equation estimations respectively. The semi-parametric coefficients reported in the following tables are the bias corrected least squares coefficients, i.e. the mean coefficient value from the 100,000 bootstrap replications. For comparison, the traditional parametric results are also reported.

Table 7.1 presents the results of the single-equation, or single equation model estimates for both the car use and car ownership models. Only the preferred models are reported due to the large number of alternative specifications available which would make the analysis cumbersome. Focussing firstly upon the car ownership models, the best specified model given the data available is that where car ownership is a function of time and household type. Though, generally speaking, household type indicators prove to be insignificant, but for the exception of household types B (houses for rent) and G (student accommodation). However this significance is revealed only when the assumption of normality is relaxed, i.e. in the semi-parametric model but not in the parametric. The general suggestion is therefore that being grouped in a particular type of housing does not ‘causally’ relate to increased levels of car ownership. The result that students are significantly associated with generally higher levels of car ownership, relative to nurses housing, is likely testament to
the reduced level of car ownership in nurses accommodation. The demand to own a car for nurses is likely lower than for residents living in many other residential development types due to the close proximity nurses must necessarily have with their place of work. Conversely students are often located at university large distance from their family homes, with much need to travel to and fro throughout the course of the academic year, and thus likely have a higher demand. The finding that houses for rent have a higher demand for car ownership may well be part of an overall lifestyle effect, though is equally likely due to be the traditionally high level of parking provision available at households in the UK.

The time variable would imply that the overall trend, year on year, is a reduction in overall levels of car ownership by around 0.05%. This trend is likely to reflect numerous exogenous factors such as increasing congestion, similar to the trend function used by Romilly et al. (2001), which reduces demand by increasing the generalised cost of travel. Further the trend function will reflect improvements in access to public transport and/or the inception of residential travel plans seeking to reduce car dependency within the household. Such policy initiatives would likely involve *inter alia* increased awareness about public transport services - such as at ‘The Bridge’ development in Dartford, or the introduction of car sharing schemes.
Table 7.1: Single-equation model results

<table>
<thead>
<tr>
<th></th>
<th>Parametric (Ownership)</th>
<th>Parametric (Use)</th>
<th>Semi-parametric (Ownership)</th>
<th>Semi-parametric (Use)</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>99.942***</td>
<td>53.878**</td>
<td>100.964**</td>
<td>57.183**</td>
</tr>
<tr>
<td>lnmsa</td>
<td>0.126</td>
<td>-0.134</td>
<td>0.121**</td>
<td>0.133**</td>
</tr>
<tr>
<td>lnnavhhs</td>
<td>-0.134</td>
<td>-0.295</td>
<td>-0.255</td>
<td>-0.293</td>
</tr>
<tr>
<td>lnnavhemp</td>
<td>-0.134</td>
<td>-0.295</td>
<td>-0.255</td>
<td>-0.293</td>
</tr>
<tr>
<td>lnpubtrans</td>
<td>0.121**</td>
<td>-0.295</td>
<td>-0.255</td>
<td>-0.293</td>
</tr>
<tr>
<td>lnη</td>
<td>0.884***</td>
<td>-0.029**</td>
<td>-0.048**</td>
<td>-0.03***</td>
</tr>
<tr>
<td>YEAR</td>
<td>-0.048**</td>
<td>-0.029**</td>
<td>-0.048**</td>
<td>-0.03***</td>
</tr>
<tr>
<td>HTA</td>
<td>-0.254</td>
<td>0.170</td>
<td>-0.183</td>
<td>0.454</td>
</tr>
<tr>
<td>HTB</td>
<td>-1.503</td>
<td>0.337</td>
<td>-1.320*</td>
<td>0.598</td>
</tr>
<tr>
<td>HTC</td>
<td>-0.914</td>
<td>-0.579</td>
<td>-0.799</td>
<td>-0.293</td>
</tr>
<tr>
<td>HTD</td>
<td>-0.870</td>
<td>-0.646</td>
<td>-0.761</td>
<td>-0.357</td>
</tr>
<tr>
<td>HTE</td>
<td>-1.108</td>
<td>-1.351**</td>
<td>-0.981</td>
<td>-1.054</td>
</tr>
<tr>
<td>HTF</td>
<td>-1.158</td>
<td>-1.207**</td>
<td>-0.921</td>
<td>-0.923</td>
</tr>
<tr>
<td>HTG</td>
<td>1.945</td>
<td>-2.209**</td>
<td>1.943**</td>
<td>-1.822</td>
</tr>
<tr>
<td>HTI</td>
<td>0.145</td>
<td>-0.486</td>
<td>0.149</td>
<td>-0.150</td>
</tr>
<tr>
<td>HTJ</td>
<td>1.287</td>
<td>-0.496</td>
<td>1.358</td>
<td>-0.158</td>
</tr>
<tr>
<td>HTK</td>
<td>0.209</td>
<td>0.186</td>
<td>0.278</td>
<td>0.468</td>
</tr>
<tr>
<td>HTL</td>
<td>-0.025</td>
<td>-0.037</td>
<td>0.075</td>
<td>0.250</td>
</tr>
<tr>
<td>HTM</td>
<td>0.669</td>
<td>0.198</td>
<td>0.771</td>
<td>0.486</td>
</tr>
<tr>
<td>HTN</td>
<td>-1.137</td>
<td>-1.089*</td>
<td>-0.975</td>
<td>-0.767</td>
</tr>
<tr>
<td>Obs</td>
<td>146</td>
<td>146</td>
<td>146</td>
<td>146</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.122</td>
<td>0.483</td>
<td>1.063</td>
<td>0.449</td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>0.29</td>
<td>0.80</td>
<td>0.29</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Notes:***=1%, **=5%, *=10%
Table 7.2: Kernel Distributions of 'Household Type' variables from Univariate Car Use Model
The preferred car use models derived from the single-equation specifications relate the demand for travel by car to household type, time and also socio-economic indicators. Once again there is a negative and significant coefficient on the time variable, thus indicating that over time the demand for car use is reducing by around 0.03% per year. A 1% increase in car ownership at the average residential development is met with a 0.8% increase in car use. The results of the parametric model would lead to the conclusion that several household types have significant non-zero coefficients, however analysis of the results from the semi-parametric model shows that this does not hold true under a more flexible inference procedure. This is likely a result of the nature of the dummy variables where, particularly when groups are small in absolute terms, the assumption of normality on the distribution of the coefficients is simply inaccurate. By way of illustration, Figure 7.3 shows the actual coefficient distribution derived from the 100,000 replications along with normal distributions for the household type variables only. From this figure it is evident that some of the household types are clearly producing bi-modal distributions, reflecting the two outcomes available with this type of variable, whereas others show more continuous distributions, and some are merely sporadic.

From a policy perspective, an interesting, albeit unfavourable finding (in terms of achieving sustainability objectives), is that the increased provision of public services is positively and significantly related to increased car use. There are several potential reasons for this, one of which being the argument discussed in Chapter 6 that perhaps public services are not provided for the residential development per se, but rather pass through between two other major destinations. An alternative would be that the provision of bus services fosters a greater overall demand for travel, and the level of demand for car use increases alongside rising demand for bus patronage. Such a phenomenon is perhaps not intuitively obvious but is quite easily rationalised,
the argument is that the substitution of mode of transport by some residents in a
development reduces the level of traffic on the road. This reduction in traffic reduces
the generalised cost of travel and has a resultant demand inducing effect for other
residents dwelling in the area. Although the argument is a little more long winded
than this simple description gives rise to, the simple result is that the use of bus
services as a car travel demand reduction tool, can potentially be ineffective. In the
worst scenario this could potentially result in a higher demand for travel, and in the
present instance there is evidence of this in that the coefficient is positive, however
given that the multi-equation specifications will be shown to statistically outperform
the single-equation models, then this phenomena will be discussed accordingly later.

Turning attention now to the multi-equation models, the first point of discussion
should be the Hausman test results for endogeneity. When applied to the parametric
model indicate that the treatment of car ownership as an endogenous variable in this
two stage system, there is a statistical advantage in the explanation of car use. Thus,
the multi-equation approach is preferred to the single-equation. Over-identification
was allowed for as a feature of the model, and the preferred specification shown in
Table 7.3 is indeed over-identified due to the extra regressors in the first stage equa-
tion for car ownership. The final preferred model form specifies that car ownership
is a function of land zone characteristics, household type, time and socio-economic
factors and that car use is in turn determined by car ownership, household type,

\footnote{Land zone types 6, 7 and 8 are omitted from the results in Table 7.3 as there are no instances of residential developments in these locations. This is not too surprising given that they represent commercial, industrial and development zones respectively. Although it is recognised that some development zones include residential land use, such as the London Docklands redevelopment, as yet these have not featured within the TRICS database in a manner amenable to econometric specification.}

\footnote{Car use is also featured within the socio-economic factors in the standard application of the instrumental variables methodology.}
Chapter 7: Residential Site Car Ownership and Use

Table 7.3: Two-Stage Least Squares model results

<table>
<thead>
<tr>
<th></th>
<th>Parametric</th>
<th>Semi-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 1 (Ownership)</td>
<td>Stage 2 (Use)</td>
</tr>
<tr>
<td>lnT</td>
<td>0.903***</td>
<td>0.914***</td>
</tr>
<tr>
<td>lnη</td>
<td>-0.104</td>
<td>0.046</td>
</tr>
<tr>
<td>YEAR</td>
<td>0.007</td>
<td>-0.012</td>
</tr>
<tr>
<td>HTA</td>
<td>-0.416</td>
<td>0.261</td>
</tr>
<tr>
<td>HTC</td>
<td>0.478</td>
<td>-0.238</td>
</tr>
<tr>
<td>HTD</td>
<td>0.418</td>
<td>-0.447</td>
</tr>
<tr>
<td>HTE</td>
<td>0.949</td>
<td>-1.103*</td>
</tr>
<tr>
<td>HTF</td>
<td>0.971*</td>
<td>-1.048*</td>
</tr>
<tr>
<td>HTG</td>
<td>2.478***</td>
<td>-2.541***</td>
</tr>
<tr>
<td>HTI</td>
<td>0.458</td>
<td>-0.482</td>
</tr>
<tr>
<td>HTJ</td>
<td>0.750</td>
<td>-0.804</td>
</tr>
<tr>
<td>HTK</td>
<td>-0.047</td>
<td>0.156</td>
</tr>
<tr>
<td>HTL</td>
<td>0.101</td>
<td>-0.026</td>
</tr>
<tr>
<td>HTM</td>
<td>0.007</td>
<td>0.030</td>
</tr>
<tr>
<td>HTN</td>
<td>0.779</td>
<td>-0.816</td>
</tr>
<tr>
<td>LU2</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td>LU3</td>
<td>-0.061</td>
<td></td>
</tr>
<tr>
<td>LU4</td>
<td>-0.023</td>
<td></td>
</tr>
<tr>
<td>LU5</td>
<td>0.025</td>
<td></td>
</tr>
<tr>
<td>LU9</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>lnmsa</td>
<td>-0.178</td>
<td></td>
</tr>
<tr>
<td>lnavhhs</td>
<td>-0.690</td>
<td></td>
</tr>
<tr>
<td>lnavhhemp</td>
<td>0.927*</td>
<td></td>
</tr>
<tr>
<td>lnpubtrans</td>
<td>-0.062</td>
<td></td>
</tr>
<tr>
<td>Obs</td>
<td>146</td>
<td>146</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.495</td>
<td>0.539</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.86</td>
<td>0.87</td>
</tr>
<tr>
<td>Sargan</td>
<td>0.154</td>
<td></td>
</tr>
<tr>
<td>Hausman</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Notes: *** = 1%, ** = 5%, * = 10%
CHAPTER 7. RESIDENTIAL SITE CAR OWNERSHIP AND USE

time and socio-economic factors. Note that these functional forms are different to those obtained from the single-equation specifications in particular. land zone features are no longer rejected from the model.

Qualitatively, the models tell a very similar story to that of the single-equation models in that the significant variables would indicate that, for the car ownership model household type G is found to be significantly and positively associated with car ownership. Household type F is found to be significant in the parametric specification, but this significance does not hold when the semi-parametric model is estimated. Unlike in the single-equation models, employment is found to have a significant effect in the parametric mode, though falls just outside of the 10% achieved significance level for the semi-parametric model. Therefore, there is little evidence to support an income effect in determining car ownership levels at residential developments. Conversely, the results of the first stage model would suggest that the demand to own a car is nearly completely determined by the level of car use at a development. There are perhaps other external factors influencing this result, in that the development itself either will or will not be associated with high car use, and the factors which determine this are not necessarily encapsulated within this model. Perhaps, alternatively there is a sense of keeping up with the Jones's at play, in that you are more likely to own a car because the people around you use cars. This however is a conjecture that cannot be qualified with the available data.

With respect to car use, the second stage regression, the results conclude that demand for travel by car is nearly entirely determined by car ownership levels. The seemingly crude nature of this result, i.e. you have a car because you use one and you use a car because you have one, may partially be a result of an aggregation problem. However it is also partially due to the bias correction associated with the
semi-parametric approach, which corrects by using the mean value of the 100,000 bootstrap replications. However the results are perhaps a little misleading, as the shape of the distribution almost imposes this insignificance. By way of illustration, a t-test is conducted on the employment coefficient in the semi-parametric model under the hypothesis that the coefficient on employment is equal to unity;

\[ t(\varepsilon^*) = \frac{\varepsilon^* + (1 - \overline{\varepsilon})}{\sigma^*/\sqrt{n}} = \frac{0.121}{\sigma^*/\sqrt{n}} \]  

(7.8)

i.e. the distribution of the 100,000 bootstrap coefficients is centered around zero, then re-centered around the hypothesised value. Then the data is tested to see whether there is sufficient evidence to reject that this is a statistically valid hypothesis in the same manner as with the ASL test procedure outlined in Chapter 4.

Under this assumption the test result shows that the coefficient is equal to one, with an ASL=0.099, i.e. significant at the 10% level. Given the structure of the distribution for this coefficient this is perhaps not entirely surprising, as the distribution has a marginally fatter tail, both taller and wider, to the right of the mean than to the left. Hence using the mean as the bias corrector for the coefficients may not be the most appropriate way in every circumstance, and in this instance may be underestimating the true effect. Given that this thesis is empirical in nature rather than theoretical this conjecture is not followed through to conclusion, although it may be an interesting avenue for future research.

This (partial) evidence of a positive income elasticity of car ownership, albeit not statistically different from unity, is in accordance with the general literature. For example Dargay & Giuliano (2005) and Dargay (2001) show at a (pseudo) individual level of aggregation for the UK, income effects are significant. Similarly, Romilly, Song & Liu (1998) and Romilly et al. (2001) show that for the aggregate level in
the UK, there is also strong evidence of a positive income effect on car ownership. Therefore the concept of a positive income elasticity should not be bluntly rejected, but at the level of aggregation which this study operates this effect may be masked. Income in the present study as not only aggregated, but also proxied, which may be the cause of the insignificant result. For households, employment levels will likely be correlated with the demand to travel by car at the same time, i.e. the morning and evening peaks. As such there may well be increased congestion, and subsequently a reduced demand to use the car. Consequently a proxy such as employment may be capturing two competing effects and thus cancelling each other out.

Given the data available, and the modelling approach taken, the conclusion is firm that the relationship is bi-directional in nature between car ownership and car use. Although the socio-economic indicators, time effects and land zone type all prove to be individually insignificant, the evidence suggests that they are a necessary feature of the model. From a development control perspective, the present model clearly has some potentially beneficial insights in that the economic, social and geographical features clearly should not be entirely overlooked for development purposes. But when considering the overall environmental impact of a development, if car ownership levels are known, then traffic impact can be easily predicted. A clear policy objective then emerges to reduce car ownership at residential developments in order to curtail their sustenance of pollutants and emissions as well as other externalities generated by traffic.

One further interesting empirical feature arising from the model is the positive and significant coefficient on household type G, suggesting that students have a preference towards higher levels of car ownership. Perhaps this is part of the coming of age ritual, and or a desired status level while at university. Or perhaps it is a
by-product of the funding structure of universities meaning that the students are inherently from wealthier families. Unsurprisingly, this group is also associated with significantly lower car use when compared to the other groups, likely due partially to a lower income effect, are partially due to the close proximity student accommodation normally shares with necessary amenities. This therefore reduces the need to use a car dramatically.

As with the evidence from the single-equation models, the multi-equation approach does not reveal any significant coefficients on the provision of public transport services. Rather, the coefficient on these variables cannot be considered to be different from zero, thus implying that an increase in bus services at a residential development is not conducive with a reduced traffic impact. The suggestion is therefore that traffic levels remain unchanged which implies one of three things; (i) the bus service provided to the area is not being efficiently utilised, and hence there is no substitution effect. This could occur if the service is poorly established providing either a limited number of services at times not suitable to the needs of the residential site, or perhaps that it does not reach the destinations that the residents wish to visit by bus. (ii) the bus service is an effective instrument for providing low cost access to certain amenities/locations which residents wish to use, so there is a positive bus patronage, but coupled with this there is an increase in trips by car from the development due to the lower cost of motoring in the area and/or a potential income effect. (iii) in the context of the findings of Bresson, Dargay, Madre & Pirotte (2003), this could be potential supporting evidence of the inferior good status of public transport in the UK, and substitution effects could be outweighed by rising income effects which are subsequently diminishing the absolute demand for public transport. It is not possible to directly ascertain the root cause of this finding without conducting more qualitative analysis, but the general result, irrespective of
cause, is the effect that public transport is not conducive with low traffic impact developments within the UK to date.

7.4 Summary and Conclusions

The results reveal some interesting issues from a modelling perspective, relating to the general approach to modelling trip generations from residential developments and also to parametric modelling in general. Taking the former issue first, the modelling of trip generations emanating from residential developments is found to share some bi-directional causality with car ownership, as such is generally in accord with some of the more recent empirical applications. Thus at the pseudo-household level of aggregation, car use cannot be rationally treated as independent of the decision to own a car. Given that this is a cross section application, it is difficult to ascertain the true direction of causality between the two, but certainly there is unquestionable evidence of some relationship.

As for the second issue, this analysis reveals that even in instances where standard parametric inference assumptions are not violated, as is the case with the multi-equation model, standard inference techniques can potentially produce misleading results. Although the general error structure of the parametric model satisfies normality assumptions, the assumption does not appear to hold true across the dummy variables in particular, thus resulting in different outcomes when a semi-parametric approach is adopted. Moreover, even for some continuous variables, the specific semi-parametric methodology adopted here, in attempting to present the true parameter estimates given the data, provides potentially misleading econometric results.
Unlike in previous studies such as Hanson & Hanson (1981), or more recently Dargay & Giuliano (2005), the socio-economic indicators are not observed to have any significant effect on travel behaviour in either of the modelling approaches. This itself may be a result of the aggregation of trip activities as well as households. Given the evidence available from extant literature, it would be unwise to state that socio-economic characteristics do not help determine travel behaviour, and indeed they are a statistically valid feature of this model. However the coarse nature at which they are defined here is likely one of the reasons precluding individual significance.

With respect to the general hypothesis to which this thesis adheres to, this chapter provides statistically significant evidence to support the view that land zone placement plays an important role in determining trip generations. Therefore the findings of this chapter are consistent with the preceding analyses in Chapters 5 and 6. The implications of these results in relation to the remainder of this thesis and to the supporting research questions will be summarised in the following concluding chapter of the thesis.
Chapter 8

Summary and Conclusions of the Thesis

This chapter brings together and summarises the general conclusions of this thesis including reviewing the specific results obtained within the three analytical chapters. The first analytical chapter, Chapter 5 analysed trip generations at Office blocks around the UK which dealt with missing data problems in the course of the analysis. The following chapter focussed on the analysis of food superstores, providing in innovative analysis to underpin discussions surrounding food deserts. Chapter 7 presented an analysis of residential sites, broadening the analysis to focus on issues of car ownership and its interaction with trip/travel generating behaviour. The thesis has engaged a scientific approach to test the key hypothesis that traffic behaviour is, in part, determined by land use type. Given this general remit and the research questions identified to support the answering of the research hypothesis, it is possible to draw a number of conclusions, as will be discussed in the remainder of this chapter.
8.1 The Research Questions Revisited

In this section the specific research questions will be answered based upon the evidence presented within this thesis. This will be done in a logical manner, such that the answers to the main research questions will be provided in order, but intervened by answers to sub-research question as appropriate.

(MQ1) What determines vehicle trips?

Given the modelling approach applied to the data, the statistical analysis strongly implies that geographical features such as land zone placement and site specific characteristics determine vehicle trip rates at the three site types considered. In particular, for Offices there is clear evidence that the parking characteristics of the site play the largest determining factor in travel demand. At these site types land zone placement is seen to be a significant feature of the model, as our other socio-economic characteristics, although they are not individually significant. Thus there is evidence that land zone placement is an important determinant, but it is fairly weak.

For food superstores, the evidence is a little clearer, with the results indicating a clear and significant relationship between land zone placement and travel behaviour. Furthermore, the other indicators featured into the model all feature significantly, indicating that for the purpose of food shopping, the decision to travel by car is determined by a rather complex interaction with many factors.

For residential developments, the evidence is a little less clear than with the other Chapters, in that there are no individually significant land zone variables, yet together they are a statistically significant feature of the final preferred model. This
would again imply only weak evidence of a relationship between land zone placement and travel behaviour. However, given that the land zone indicators are important in all three separate models, there is consistent evidence of their validity within trip generation models. Although individually statistically insignificant (for this particular development type), their omission could lead to bias in the coefficient parameters and hence inappropriate policy conclusions. The same is also true of socio-economic indicators and site specific characteristics, which are equally validly accepted to be a necessary feature of trip generation models, in all three land use types.

The results therefore imply that the set of determinants, from a functional perspective, are largely consistent across land use types, though the relationship they actually bear with travel behaviour is very different.

The response to for MQ3, given below, further elucidates the qualitative differences between each of the three site types considered in this thesis.

(MQ2) *What is the best way to model vehicle trips?*

This thesis provides in some senses only a partial response to this question given that the answer must necessarily be double-barreled, as the answer is very much dependent upon circumstance. This thesis was borne, among other things, of a desire to attempt to model the TRICS data in a more rigorous manner than previously done - to that extent the answer to which is the best approach must be caveated with a '...given the use of the TRICS data...'.

A general answer would be to say that there is no definitive answer made apparent by this thesis. The majority of recent work points towards the application of
large scale input-output type models, due to the way they can be incorporated into much larger models of economic behaviour - therefore at a coarse level of aggregation they seem to offer obvious benefits. However such models are often borne of incomplete data and as this thesis has identified, the assumptions made about the incomplete elements of data may substantially impair the accuracy of such models.

When answering the question specific to the work contained within this thesis, it is possible to assert that the regression based trip generation model is the only natural candidate. The form of the data precludes the application of more standard origin-destination analysis, as only one end of the journey is available from the TRICS data. Similarly discrete choice based analysis of individual behaviour is not feasible due to the level of aggregation of the data. Given the depth of the data and spread of observations over time, the possibility of applying time series and/or panel methods cannot be considered appropriate at present - more data, as will be collected over coming years, will however make these estimation technologies feasible in the future.

Table 8.1 summarises the results of the key econometric results of the thesis, coming from the preferred model, and allows for subquestions SQ1, SQ2 and SQ3 to be answered;

(SQ1) What determines Trip rates at Office blocks/Employment sites?

The analysis revealed that the main determinant of trips at office developments is parking provision. Given that incumbent development guidelines allow a set provision of car parking spaces per square footage of development, this is a logical conclusion. Land use features are further found to be a significant feature within
trip generation models for office developments, however they are not individually significant.

(SQ2) What determines Trip rates to food superstores?

For food superstores it has been shown that the level of vehicle trips generated by a site are positively influenced by a number of demand inducing factors such as the provision of a petrol filling station, site size and household composition at the origin of the journey. Also supply side features such as the amount of parking space there are at the site are also positively influencing the demand to travel by car.
The role of land zone placement within the model is seen to have an statistically significant effect, which is either positive or negative depending on the base land zone type that it is compared to.

An interesting further finding is the positive coefficient on bus services, which as discussed elsewhere in this thesis, may be a result of several different factors, in brief these reasons are increased overall travel activity and/or a potential latent demand effect. However the result that bus services are consistently not conducive with lower traffic impacts was an unexpected, albeit rationalisable finding.

**(SQ3)** *What determines Trip rates in residential zones?*

Trip generations from residential developments are determined predominantly by the decision to own a car, with the conclusion that if you have a car, you will use it. There is evidence of a positive income effect, thus supporting standard economic reasoning, and also evidence that household type influences travel patterns.

**§Q1** *What significant differences exist within the various individual segments of the residential sector, of the determinants of vehicle trip making behaviour? i.e privately owned accommodation versus rented accommodation.*

The results imply that almost no significant differences exist between different types of residential dwelling, with the exception of students accommodation. This may to some extent be a result which manifests because the focus of the analysis in this thesis is necessarily upon car trips only. If the analysis were opened up to account for the full range of multi-modal travel behaviour one may well find that
dwelling type is significantly related to the demand for bus services, foot travel or cycling as they will likely relate in some manner to the level of access to different mode types as well as provide a coarse indicator of the level of sprawl within the specific site. However, in the absence of a wide range of multi-modal surveys for this specific site type, such conjectures cannot as yet be empirically contested, though this will serve as a good basis for future research and a solid research question with which to initiate such work.

Following this methodical and scientific approach to the thesis allows the final main research question to answered;

(MQ3) What Qualitative differences are observed in the determinants of trip making behaviour across the main land-use types? e.g. Food Superstores, Office and Residential.

There are both similarities and differences across the three land uses analysed. By way of similarity, in all three substantive analyses it was found that land-use placement is an important feature within these forms of model, at least in the context of the United Kingdom. This is perhaps a relatively unsurprising finding given the amount of literature on the topic, but it is a positive re-enforcement that the proposed directions of government policy are valid.

It was consistently found that public transport services at the developments used for analysis in this thesis were not associated with reduced levels of car travel.
CHAPTER 8. SUMMARY AND CONCLUSIONS OF THE THESIS

8.1.1 Policy questions

This section of the Chapter reviews some of the broader policy context issues which arose during the research undertaken for this thesis. The results of these question are not borne directly of the results of the analytical sections of this thesis, however they form a crucial underpinning of the policy based discussion that surrounds the results contained within. The following therefore briefly recaps the critical issues involved in contemporary transport analysis and also considers the extent to which this thesis provides supporting evidence.

(p1) Prior to answering (p2), what are the key issues in contemporary transport planning?

The key issue in contemporary transport is arguably the generation of a transport system which allocates resources efficiently both for new developments/projects as well as for existing ones. Furthermore this should intentionally be an equitable process which ensures access to transport for all members of the community, particularly those segments of society that are inherently transport poor. As a final addition to this, development of transport solutions should not be inhibiting to the quality of life of individuals, nor to the current or future trading vitality of business entities.

(p2) How can the information from (SQ2) help guide development of work based travel plans?

Based upon these general tenets of what a travel plan (and its associated demand reducing transport solutions) should strive to achieve, the following remarks can be made;
CHAPTER 8. SUMMARY AND CONCLUSIONS OF THE THESIS

With respect to work based travel plans, the burden of evidence indicates firstly that provision of public services *ceteris paribus*, is not consistent with mode switching behaviour and hence is not a feasible instrument to constrain demand for car use. It is the provision of parking spaces at the site in question along with the scale of the business activity at that site that determines travel behaviour - or at least this is what the present data capture within TRICS would conclude. Therefore in the absence of trying to constrain business behaviour by directly controlling site activity, the implication is that parking policies are presently the most direct route to controlling demand for car use at office employment sites.

The TRICS consortium has pre-empted the need for more detailed data, for this site type and others, and the consortium are presently collecting data on travel plan details with every survey they conduct. Within the coming years this will produce a rich source of information to respond to this particular policy question with the level of detail it deserves.

(p3) What insights do the results from (SQ1), (SQ2) and (SQ3) offer in relation to spatial planning policy? i.e. how does land-zone choice (e.g. town center or free standing) affect vehicle trip rates.

The insights made available through this research still leave this answer to this issue somewhat open ended. The general conclusion of this research is that spatial measures consistently prove to be a statistically beneficial addition to trip generation regressions. However it is necessary to re-emphasise the caveat that although improving the overall model specification, it is only for food superstores that individual land zone types achieve significance. This would tend to suggest that more data is needed before any form qualification can be made with respect to this question.
as there are three potential outcomes that may be behind this ambiguity. From a policy perspective this may be indicative of an inertia towards car use, and it may be the case that land zone placement effects, as adhered to above, may be less obvious upon the behaviour of car users, but may have a far greater impact on modes such as cycling or walking. This is a conjecture that should be dealt with empirically in coming years as and when more data becomes available.

\textbf{(p4)} \textit{How suitable is TRICS for this type of research to help inform local level policy?}

This thesis has shown that the TRICS database contains plentiful information to create well specified models of trip generation for certain land use types. Moreover, the results of these models raise some potentially interesting issues for local level planners and policy makers, such as the existence of scale and scope economies for food superstore trips, and the seemingly inadequate manner in which public transport services have been historically implemented.

The analytical framework is unable to really comment on the feasibility of policy instruments, \textit{ex ante} their implementation, this is work better suited to theoretical modelling. However it does provide a clear basis for a model that could control for \textit{ex post} policy implementations - so long as they are transparently recorded within the database and/or associated and easy to access sources. In particular, keeping an eye to the future, the data which TRICS has begun to collate on travel plan policies should soon be able to feature into these kinds of models. Once again the spread of data at the time of writing precludes the inclusion of this into the thesis research.
8.2 Closing comments and discussion

This thesis has sought to answer a number of questions and particularly, establish a scientifically grounded response to two key issues (i) can the data held in the TRICS database be used for 'sound' econometric modelling and (ii) using appropriate econometric methods, can any statistically significant relationship be found to help support discussions surrounding the interaction between land use type and traffic generation in the UK.

The first of these points has been responded to rather emphatically, with the generation of three separate models, largely adhering to a-priori expectations, and satisfying the necessary statistical pre-requisites, albeit with the assumption of normality needing to be dropped in many cases. Based on the results of these models the issue as to whether or not land use type has an impact on trip generation behaviour was answered. Based on the differences identified in the answer to MQ3, it is easy to see that aside from the need for different sets of explanatory variables for different land use types, the magnitude on mutually consistent regressors changes across the different land zone type. Therefore at a general level of geographical consideration, there is a clear distinction between land use types. When considering the effect of land zone types within this, the evidence is somewhat less clear. Land zone placement comes across as being a significant feature within all models, though the significance on individual coefficients providing mixed results.

En route to arriving at these results the thesis touched upon three major concepts, the notion of missing data inside analytical datasets, the issue of food deserts in the UK and finally the implied causality between car ownership and car use for UK residencies.
Bibliography


BIBLIOGRAPHY


Brookes, L. (1978), 'Energy policy, the energy price fallacy and the role of nuclear energy in the uk', Energy Policy 6(2), 94-106.


BIBLIOGRAPHY


Department of Transport (1981), *Junctions and Accesses: Determination of Size of Roundabouts and Major/Minor Junctions*, Department of Transport Highways and Traffic Departmental Advice Note TA 23/81, Department of Transport, UK.


Evans, K. (2006), *Funny weather: Everything you didn’t want to know about climate change but probably should*, Myriad Editions, UK.


Khan, A. & Willumsen, L. (1986), 'Modelling car ownership and use in developing countries', *Traffic Engineering and Control* 27(11), 554-560.


Monbiot, G. (2003), ‘Bottom of the barrel’, *The Guardian Newspaper* (Tuesday December 2\(^{nd}\)).


Office of Communications (OfCom) (2004), *Spectrum Framework Review; A Consultation*.


BIBLIOGRAPHY


STATA (2004), *STATA/SE 8.2 for Windows*, StataCorp LP, USA.


Varian, H. (1992), Microeconomic Analysis, Norton and Company, USA.


Yim, Y. (1992), *The Effects of Transportation Services on the Scale of Food Retailing*, Working Paper, No 112, University of California Transportation Centre (UCTC).