MODELLING COMMUTERS' MODE CHOICE IN SCOTLAND

Arne Risa Hole

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2005

MODELLING COMMUTERS' MODE CHOICE IN SCOTLAND

Arne Risa Hole

Thesis submitted for the degree of

Ph.D.

In Economics and Geography at the

University of St. Andrews

July 2004

(Revised December 2004)

Abstract

This thesis contributes to the literature on the choice of transport mode for commuting trips, with special focus on the difference between urban and rural commuting in Scotland. The thesis begins by giving an overview of discrete choice theory and some empirical models consistent with this theory, before reviewing the literature on empirical applications of mode choice models for commuting trips. In the following, multinomial, nested and mixed logit models using data from a survey of commuters in the University of St Andrews are developed. The models are used to estimate aggregate mode-choice elasticities that can assist the development of efficient car reduction policies in St Andrews and other small towns in rural areas. The direct elasticities of the car mode are found to be comparable to estimates reported in studies of urban commuting, while the demand for public transport is found to be considerably more elastic. The value of in-vehicle travel time is found to be lower than in most studies of urban commuting, reflecting that the roads in the St Andrews area are relatively uncongested. Subsequently, current car drivers' willingness to use a Park and Ride service prior to the implementation of such a service are examined. The results show that the modal shift away from parking on-site will be small unless the new service is accompanied by measures aimed at making parking on-site less attractive such as introducing parking charges. Finally, the effect of the 'compact city' on modal split and congestion are examined. As well as making urban transport more sustainable as a result of an increase in the use of public transport, making cities more compact is found to contribute to lower levels of congestion in urban areas through a reduction in complex trip chains.

Declarations

I, Arne Risa Hole, hereby certify that this thesis, which is approximately 50.000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

December 14, 2004 ....

Arne Risa Hole

I was admitted as a research student in September 2001 and as a candidate for the degree of Ph.D. in September 2001; the higher study for which this is a record was carried out in the University of St Andrews between 2001 and 2004.

December 14, 2004 .. ....

Arne Risa Hole

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

December 14, 2004 ....

Chris Jensen-Butler

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

December 14, 2004 ....

Arne Risa Hole

Acknowledgements

Firstly, I would like to thank my supervisors, Paul Boyle, Felix FitzRoy and Chris Jensen-Butler for their support and advice throughout the process of writing this thesis. I would also like to thank Ian Smith and Otto Anker Nielsen for their useful comments on the first draft of the thesis.

I am grateful to the University of St Andrews, Schlumberger and The Norwegian State Educational Loan Fund for financial support.

I would also like to thank Gerald, Jim, Bram and María for their friendship and encouragement. Having friends around who have already been through the process of writing a Ph.D. helped putting it all in perspective as well as realizing that the project would, at some point, come to an end.

I am also grateful to my uncle, Alf Erling, both for inspiring me to study economics and for his advice throughout my time as a student, and to my grandparents, Arne, Kristin and Eli for their support.

Finally, I would like to thank my parents, Marit and Bjørn, for their continuous support and encouragement.

TABLE OF CONTENTS

Chapter 1

Introduction

Increasing levels of congestion in urban areas as well as a growing recognition of the adverse environmental impacts of increased growth in traffic led to a rethinking of UK transport policy in the 1990s (Goodwin, 1999). As it became increasingly apparent that new road programmes could not keep up with the forecasted increase in the demand for travel by car, supply management (or 'predict and provide') was replaced by demand management as the dominant policy position in the UK. This position was manifested by the 1998 White Paper on transport (DETR, 1998), which emphasised the importance of encouraging the use of more environmentally friendly modes of transport (public transport, walking and cycling), as well as discouraging the use of the private car. The then Secretary of State for Transport, John Prescott, famously proclaimed that:

> I will have failed if in five years' time there are not many more people using public transport and far fewer journeys by car. It's a tall order but I urge you to hold me to it.

Although there has been some progress in achieving an increase in the use of public transport (figure 1.1), few policies aimed at discouraging the use of the private car have been implemented, with the notable exception of the London congestion-charging scheme. Indeed, one change in policy has done exactly the opposite: the fuel tax escalator (the annual 6% increase in the tax on fuel above inflation which was introduced by the government at the advise of the Royal Commission on Environmental Pollution) was abandoned in 2000, and later in the same year the government actually lowered the fuel tax as a result of growing unrest among motorists, contrary to the advice of transport specialists (Begg, 2001). Although the price of petrol has been increasing over the past decade, the total cost of motoring[1] has declined, while the cost of travelling by public transport has increased substantially (figure 1.2). As a consequence of the lack of measures aimed at reducing driving, car use has continued to increase (figure 1.3). In terms of modal split for work trips, the lack of policies in place to discourage car use has led to an increase in the share of work trips undertaken by car over the past decade (figure 1.4).

[1] The total cost of motoring includes purchase, maintenance, petrol and oil, and tax and insurance costs.

Figure 1.1 Passenger travel by public transport in billions of passenger kilometres (source: DfT, 2004)

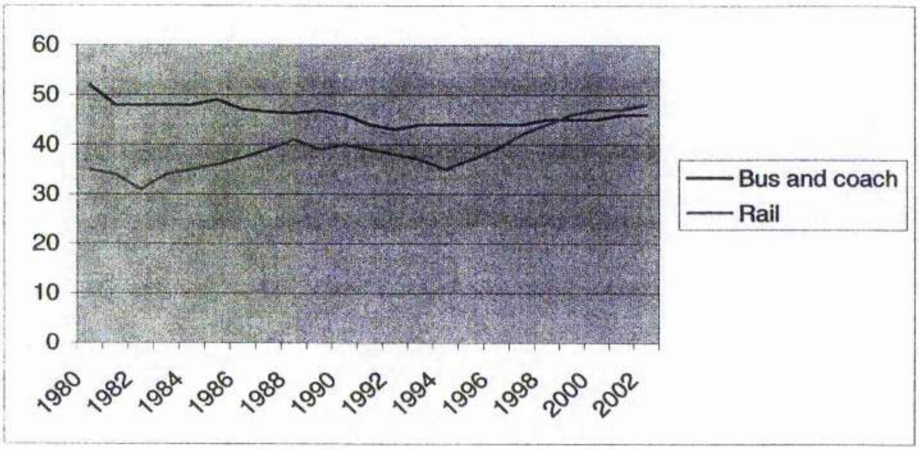


Figure 1.2 Changes in the real cost of transport and in income (source: DfT, 2004)

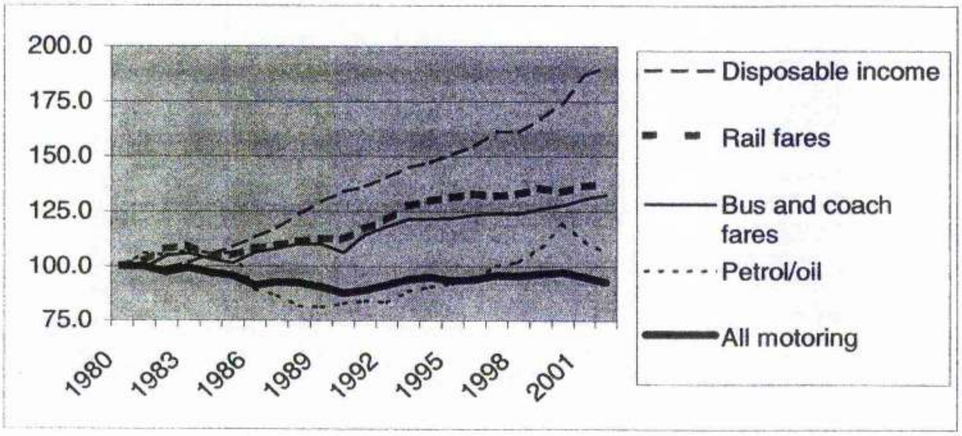


Figure 1.3 Passenger travel by car, van and taxi in billions of passenger kilometres (source: DfT, 2004)

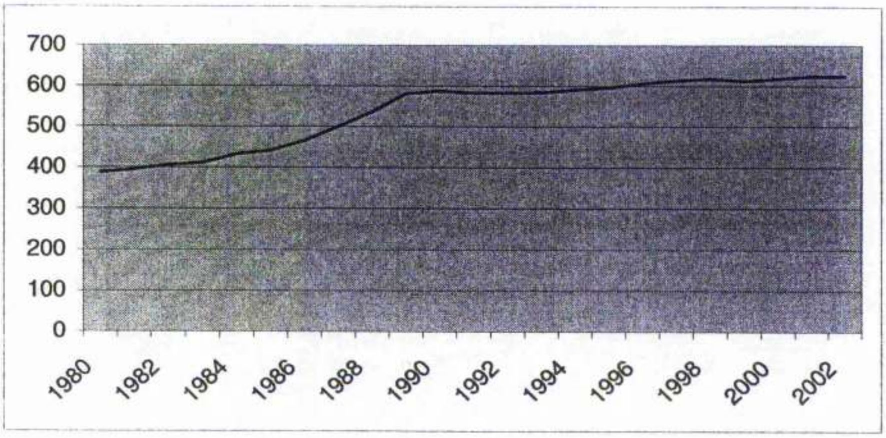
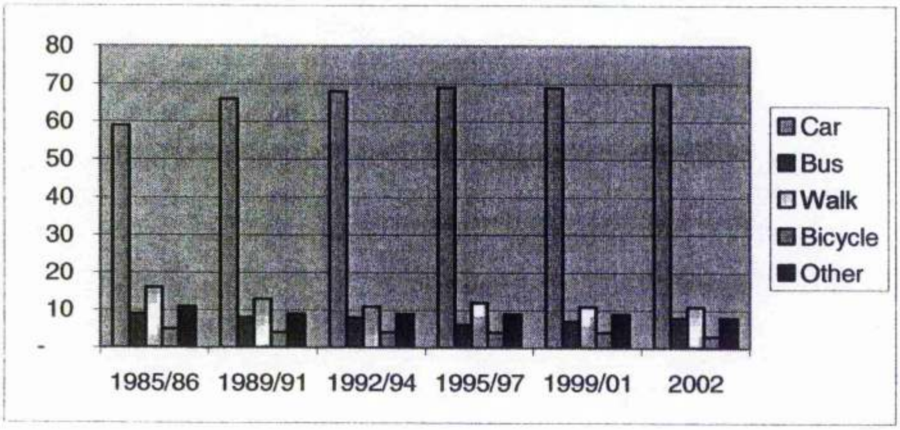


Figure 1.3 Main mode of travel to work – percentage of trips (source: DfT, 2004)



This thesis contributes to the literature on the choice of transport mode for work trips, with special focus on the difference between urban and rural commuting in Scotland.[2] Rural commuting differs from urban commuting in several important respects: there is little or no road congestion, a parking space is usually provided free by the employer and the supply of convenient public transport is often limited. As a result a high share of rural commuters will depend on the private car to get to their workplace. An important consequence of these differences is that car reduction policies designed for large cities with ample public transport may be unsuitable for smaller towns. Relatively little research has been done on commuting in rural areas, however, and the present thesis contributes to filling this gap in the literature. The data used for the analysis include original data collected by the author on staff commuting in the University of St Andrews, as well as data from the Scottish Household Survey Travel Diary. To the author's knowledge the latter dataset has not been used previously to explore the difference between rural and urban commuting. It should be noted the analysis focuses on the *demand* for transport, and that in line with the majority of the literature on commuters' mode choice supply-side characteristics are assumed to be exogenously given. Further, the thesis does not form the whole of an analysis that aims to evaluate changes in consumer welfare arising from the various

[2] In spite of the fact that the recently established Scottish Parliament has legislative control over most aspects of transport in Scotland, the development of transport policy in Scotland has been similar to that in the rest of the UK (Smyth, 2003). As Westminster is still responsible for UK fiscal policy, however, the flexibility of introducing 'Scotland specific' transport policies is somewhat limited.

policies discussed, as this analysis would also have to take the costs of implementing the policies into account.

The structure of the thesis is as follows:

The thesis starts by giving an overview of microeconomic choice theory, with focus on the random utility model and the theory of time allocation that underpins the empirical models of discrete choice analysis (chapter 2). Chapter 2 also presents some of the criticisms raised by behavioural scientists against the microeconomic model.

Chapter 3 presents the empirical methodology used for modelling discrete choices, with emphasis on the multinomial logit, nested logit and mixed logit models. Some basic hypothesis tests and specification criteria as well as standard procedures for deriving aggregate predictions from disaggregate models are also discussed.

Chapter 4 describes the benefits and drawbacks of stated and revealed preference data, which are the two data types most commonly used for travel demand analysis. I describe tests for two phenomena that may arise when using stated preference data (fatigue/ learning effects and the repeated measurements problem) as well as some ways to combine revealed and stated preference data to obtain more robust estimates of the model coefficients.

Chapter 5 provides a comprehensive literature review of studies modelling commuters' mode choice. The chapter starts by reviewing McFadden's (1974, 1978) seminal work on commuting in the San Francisco Bay Area and goes on to discuss more recent contributions to the literature.

As mentioned above, car reduction policies designed for large cities with ample public transport may be unsuitable for smaller towns. In particular, pricing policies designed to encourage public transport use may be less effective, as commuters with no convenient substitute to driving will be unable to switch. Chapter 6 develops multinomial, nested and mixed logit models using data from a survey of commuters in the University of St Andrews. The models are used to estimate aggregate mode-choice elasticities that can assist the development of car reduction policies in St Andrews and other small towns in rural areas. The direct elasticities of the car mode are found to be comparable to the estimates reported in studies of urban commuting, while the demand for public transport is found to be considerably more elastic. Although this is partially a result of the fact that bus has a substantially lower market share in St Andrews compared to larger towns and cities, the finding nevertheless indicates that there is scope for increased use of public transport for commuting in St Andrews and other small towns in rural locations. The value of in-vehicle travel time is found to be lower than in most studies of urban commuting, reflecting that the roads in the St Andrews area are relatively uncongested. The value of walking time is found to be about 7-8 times higher than the value of in-vehicle time, while the value of cycling time is about 60-80% of the value of walking time.

Travel plans are an important tool in making transport more sustainable at workplaces in the UK. One of the measures that can be taken by employers in order to reduce the need for employees to take their cars to the workplace is setting up a Park and Ride service. Chapter 8 examines current car drivers' willingness to switch to Park and Ride prior to the implementation of such a service. Since there will be no revealed preference (RP) data available in this case, data derived from a stated preference (SP) experiment are used to calibrate the models. The models are subsequently used to forecast the demand for Park and Ride. Since it is well known that SP data contain sources of variation not present in RP data, special attention is paid to the scaling of the SP model. The results show that the modal shift away from parking on-site will be small unless the new service is accompanied by measures aimed at making parking on-site less attractive such as introducing parking charges.

One of the often-cited benefits of the 'compact city' is that it offers the potential for developing an efficient public transport system, which in turn encourages commuters to travel by public transport. In chapter 9 I argue that a potential second benefit of making cities more compact is a reduction in peak hour congestion on urban roads. Since urban dwellers are expected to be less likely to link non-work activities to the commute than those who live outside the city and commute to the city to work as the gain from trip chaining is lower for those living close to facilities, urban residents contribute relatively less to peak hour congestion. This is confirmed by the modelling results: as well as making urban transport more sustainable as a result of the increased use of public transport, making cities more compact is found to contribute to lower levels of congestion in urban areas, since the reduction in complex trip chains implies that fewer trips will be undertaken during peak hours.

Finally, Chapter 10 offers some concluding remarks.

Chapter 2

Choice Theory

In this chapter the theoretical foundations of disaggregate travel demand models are discussed. Although the focus of this section will be on mode choice, the theories outlined in section 2.1 and 2.2 can be applied to all types of travel-related choices such as the choice of whether or not to make a trip, departure time and travel route. Section 2.1 provides an overview of economic choice theory, section 2.2 outlines the structure of the random utility model that underlies the specification of the empirical models described in chapter 3, while section 2.3 discusses the relationship between conventional microeconomic consumer theory and random utility models. Section 4 gives an overview of some of the criticism raised by behavioural scientists against the standard model outlined in the first three sections.

2.1 An outline of economic choice theory

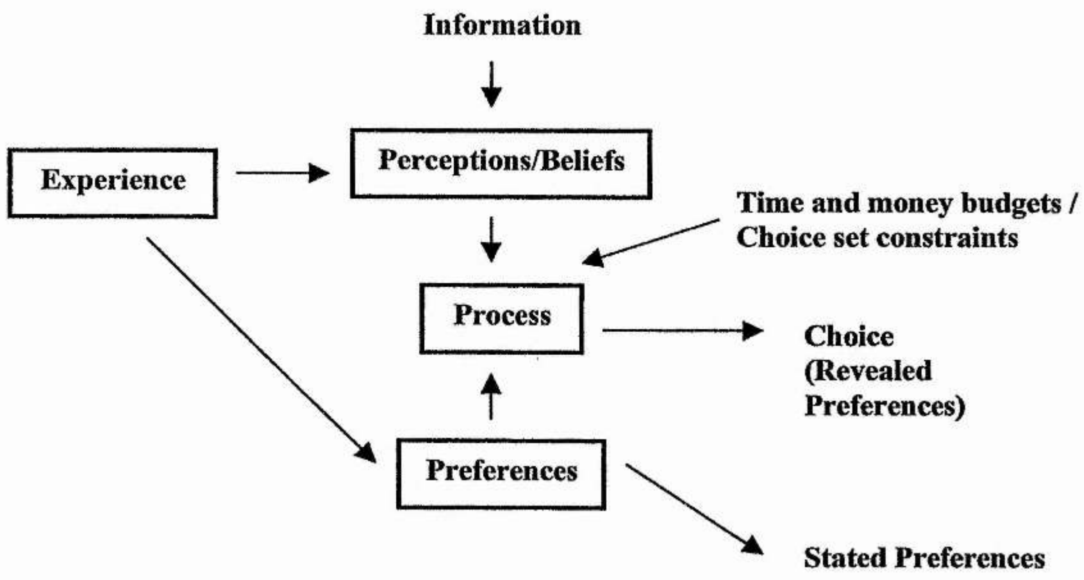
The theoretical foundation of disaggregate travel demand models has its roots in Lancaster's (1966) microeconomic theory of consumer demand and the Random Utility Theory developed by Thurstone (1927), Marschak (1960) and McFadden (1973). In his theory, Lancaster postulated that the demand for goods depends on the characteristics or attributes of the goods rather than the goods *per se*. The basic structure of a random utility model is outlined in section 2.2.

Figure 2.1 illustrates the choice process. The individual receives information about the alternatives in a choice context (for instance the modes available for the work-trip) and processes the information to form perceptions of the attributes of the alternatives. The individual is assumed to have perfect information about the attributes and attribute levels that are relevant for each alternative available. Given her perceptions of the attributes the individual is assumed to behave as if she translates this information into a utility index based on her preferences and chooses the alternative with the highest utility given financial and time constraints. Since a satisfactory level of one attribute can compensate an unsatisfactory level of another attribute this type of behaviour is called "compensatory". Not all alternatives will be available to all individuals since each individual faces time and income constraints (such that a particularly slow or costly mode may not a feasible option) as well as socio-demographic constraints (for instance the individual's car ownership level can restrict her from using the private car).

The individuals' preferences are assumed to be stable and innate, but the way preferences are expressed in a choice situation can change with experience. An important consequence of this assumption is the "consumer sovereignty" property which states that preferences do not depend on the alternatives available in a given choice situation. In other words "desirability precedes availability" (McFadden, 2001). The individual's perceptions and beliefs about the alternative attributes are modified through experience. The influence of experience on choice behaviour, however, is not usually incorporated in applied models of discrete choice due to demanding data requirements. In order to investigate how experience influences choice behaviour several observations per individual at different points in time are required. Experience is not likely to play a role in a stated preference experiment (see chapter 4) where repeated choices are made because of the short time span between the choice tasks performed (note that experience is not the same as the learning effect discussed in chapter 4, since this is related to learning in the experimental setting and thus not relevant to actual choice behaviour).

Both the strength and the weakness of the standard economic model lie in its simplicity. The strength of the model is that it is straightforward to make operational and thus it is very useful as a practical tool in travel demand analysis. It is also successful in explaining and predicting many types of market behaviour. On the other hand there is much behavioural evidence suggesting that people in many situations do not behave in a way which is consistent with the model. This seems to particularly be the case in hypothetical choice situations, but also under some market conditions (McFadden, 1999). As a result of this evidence many behavioural economists and psychologists claim that the assumptions made in the standard economic model are unrealistic. We will discuss some of the criticisms raised against the standard economic model in section 2.4.

Figure 2.1. The choice process



Source: McFadden (2001)

2.2 Random utility theory

The random utility theory developed by Thurstone (1927), Marschak (1960) and McFadden (1973) is the theoretical cornerstone of disaggregate travel demand models. The difference between random utility theory and conventional microeconomic consumer theory is that the researcher is assumed to have incomplete information about the factors that influence the individuals' choices. As a result the choice outcome is probabilistic, or random, rather than deterministic as in the conventional theory of consumer behaviour. The basic structure of the random utility model is outlined below. For a more complete overview of the random utility model see Ben-Akiva and Lerman (1985) or Train (2003).

Let U_{ni} be the utility individual n derives from choosing transport mode i.[1] The individual is assumed to choose the mode that maximises her utility from a set of J alternatives. Furthermore it is assumed that the utility U_{ni} can be partitioned into a systematic component or "representative utility", V_{ni}, and a random component, ε_{ni}. Hence, utility is given by:

$$U_{ni} = V_{ni} + \varepsilon_{ni} \qquad (2.1)$$

The representative utility V_{ni}, is a function of the attributes of mode i and the individual's observable socio-demographic characteristics, while ε_{ni} represents characteristics and attributes unknown to the researcher, measurement error and/or heterogeneity of tastes in the sample. Since the unknown variable, ε_{ni}, is treated as random by the researcher, this class of utility models is called ***random*** utility models. Note that the individual does not maximise utility in a random manner, the randomness occurs because the researcher cannot accurately observe all the variables that influence the individual's choice.

From the researcher's point of view, however, the maximization process is random and therefore probabilistic rather than deterministic. Specifically, the probability that individual n chooses mode i rather than mode j is the probability that the utility of choosing i is higher than the utility of choosing j:

$$\begin{aligned} P_{ni} &= P(U_{ni} > U_{nj}) \\ &= P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) \\ &= P(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}) \end{aligned} \qquad (2.2)$$

[1] U_{ni} is the indirect utility function, rather than the direct utility function (see section 2.3).

Denoting the joint densitiy function of the random terms by $f(\varepsilon_n)$, the probability that alternative i is chosen is given by:

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}) f(\varepsilon_n) d\varepsilon_n \qquad (2.3)$$

where $I(\cdot)$ equals 1 when the expression in parenthesis is true and 0 otherwise. Hence the probability of choosing alternative i depends on the distribution of the random terms, and different choice models arise from different assumptions regarding this distribution.

In chapter 3 some of the econometric models consistent with the random utility theory are presented. The following section takes a closer look at the microeconomic theory underlying the specification of the representative utility function in discrete choice models.

2.3 Time allocation and the value of time

Since travel is essentially a time consuming activity the choice of transport mode can be incorporated into the more general microeconomic theory of time allocation. The inclusion of time in the consumers' maximisation problem was originally motivated by the need to understand the supply side of the labour market, which can be viewed as the individuals' choice between working and spending time on leisure. Becker (1965) was the first to introduce time as a central component in a model of consumer behaviour with later important contributions by DeSerpa (1972) and Evans (1972). In the transport field the goods/ leisure framework introduced by Becker has been used to develop operational models to give a sound theoretical foundation to empirical discrete choice models (Train and McFadden, 1978; Troung and Hensher, 1985; Bates, 1987; Jara-Díaz and Farah, 1987; Jara-Díaz and Videla, 1989 and Jara-Díaz and Guevara, 2003).

We will pay particular attention to the contributions by Train and McFadden (1978), and Jara-Diaz and Farah (1987) since they have become the standard departure for the specification of representative utility in applied mode choice modelling. In Train and McFadden (1978) the individuals are assumed to maximise their utility by choosing an optimal level of goods consumption and time spent on leisure subject to time and budget constraints. Since income depends on the time spent working, the individual must trade off time spent on leisure with consumption according to her preferences. Formally, this maximisation problem can be written as:

$$\text{Max } U(G,L) \tag{2.4}$$

subject to

$$G = V + wW - c_i \tag{2.5}$$

$$L = T - W - t_i \tag{2.6}$$

where G is the value of the goods consumed (assuming that the price index is constant and normalised to 1), L is the time spent on leisure, V is non-labour income, w is the wage rate, W is the time spent working, T the total time budget and c_i and t_i the cost and time spent commuting by mode i respectively. Since both G and L can be expressed as a function of W, utility can be restated as a function of the variables that are assumed to be under the individuals control: the amount of time spent working and the choice of transport mode (the individual is assumed to have no influence over non-labour income, the wage rate and the total time budget, so V, w and T are exogenous). This maximisation problem can then be solved in two steps: first the individual decides on the optimal level of W *conditional* on the mode of transport, and subsequently the mode of transport that maximises utility given the conditional demand for working time. The maximisation problem can then be restated as follows:

$$\underset{i}{Max}\{\underset{W}{Max}\, U[G(W,c_i), L(W,t_i)]\} \tag{2.7}$$

Maximising U with respect to W yields the demand for working time conditional on mode choice, $W^*(c_i,t_i)$. By substituting $W^*(c_i,t_i)$ back into (2.7), the conditional indirect utility function is obtained:

$$U_i \equiv U\{G[(W^*(c_i,t_i),c_i], L[W^*(c_i,t_i),t_i]\} \tag{2.8}$$

Denoting the set of available modes by J, the mode $j \in J$ which maximises U_j is chosen by the individual. Train and McFadden present three functional forms of the utility, showing in each case how they lead to different forms of the indirect utility function. Here only the more general functional form, the Cobb-Douglas function, will be presented. In this case utility is given by $U = KG^{1-\beta}L^{\beta}$ where K is a constant. Solving the utility maximising problem conditional on mode choice yields:

$$U_i = K(1-\beta)^{1-\beta}\beta^{\beta}[w^{-\beta}(V-c_i)+w^{1-\beta}(T-t_i)] \tag{2.9}$$

which is the conditional indirect utility function, corresponding to (2.7) in the general case. Since only the variables associated with the alternatives (cost and time) will influence mode choice, this function can be rewritten by omitting the terms that are constant across modes:

$$\bar{U} = K(1-\beta)^{1-\beta}\beta^{\beta}[-w^{-\beta}c_i - w^{1-\beta}t_i] \tag{2.10}$$

which is what Jara-Diaz (1998) calls the ***truncated conditional indirect utility function***. It can easily seen from (2.10) that if $0<\beta<1$, the term in the square brackets is $-w^{-\beta}c_i - w^{1-\beta}t_i$, when $\beta=0$ it is $-c_i - wt_i$, and when $\beta=1$ it is

$-\frac{c_i}{w}-t_i$. These functional forms can then be adopted when specifying the representative utility function, V_i, in order to be consistent with utility maximising behaviour. Train and McFadden point out the choice of β is an empirical issue, and hence recommend testing the various specifications consistent with different values of β. Most applied work, however, has settled for the cost over wage specification of the representative utility function (Jara-Díaz, 1998).[2]

Jara-Diaz and Farah (1987) argue that a drawback of Train and McFadden's model is that the time spent working, W, cannot necessarily be realistically assumed to be an endogenous variable, as many individuals have fixed working schedules with little or no possibility of working longer hours. They therefore rephrase the maximisation problem (2.4) – (2.6) as follows:

$$\text{Max } U(G, L) \tag{2.11}$$

subject to

$$G = I - c_i \tag{2.12}$$

$$L = T - W - t_i \tag{2.13}$$

where I is the income earned by working a fixed number of hours (W) (non-labour income V has been dropped for simplicity). Since working time is fixed in this case, the conditional representative utility function can be obtained directly by substituting the constraints into the utility function:

$$U_i = K(I - c_i)^{1-\beta}(T - W - t_i)^{\beta} \tag{2.14}$$

[2] Since data on wages is usually not available it is customary to specify the variable as cost over income, where income is a proxy for the wage rate.

Since (2.14) is non-linear in cost and time, however, it is not very helpful for model specification. If utility is approximated to a first order Taylor expansion around (I,T-W), replacement of G and L yields (Jara-Diaz and Ortuzar, 1989):

$$U_i = K[I^{1-\beta}(T-W)^{\beta} - (1-\beta)\left(\frac{T-W}{I}\right)^{\beta}]c_i - \beta\left(\frac{I}{T-W}\right)^{1-\beta} t_i] \qquad (2.15)$$

In this case the truncated conditional indirect utility function can be written:

$$\bar{U}_i = -\theta(1-\beta)\frac{c_i}{g} - \theta\beta t_i \qquad (2.16)$$

where, g is an expenditure rate $I/(T-W)$ and θ is $Kg^{1-\beta}$. This implies a specification of the representative utility function which is similar to the cost over wage specification derived by Train and McFadden, only that the wage rate in the latter is replaced by the expenditure rate, g. Jara-Diaz and Ortuzar (1989) compares the wage rate and expenditure rate specifications using data on commuters in Santiago, Chile, and finds that the expenditure rate specification results in the superior model fit.

Equations (2.15) and (2.16) are only valid representations of (2.14) if the conditional representative utility function can be sufficiently closely approximated by a first-order Taylor expansion. As pointed out by Jara-Díaz (1998), it may be that a second order expansion results in a better approximation. In this case the truncated conditional representative utility function is given by (Jara-Díaz, 1998):

$$U_i = -\theta(1-\beta)\frac{c_i}{g} - \theta\beta t_i + \frac{1}{2}\theta\beta(1-\beta)(S_T - S_I)(\frac{c_i}{g} - t_i) \qquad (2.17)$$

where S_I and S_T is the share of income $\left(\frac{c_i}{I}\right)$ and free time $\left(\frac{t_i}{T-W}\right)$ spent on commuting respectively. This expression has a number of interesting implications.

Firstly, if either S_I or S_T are significantly different from zero, a second-order approximation to (2.14) is appropriate (it can easily be seen that 2.17 reduces to 2.16 when $S_T = S_I = 0$). Secondly, it can be seen that if S_I and/or S_T are significantly different from zero, second order terms in travel time, travel cost or both should be included in the specifiation of the representative utility function (Jara-Díaz, 1998). Jara-Díaz and Videla (1989) point out that the second order term in cost represents an income effect, or that the marginal utility of income is a decreasing function of transport costs. Intuitively, this effect reflects that when travel costs increase, income falls and, given that travel costs represent a non-trivial proportion of overall income, an additional unit of income becomes more valuable to the individual.

It should be clear from the preceding discussion that theoretical analysis alone cannot determine which form of the representative utility function is the most appropriate, and that it is necessary to test various functional forms before determining which specification of model is preferred, based on various model specification criteria such as data fit (see chapter 3) and whether the sign of the coefficient estimates are logical. To the extent that the data allows it, the results of this discussion will be taken into account when specifying the models in the empirical sections of the thesis.

Apart from model specification, the theoretical literature on travel demand analysis has focused extensively on the derivation of the subjective value of time (*SVOT*), which is given by the rate of substitution between the time and cost of travelling by mode ***i*** (Jara-Díaz and Ortúzar, 1989):

$$\left.\frac{-dc_i}{dt_i}\right|_{V_i} = \frac{\partial V_i / \partial t_i}{\partial V_i / \partial c_i} \tag{2.18}$$

Using equation (2.18) it is easy to see that the value of time in Train and McFadden (1978) is given by:

$$SVOT = \frac{\partial V_i / \partial t_i}{\partial V_i / \partial c_i} = w \tag{2.19}$$

In other words the subjective value of time equals the wage rate. This result is rarely supported by empirical evidence, and Train and McFadden show how to generalize their model to include time and cost specific coefficients, such that *SVOT* equals the ratio of the coefficients times the wage rate:

$$SVOT = \frac{\partial V_i / \partial t_i}{\partial V_i / \partial c_i} = \frac{\gamma_t}{\gamma_c} w \tag{2.20}$$

where γ_t and γ_c are the coefficients for time and cost, respectively. In this case it is easy to see that $\frac{\gamma_t}{\gamma_c}$ represents *SVOT* as a percentage of income ($SVOT / w$). This result has been widely used in empirical analysis.

In Jara-Diaz and Farah (1987), the value of time is given by:

$$SVOT = \frac{\partial V_i / \partial t_i}{\partial V_i / \partial c_i} = \frac{\beta}{(1-\beta)} \frac{I - c_i}{T - W - t_i} \tag{2.21}$$

An interesting implication of equation (2.21) is that the value of time is a decreasing function of travel costs, and an increasing function of travel time. Operationally, this effect can be captured by entering second order terms in the representative utility function as implied by equation (2.17). As pointed out by Jara-Díaz and Videla (1989) the negative relationship between *SVOT* and transport costs reflects that the marginal utility of income decreases with transport costs. The positive relationship between *SVOT* and time, on the other hand, reflects that as travel times increase, the time spent on leisure falls and hence leisure time becomes more valuable. It should be pointed out that it is possible from a behavioural point of view that an increase in travel times from, say, 5 to 10 minutes, is perceived to be more costly, than an increase from 30 to 35 minutes. This result is not consistent with the utility maximising model, however. The issue of model specification and value of time estimation will be discussed further in the empirical applications in chapters 6 and 7.

2.4 Behavioural criticism of the standard model.[3]

As mentioned in section 2.1 the standard model has been met by much criticism, particularly from behavioural psychologists who have shown that several of the assumptions of the model are not supported by experimental evidence. One of the criticisms raised is that the assumption that individuals have perfect information is not likely to hold in many circumstances. Simon (1955) was the first to argue that if collection of information is costly individuals are likely to collect information on alternatives only up until the point where the added benefit of collecting more information (the possibility of finding a better alternative) outweighs the cost of collecting the information. Thus the information collection stops when a satisfactory alternative is found. There are several claims in the literature that individuals use such simplified decision rules, or "heuristics", when making their choices. In the "elimination by aspects" theory (Tversky, 1972) individuals are assumed to focus on the attribute that they find most important (such as the time of the mode) and choose the alternative that is best in terms of this attribute. If two or more alternatives are equal in respect to this attribute the individual compares those alternatives with regard to the second most important attribute. This process goes on until a preferred alternative is found. This choice procedure is "non-compensatory" as opposed to the standard economic model since only the level of the most important attribute decides the choice outcome.

In addition to the claim that individuals are likely to use simplified decision rules when making choices there may be circumstances where individuals do not make a deliberate choice at all due to factors such as habit and inertia (Peter and Olson, 1994; Verplanken *et al.*, 1997). The argument is that an individual will only make a conscious choice when there is a major change in the travel conditions (such as moving house or changing jobs) and then stick to that alternative without considering all available alternatives each time she travels. This is a potentially important point to bear in mind in terms of policy analysis, as a policy change might not be effective unless it's marketed in such a way that the targeted individuals are made aware that there is a change in the travel conditions and hence feel the need to reconsider their travel choices. These factors can be investigated in a stated preference model where individuals provide several responses (see chapter 4) or a revealed preference model estimated using panel data. Since panel data is very costly to obtain the first approach will be the more practical solution in most circumstances.

A further criticism of the standard choice model is that preferences may not be stable over time. Tversky and Kahneman (1974, 1981) argue this assumption is unrealistic due to what they call "framing effects". In short the "framing" of the choice situation, or how the choice situation is presented to and perceived by the individual, influences her behaviour. According to Tversky and Kahneman individuals seem to be more concerned about reducing risk than making an optimal

[3] The literature on this subject is so extensive that it would require a separate PhD thesis to do it justice. Here I will try to summarize some of the main ideas in the literature, particularly those relevant for improving the specification of the standard model.

choice. Because of this attitude towards risk they demonstrate in an experimental setting that in the choice between two alternatives individuals choose the alternative they perceive to be less "risky" even though the objective conditions remain the same.

Behavioural psychologists use the term "attitude" to describe the individuals' overall evaluation of an alternative (see Peter and Olson, 1994). The overall evaluation consists of the perception and beliefs about alternative attributes as well the *emotional appeal* of the alternative. Peter and Olson argue that when individuals are more involved in the decision making process, when the decision is of importance to the individual, the cognitive process (evaluation of alternative attributes) plays a larger role. Emotions seem to play a larger role in the choice of some products such as ice cream or sports cars (Nerhagen, 2001). In the mode choice context there is evidence that commuters have an emotional attachment to the car mode (Stradling *et al.*, 1999).

The important question that these studies implicitly raise for a practitioner is how to take this evidence into account when developing travel demand models. McFadden (1999) summarizes the task in the following quote:

> The challenge is to evolve [the standard model] in the direction of [the psychological views of decision making], adopting those features needed to correct [the standard model's] most glaring deficiencies as a behavioural model, and modifying economic analysis so that it applies to this hybrid.

McFadden argues that there is scope for modifying the standard model in such a way that it takes some of the criticisms raised by behavioural scientists into account, for instance by incorporating the idea of bounded rationality into the decision framework.

Nerhagen (2001) suggest that more effort should be made when estimating discrete choice models in 1) the formation of the individuals' choice set and 2) the specification of the functional form of the utility function. She argues that choice set formation in discrete choice models is under-researched in the literature given its behavioural importance, although there are exceptions (Swait, 2001; Swait and Ben-Akiva, 1987, Ben-Akiva and Boccara, 1996). Furthermore the traditional discrete choice model incorporating the time and cost of the modes only may be biased as they omit other important factors influencing choice behaviour such as comfort and reliability. There are, however, some examples of work taking such attributes into account (DePalma and Rochat, 2000; Noland and Kunreuther, 1995).[4] The problem of including "softer" attributes in the individuals' utility function, however, is that they are difficult to quantify. One solution is to ask individuals about their perception of these attributes, although since the link between perceptions and objective values is unclear this may be of limited usefulness if one wishes to investigate how changes in these variables influence choice (Small, 1992).

[4] See chapter 5.

Chapter 3

Empirical Methodology

This chapter gives an account of three econometric models that are consistent with the random utility framework outlined in chapter 2. Section 3.1 describes the simplest and by far the most popular model of discrete choice, the multinomial logit model, before going on to describe the more flexible nested logit and mixed logit models in sections 3.2 and 3.3 respectively. Section 3.4 presents some standard goodness of fit measures and hypothesis tests, section 3.5 describes how to produce forecasts of aggregate behaviour, section 3.6 is devoted to heteroscedasticity and section 3.7 describes the various issues concerning the derivation of value of time estimates from discrete choice models.

3.1 The multinomial logit model

3.1.1 Multinomial logit choice probabilities

McFadden (1973) shows that if the unobservable components[1] in the random utility function, ε_{ni}, are distributed independently, identically (IID) extreme value, the probability that individual **n** chooses alternative **i** from a set of **J** alternatives is given by:

$$P_{ni} = \frac{e^{\frac{1}{\mu}V_{ni}}}{\sum_j e^{\frac{1}{\mu}V_{nj}}} \qquad \text{where } j = 1,\ldots,J \text{ and } n = 1,\ldots,N \tag{3.1}$$

where $\frac{1}{\mu}$ is a positive scale parameter and V_{nj} is the representative utility function described in chapter 2. McFadden called this model the conditional logit model, since it has the form of a conditional probability and the error difference follows the logistic distribution in the two alternative case. It is now more commonly referred to as the multinomial logit (MNL) model. The representative utility function is normally specified to be linear in parameters, $V_{nj} = \alpha_j{}'c_n + \beta' x_{nj}$. In this case the model takes the form:

$$P_{ni} = \frac{e^{\frac{1}{\mu}(\alpha_i{}'c_n + \beta' x_{ni})}}{\sum_j e^{\frac{1}{\mu}(\alpha_j{}'c_n + \beta' x_{nj})}} \tag{3.2}$$

[1] In econometric terminology the unobserved components are often referred to as *errors*. I will use both terms throughout the thesis.

where α_j and β are vectors of coefficients, x_{nj} is a vector of observed attributes relating to alternative j and individual n and c_n is a vector of observed characteristics of person n.[2]

The scale parameter, $\frac{1}{\mu}$, can be shown to be inversely proportional to the error variance, σ_ε^2 (see Ben-Akiva and Lerman, 1985):

$$\frac{1}{\mu}=\frac{\pi}{\sqrt{6\sigma_\varepsilon^2}} \tag{3.3}$$

Since the overall scale of utility cannot be identified in estimation it is customary to impose the normalization $\frac{1}{\mu}=1$, which is equivalent to assuming that the error variance equals $\frac{\pi^2}{6}$. The consequence of this normalisation is that the true scale parameter will be confounded with the α_j and β parameters. In other words $\frac{\alpha_j}{\mu}$ and $\frac{\beta}{\mu}$ will be estimated, not α_j and β. This causes some problems when using stated preference models for forecasting which will be discussed in chapter 4.

The parameters in the MNL model are normally estimated using the method of maximum likelihood. The log-likelihood function is given by:

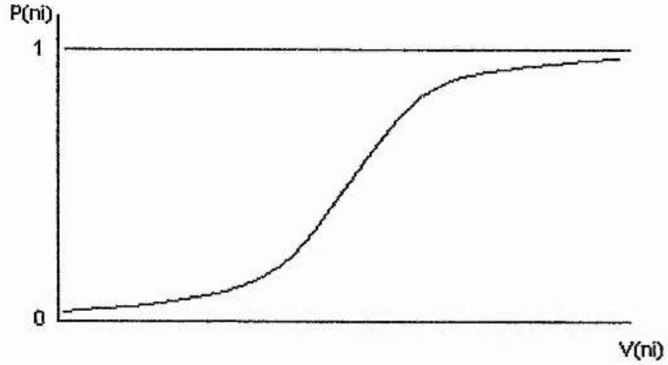
$$LL(\alpha_j,\beta)=\sum_n\sum_j y_j \ln(P_{nj}) \tag{3.4}$$

[2] Note that β is assumed to be constant over individuals and alternatives while α_j are constant over individuals but not alternatives. This is necessary since there is no variation in c_n over alternatives and implies that characteristics of the individual affect the utility of alternatives differently (see Griffiths *et al.*, 1993). Furthermore, since the level of utility cannot be identified in estimation, one of the α_j coefficients needs to be normalised to zero for identification purposes.

where $y_{nj} = 1$ if individual *n* is observed to choose alternative *j* and 0 otherwise. The values of α_j and β which maximises this function gives the maximum likelihood estimates (MLE) of α_j and β.

The relationship between the probability of individual *n* choosing alternative *i* and the representative utility V_{ni}, is illustrated in figure 3.1 below.

Figure 3.1



It can be seen from the diagram that a change in representative utility will have the greatest impact on the probability of choosing an alternative when the probability is initially around 0.5 (where the curve is steepest). From a policy perspective this means that a quality improvement, price reduction etc. will be most effective when there is an initial 50-50 chance of the alternative being chosen. If the representative utility of an alternative is initially very high or very low compared to other alternatives, however, a small change in utility will not have a great impact on the probability of its being chosen.

The multinomial logit model has many advantages: it ensures that the probability of choosing an alternative always lies between 0 and 1 and that the probabilities of all available alternatives always sum to 1 (as opposed to the linear probability model, see Greene, 2003a). Furthermore, when the model takes the form of (3.2) its log-likelihood function is globally concave in the coefficients α_j and β, which simplifies numerical optimisation.

However, because the unobserved component of utility is assumed to be IID over individuals, alternatives and time, the logit probabilities also exhibit some fairly restrictive properties. The most prominent is the independence from irrelevant alternatives property (IIA) first described by Luce (1959), which states that the probability ratio of two alternatives is independent of the other alternatives available to the individual. It follows from the IIA property that if an attribute of some alternative improves, the cross elasticity with respect to this attribute is the same for all other alternatives. For example, if the price of alternative i decreases this is expected to increase the probability of an individual choosing i and decrease the probability of him choosing a different alternative, j. When the IIA property holds the probability of choosing either of the other alternatives decreases by the *same percent* for all $j \neq i$. This implies a substitution pattern that might not always be realistic. Consider for example a situation where the government wishes to introduce a policy to reduce the reliance on petrol for cars and there are three kinds of vehicles: large petrol cars, small petrol cars and electric cars.[3] Under current conditions the probabilities that a household will choose each of these vehicles are 0.66, 0.33 and 0.01 respectively. Suppose a subsidy on electric cars raises the probability for the electric car. The logit model would predict the probability for the petrol cars to drop by the same percent, so that the increase in electric cars comes twice as much from large petrol cars as from small petrol cars. This pattern of substitution is clearly unrealistic, since one would expect an increase in (small) electric cars to draw more from small petrol cars than large petrol cars. The logit model will thus over-predict the petrol savings resulting from the subsidy.

The IIA property is a consequence of the assumption that the errors are IID (uncorrelated) over alternatives. As mentioned above the multinomial logit model specification assumes that the unobserved part of utility is also IID over time periods. If each individual is observed making several choices over a period of time or the data is the result of a stated preference (SP) experiment where repeated choices are made (see chapter 4 for a discussion), some of the unobservable variables that enter the individuals utility function may be correlated over time (choices). In these cases the errors are not IID and hence the multinomial logit model is an inappropriate specification.

Because of the restrictive properties of the multinomial logit model, alternative econometric models that are consistent with random utility maximisation have been developed. We will present two of them in sections 3.2 and 3.3: the nested logit model and the mixed logit[4] model. In the following sub-sections it is described how the MNL model can be used to evaluate the change in demand for a mode following a change in one or more explanatory variables, and how the commuters' welfare is affected by such a change.

[3] This example is due to Train (2003)

[4] Also called error components logit, random parameters logit and random coefficients logit

3.1.2 Marginal effects and elasticities in the multinomial logit model

It is useful from a policy perspective to know how changes in the levels of attributes influence the probability of an alternative being chosen. In a choice model which is linear in the alternative attributes the estimated coefficients are the marginal ***utilities*** of an attribute, $\frac{\partial V_{ni}}{\partial z_{ni}} = \beta_z$, where z_{ni} is the attribute of alternative ***i*** as faced by individual ***n***. The marginal utility measures how individual ***n***'s utility of choosing alternative ***i*** changes in response to a unit increase in attribute z_{ni}. This is not the same as the marginal ***probability***, or the change in the probability of an alternative being chosen following a unit increase in the attribute. In the multinomial logit model, the change in the probability of individual ***n*** choosing alternative ***i*** following a unit increase in z_{ni} is given by:

$$\frac{\partial P_{ni}}{\partial z_{ni}} = \mu \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni}) \tag{3.5}$$

It can be seen from (3.5) that the magnitude of the marginal probability depends on P_{ni} which is determined by the estimated coefficients and the initial attribute levels. Specifically, $\frac{\partial P_{ni}}{\partial z_{ni}}$ is highest when $P_{ni} = 0.5$ and becomes smaller as P_{ni} approaches zero or one. This is directly related to figure 3.1 above: the effect of a change in the level of an attribute is highest when there is an initial 50 percent chance of the alternative being chosen. It should be noted that the marginal effect is negative given that the marginal utility of the attribute is negative ($\frac{\partial V_{ni}}{\partial z_{ni}} < 0$).

The size of the marginal probabilities depends on the units in which the attribute is measured. An alternative "unit free" statistic is the ***elasticity***, which measures the percentage change in the probability of individual n choosing alternative i following a 1 percent increase in z_{ni}. The elasticity is given by:

$$E_{iz_{ni}} = \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}} = \mu \frac{\partial V_{ni}}{\partial z_{ni}} z_{ni}(1 - P_{ni}) \tag{3.6}$$

It may also be of interest to know how the probability of individual n choosing alternative i changes in response to a change in the level of an attribute relating to another alternative (j). The ***cross elasticity***, which measures the change in the probability of individual n choosing alternative i following a percentage increase in z_{nj}, is given by:

$$E_{iz_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} = -\mu \frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj} \tag{3.7}$$

The cross elasticity is positive given that the marginal utility of the attribute is negative ($\frac{\partial V_{ni}}{\partial z_{ni}} < 0$). It can be seen from (3.7) that in the MNL model the cross elasticity is equal for all $i \neq j$. This is a manifestation of the independence from irrelevant alternatives property described earlier.

3.1.3 Welfare analysis and the multinomial logit model

It is often of interest to the researcher to determine how the welfare of one or more individuals is affected by a change in commuting conditions. McFadden (1981) shows that the expected utility of making a choice between a set of alternatives is given by the log of the denominator in (3.1):

$$E(MaxU_{nj}) = \ln(\sum_j e^{\mu V_{nj}}) \tag{3.8}$$

Ben-Akiva and Lerman (1979) point out that in the mode choice context (3.8) can also be interpreted as a measure of the individual's accessibility to the work location.

Williams (1977) shows that since the MNL model can be viewed as a demand function for a given alternative, the difference in consumer surplus following a change in commuting conditions can be calculated as the difference in expected utility evaluated at V^1 (the initial representative utility) and V^2 (the representative utility after the change) such that:

$$\Delta CS_n = \ln(\sum_j e^{\mu V_{nj}^2}) - \ln(\sum_j e^{\mu V_{nj}^1}) \tag{3.9}$$

A problem with William's measure of consumer surplus is that it cannot be used to compare changes in welfare across model specifications. An alternative measure of consumer welfare is ***compensating variation***, or the amount of money an individual needs to receive (or give up) following a change in her utility in order to be equally well off as before the change (see Varian, 1992). Since the welfare change in this case is measured in real units (money), the measure can be used to compare changes in welfare across model specifications. Small and Rosen (1981) shows that the compensating variation can be derived from an MNL model by multiplying (3.9) by the inverse of the marginal utility of income such that:

$$CV_n = -\frac{1}{\lambda}\left[\ln(\sum_j e^{\mu V_{nj}^2}) - \ln(\sum_j e^{\mu V_{nj}^1})\right] \tag{3.10}$$

where λ is the marginal utility of income. In practise the marginal utility of income can be calculated as the absolute value of the cost coefficient in the model.

3.2 The nested logit model

3.2.1 Nested logit choice probabilities

As discussed in section 3.1 one of the main drawbacks of the multinomial logit model is the independence from irrelevant alternatives property. The nested logit (NL) model relaxes the IIA property by dividing the choice alternatives into different subsets or nests, allowing the IIA property to hold within each nest but not across nests. In other words, the ratio of the probabilities of two alternatives in different nests may depend on the attributes of the other alternatives in these two nests. The ratio of the probabilities of two alternatives in the same nest, however, will not depend on the attributes of the other alternatives.

Figure 3.2. An example of a nested logit decision tree

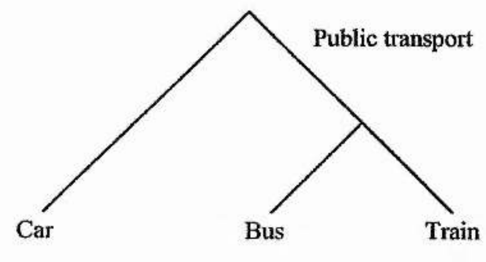


Figure 3.2 above is an example of a Nested Logit decision "tree". The alternatives that are likely to be close substitutes (bus and train) are specified to belong to the same nest. By relaxing the IIA property, the cross elasticities with respect to bus frequency are allowed to differ between the car and train modes. Hence the nested structure in

figure 3.2 accounts for the *a priori* belief that an increase in the probability of n choosing bus comes more from train than car.

In order to give a more formal description of the nested logit model it is conceptually helpful to divide the representative utility function into two parts: one which varies between nests but not between alternatives within a nest, W_{nk}, and one which varies between alternatives within the nest, Y_{nj}. The utility individual n derives from choosing alternative j belonging to nest B_l is thus given by:

$$U_{nj} = W_{nl} + Y_{nj} + \varepsilon_{nj} \tag{3.11}$$

If the income of an individual is thought to influence her choice between private or public transport but not the choice between bus and train, for example, the income variable would enter W_{nl} rather than Y_{nj}. The cost of the bus and train modes on the other hand would enter Y_{nj} since it is relevant for the choice between the modes. It should be noted that it is not uncommon to have all explanatory variables enter Y_{nj}, since they may all be thought to influence the choice between alternatives within nests. Since $Y_{nj} = V_{nj} - W_{nl}$ for any W_{nl}, however, (3.11) is a fully general specification (Train, 2003). The decomposition of representative utility is particularly useful when modelling multidimensional choices such as in a joint car ownership and mode choice model (see the next section and chapter 5). In this case the variables relating to the car ownership decision would be specified to enter W_{nl} while the variables influencing mode choice enter Y_{nj}.

McFadden (1978a) shows that if the unobserved components of the random utility function, ε_{ni}, are assumed to be distributed according to a particular generalised extreme value (GEV) distribution, the probability that individual n chooses alternative i belonging to nest B_k is given by:

$$P_{ni} = P_{ni|B_k} P_{nB_k} \tag{3.12}$$

where,

$$P_{nB_k} = \frac{e^{\frac{1}{\mu}(\lambda_k I_{nk} + W_k)}}{\sum_l e^{\frac{1}{\mu}(\lambda_l I_{nl} + W_l)}} \tag{3.13}$$

$$P_{ni|B_k} = \frac{e^{\frac{1}{\lambda_k} Y_{ni}}}{\sum_{j \in B_k} e^{\frac{1}{\lambda_k} Y_{nj}}} \tag{3.14}$$

and,

$$I_{nl} = \ln \sum_{j \in B_l} e^{\frac{1}{\lambda_l} Y_{nj}} \tag{3.15}$$

In words the probability of choosing alternative i in nest B_k equals the marginal probability of choosing nest B_k multiplied by the conditional probability of choosing alternative i given that B_k is chosen. The forms of the marginal and the conditional probabilities are both multinomial logit, and the nested logit model is therefore the product of two multinomial logit models.[5]

The key feature of the nested logit model is that the scale of the multinomial logit models in equations (3.13 - 3.14) are allowed to differ. If the scale factors of the conditional model, $\frac{1}{\mu}$, and the marginal models, $\frac{1}{\lambda_l}$, all equal 1 the nested logit reduces to the multinomial logit model (hence the multinomial logit model is "nested" within the NL model). Equation (3.15), which is the log of the denominator in (3.14), is often called the "inclusive value" or "log-sum term" (Ben-Akiva, 1972). The

[5] It should be pointed out that the nested logit can have more than two levels. It is straightforward to describe a model with three or more levels using the framework outlined above. For the present purposes, however, the nested logit with two levels will suffice.

product of the scale factor, $\frac{1}{\lambda_l}$, and the inclusive value can be interpreted as the expected utility the individual recieves from choosing nest B_k analogous to the discussion of welfare analysis above.

In the context of the nested-logit model the inverse of the scale factor, λ_l, is often called the *dissimilarity parameter*, since it measures the (dis)similarity between the unobserved portions of utility for alternatives within the same nest. Ben-Akiva and Lerman (1985) show that $1-\left(\frac{\lambda_l}{\mu}\right)^2$ is a measure of the degree of correlation among the unobserved portions of utility for alternatives in nest B_l. Thus, when the dissimilarity parameter equals 1 the degree of correlation between the alternatives in a nest is zero (and if this is the case in all K nests the nested logit model reduces to the multinomial logit model as discussed above).

As in the multinomial logit model one of the scale parameters must be normalized to 1 for identification purposes. It is common to impose the normalisation on the scale parameter of the upper (marginal) model such that $\mu = 1$, and this normalisation will be used in the following discussion.[6] Daly and Zachary (1978) and McFadden (1978b) show that the nested logit model is globally consistent with utility maximisation if:

$$0 < \lambda_l \leq 1 \qquad \text{for all } l \in K \qquad (3.16)$$

Börch-Supan (1990) argues that this condition is unnecessarily strong given that the NL model should be viewed as a local approximation. Based on the work of Börch-Supan, Herriges and Kling (1996) and Gil-Moltó and Hole (2004) derive necessary

[6] Koppelmann and Wen (1997), Hunt (2000) and Hensher and Greene (2002) give an overview of alternative normalisations of the nested logit model, including the so-called non-normalised nested logit model (Daly, 1987).

conditions for local consistency with utility maximization for two-level and three-level NL models respectively.

3.2.2 Elasticities in the nested logit model

In the nested logit model the direct elasticity, or the change in the probability of individual n choosing alternative $i \in B_k$ following a percentage increase in z_{ni}, is given by:

$$E_{iz_{ni}} = \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}} = \frac{\partial V_{ni}}{\partial z_{ni}} z_{ni} \left[\frac{1}{\lambda_k}(1 - P_{ni|B_k}) + P_{ni|B_k} - P_{ni} \right] \tag{3.17}$$

The expression reduces to the MNL direct elasticity if $\lambda_k = 1$, illustrating that the NL model reduces to the MNL model when the dissimilarity parameters equal 1.

The nested logit cross elasticities are of special interest since they illustrate the flexibility of the nested logit model to incorporate a wide rage of substitution patterns. It is important to distinguish between the cross elasticity of an alternative belonging to the same nest as the alternative which attribute is increasing and the cross elasticity of an alternative belonging to a different nest. The change in the probability of individual n choosing alternative $i \in B_k$ following a percentage increase in z_{nj}, $j \in B_k$, is given by:

$$E_{iz_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} = -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} \left[P_{nj|B_k} (\frac{1}{\lambda_k} - 1) + P_{ni} \right] \tag{3.18}$$

which, as with the direct elasticity, reduces to the MNL cross elasticity if $\lambda_k = 1$.

Is interesting to note that both the direction and magnitude of the cross elasticity depends on the correlation between the random utility of alternatives within nest B_k, which can be derived from the dissimilarity parameter. Given that $\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} > 0$, an increase in an attribute belonging to alternative j will lead to a decrease in the probability of alternative i being chosen as long as the dissimilarity parameter of a nest is low relative to the marginal probability of choosing nest B_k such that:

$$\lambda_k \leq \frac{1}{1 - P_{nB_k}} \tag{3.19}$$

Intuitively the effect of the change in z_{nj} is two-fold: it will influence both the marginal probability of choosing nest B_k and the probability of choosing alternative i given that nest B_k is chosen. Consider for example an increase in the fare of the bus service in figure 3.2. This is likely to decrease the probability of an individual choosing public transport since the expected utility of travelling by public transport has fallen. On the other hand the probability that train is chosen given the choice of travelling by public transport is likely to increase. Thus the total effect is determined by the relative strengths of these two effects. The relative strength is determined by the degree of correlation between the utility of the alternatives in the nest since the higher the correlation the stronger the latter effect will be.

The change in the probability of individual n choosing alternative $i \notin B_k$ following a percentage increase in z_{nj}, $j \in B_k$, is given by:

$$E_{iz_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} = -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj} \tag{3.20}$$

which is the same as the cross elasticity in the MNL model. It is easy to see that the cross elasticity differs when i and j belong to different nests. In this case there is only one effect: the change in the probability of choosing the nest the alternative belongs to. The probability of choosing the alternative given that its nest is chosen remains unchanged. In the previous example, an increase in the bus fare would lead to a decrease in the probability of going by public transport and thus an increase the probability of going by car. Furthermore, it can be shown that, in the case of a nested logit model which satisfies the condition for global consistency with utility maximisation (3.16), the cross elasticity of an alternative that belongs to the same nest as alternative j will always be greater than or equal to the cross elasticity of an alternative that belongs to an alternative in a different nest (see appendix 3.1 for a proof). In other words an increase in the bus fare will lead to a greater (relative) increase in the share of people going by train than people going by car. The fact that the cross elasticity differs between alternatives belonging to different nests illustrates that the IIA property does not hold in the NL model.

It can be shown (McFadden, 1981) that a discrete choice model is consistent with utility maximisation *if and only if* all cross elasticities are non-positive (this is a necessary but not sufficient condition for consistency with utility maximisation).[7] In the case of the nested logit model this condition is always satisfied when equation (3.16) holds, as can be seen by substituting $P_{nB_k} = 0$ into equation (3.19). For values of P_{nB_k} higher than zero, however, the cross elasticity is also negative for values of λ_k higher than 1. How much higher λ_k can be without violating the utility maximising condition is determined by P_{nB_k} as well as the number of alternatives in nest B_k (see Börch-Supan, 1990; Herriges and Kling, 1996; Gil-Moltó and Hole, 2004).

[7] Given that $\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} > 0$.

3.2.3 Multidimensional choices and the nested logit model

The nested logit model is well suited to model choices that have two or more dimensions. The joint car ownership/ mode choice models in Ben-Akiva and Atherton (1977), Train (1980) and Thobani (1984) and the trip chaining/ mode choice model in Hensher and Reyes (2000) (see chapter 5) are examples of models where choices that are interrelated are modelled together using a nested structure. Another application of the nested logit to model multidimensional choices is the joint household location/ mode-choice model in Anas and Chu (1984).

The rationale behind the multidimensional models is that choices that are made simultaneously should be modelled simultaneously. For example, as Train (1980) argues, a mode choice model in which car ownership is included as an exogenous variable may be misspecified since households are likely to take commuting into account when deciding how many cars to own (which makes car ownership endogenous to the mode choice decision). As a result of this misspecification the parameters of the model may be biased and hence the forecasts produced by the model will be incorrect.

The theoretical framework of the multidimensional models is fundamentally the same as in the one-dimensional nested logit model outlined in section 3.3. A further complication in the multidimensional models is that decisions made at the household level (household location, car ownership) are mixed with those usually assumed to be made by the individual (mode choice). In practice this is dealt with by including variables relating to the household (income, household size) in the utility functions of alternatives in choices assumed to be made on the household level, and variables relating to the individual (age, gender) in the utility functions of alternatives in choices assumed to be made by the individual.

3.3 The mixed logit model

3.3.1 Mixed logit choice probabilities

As described in the previous section, the nested logit model relaxes the restrictive independence from irrelevant alternatives property by allowing for correlation between alternatives within a nest. The mixed logit (ML) model extends the flexibility to model a non-IIA substitution patterns even further by allowing *all* alternatives available to the individual to be correlated. Furthermore, the mixed logit model relaxes the restriction that all the individuals in the sample have the same tastes by allowing the coefficients to vary randomly in the population.

Following Brownstone and Train (1999) and using the same notation as above, the utility function is denoted $U_{ni} = \beta' x_{nj} + \lfloor \eta_{nj} + \varepsilon_{nj} \rfloor$, where η_{nj} is a random term with zero mean whose distribution over alternatives and people depends on underlying parameters and observed data relating to individual n and alternative j.[8] As in the multinomial logit model ε_{ni} is assumed to be IID extreme value, with variance normalised to $\frac{\pi^2}{6}$ to set the scale of utility, while η_{nj} is distributed with density $f(\eta_{nj}|\theta)$ where θ are the fixed parameters of the distribution (such as the mean and

[8] To simplify the notation socio-demographic variables are omitted from the utility function.

variance of η_{nj}). It can be seen that when η_{nj} is zero for all individuals/ alternatives, the mixed logit model reduces to the multinomial logit model.

The probability of person ***n*** choosing alternative ***i conditional*** on knowing η_{nj} is given by:

$$L_{ni}(\eta_{ni}) = \frac{e^{\beta' x_{ni} + \eta_i}}{\sum_j e^{\beta' x_{nj} + \eta_j}} \tag{3.23}$$

which is the standard logit formula. However, the researcher does not know η_{nj}, and the ***unconditional*** probability of person ***n*** choosing alternative ***i*** is given by integrating the logit formula over all values of η_{nj}:

$$P_{ni}(\theta) = \int L_{ni}(\eta_{ni}) f(\eta_{ni}|\theta) d\eta_{ni} \tag{3.21}$$

The mixed logit probability is thus a weighted average of the logit formula evaluated at different values of η_{ni}, with the weights given by density f. This expression cannot be solved analytically, and is therefore approximated using simulation methods. The algorithm used to obtain the maximum simulated likelihood (MSL) estimates can be described as follows:

1) Set starting values for the parameters of the distribution of η_{nj} (in case of a normally distributed coefficient the mean and the variance).
2) Draw values of η_{nj} from this distribution for each person/ alternative and use these values to calculate the log-likelihood function.
3) Repeat step 2) ***r*** times, obtaining ***r*** values for the likelihood function, ***LL***.

4) Compute the average likelihood function, $\frac{\sum_{i=1}^{r} LL_i}{r}$, which is the simulated value for the likelihood.

5) Change the coefficients and the parameters of the distribution of η_{nj} and repeat steps 2) – 5) until a maximum is found. The parameter values that maximises the log-likelihood function are the MSL estimates of the true parameters of the distribution.

Lee (1992) and Hajivassiliou and Ruud (1994) derive the asymptotic distribution of the MSL estimator and show that under regularity conditions the estimator is efficient and asymptotically normal.

3.3.2 Taste Variation

The specification of η_{nj} depends on whether the use of the mixed logit model is motivated by allowing for a flexible (non IIA) substitution pattern or by the flexibility to model random taste variation. If the model is motivated by the flexibility to model random taste variation in the population, utility is given by $U_{nj} = \beta' x_{nj} + \mu'_n x_{nj} + \varepsilon_{nj}$, where μ_n is a vector of coefficients for person n which represents to what extent her tastes deviates from the average tastes in the sample. It can be seen that this utility specification is consistent with the more general utility specification above given that $\eta_{nj} = \mu'_n x_{nj}$.

In most applications of the mixed logit model $f(\eta_{nj}|\theta)$ has been specified to be normal or lognormal (Revelt and Train, 1998; Walker *et al.*, 2003) but other distributions (triangular, uniform) have also been used (Hensher and Greene, 2001). The lognormal distribution is useful when a coefficient is restricted to have a specific sign for all decision-makers, such as a negative price coefficient (see the discussion in chapter 5).

3.3.3 Error components

If the mixed logit model specification is motivated by the flexibility to specify a more complex error structure rather than modelling taste variation in the population, the utility function is given by $U_{nj} = \beta' x_{nj} + \mu'_n z_{nj} + \varepsilon_{nj}$. The error component, $\eta_{nj} = \mu_n' z_{nj} + \varepsilon_{nj}$, is correlated over alternatives if z_{nj} is non-zero. Again this utility specification is consistent with the more general specification of utility given that $\eta_{nj} = \mu'_n z_{nj}$. It should be pointed out that the two specifications of the utility function are formally equal and differ only in interpretation (in the case of $z_{nj} = x_{nj}$ they are the same).

It can be shown (Brownstone and Train, 1999) that the error covariance of the mixed logit model is given by $\mathrm{cov}(\eta_{ni}, \eta_{nj}) = E(\mu_n' z_{ni} + \varepsilon_{ni})(\mu_n' z_{nj} + \varepsilon_{nj}) = z_{ni}' W z_{nj}$ where W is the covariance of μ_n. Different specifications of f and z_{nj} lead to different patterns of correlation and hence different substitution patterns. Entering dummy variables for two or more alternatives, for instance, generates correlation between those alternatives, analogous to the nested logit model (a mixed logit analogue to the nested logit example in the previous section could be specified by entering a dummy variable that equals one in the utility function for the bus and train modes and zero in the utility function for the car mode).[9] Importantly, it has been shown that any random utility model can be approximated to any degree of accuracy by the mixed logit model through appropriate specification of the distribution of the random parameters and the explanatory variables (McFadden & Train, 2000), a testimony to its virtually limitless flexibility.

3.3.4 Panel Data

The framework described above can be extended to panel data, which can either be revealed preference (RP) data where each individual is observed making several choices over a period of time or stated preference (SP) data where repeated choices are made (see chapters 4 and 7). In this case utility is given by $U_{njt} = \beta' x_{njt} + \eta_{nj} + \varepsilon_{njt}$ where t denotes the time period (choice situation). Since ε_{njt} is IID over time periods, the probability that an individual chooses a particular sequence of alternatives is given by the weighted average of the product of the logit probabilities evaluated at different values of η_{nj} (Train, 1998):

$$P_n(\theta) = \int S_n(\eta_{ni}) f(\eta_{ni}|\theta) d\eta_{ni} \tag{3.24}$$

where,

$$S_n(\eta_{ni}) = \prod_t L_{ni(n,t)t}(\eta_{ni}) \tag{3.25}$$

and,

[9] An overview of a variety of possible error structures, along with conditions for identification of the models, is given in Walker (2001) and Walker *et al.* (2003).

$$L_{ni(n,t)t}(\eta_{ni}) = \frac{e^{\beta' x_{nit} + \eta_{ni}}}{\sum_j e^{\beta' x_{njt} + \eta_{nj}}} \tag{3.26}$$

Here S_n is the probability of individual n making her observed sequence of choices and $L_{ni(n,t)t}$ the probability of individual n making her actual choice in period t. Since (3.24) cannot be solved analytically, it is solved using simulation methods as described earlier.

Since η_{nj} is constant over time for each person/ alternative, this specification generates correlation over choices made by the same individual in different time periods. The correlation is not perfect, however, since the error term also includes ε_{njt} which is IID over individuals, alternatives and time periods. For applications of the mixed logit to model panel data see Train (1998), Revelt and Train (1998) or Algers *et al.* (1998).

3.3.5 Identification of the Mixed Logit model

While the conditions for identification of the multinomial and nested logit models are well-known, identification of the mixed logit model is still an unresolved issue in the literature, with the exception of the special case of a ML model with normally distributed error components. Based on the identifying conditions for the multinomial probit model given in Bunch (1991), Walker (2002) and Walker *et al.* (2003) show that the identifying conditions in this case are given by:

1) The order condition. The upper limit on the number of estimable parameters in the variance-covariance matrix is given by $\frac{J(J-1)}{2}-1$, where J is the number of alternatives in the model. This is a necessary but not sufficient condition for identification.

2) The rank condition. The upper limit on the number of estimable parameters in the variance-covariance matrix is given by the rank of the Jacobian of the variance-covariance matrix for the ***error differences*** minus one (Train, 2003, describes a straightforward procedure for calculating the variance-covariance matrix for the error differences). If this condition is satisfied - the number of parameters in the specified variance-covariance matrix is lower than or equal to the rank of the Jacobian of the variance-covariance matrix for the error differences minus one - no further restrictions are needed.

3) The equality condition. Given that the rank condition implies that one or more of the parameters in the variance-covariance matrix must be restricted, the equality condition must also be satisfied. In short, this condition states that the probabilities of the normalised model must equal the probabilities of the unrestricted model (see Walker ***et al.***, 2003 for details).

The question remains, however, of which parameter(s) in the variance-covariance to restrict in order to ensure consistency with the equality condition when a restriction is necessary (beyond the customary normalisation of the scale parameter) and to which value the parameter(s) should be restricted. Walker *et al.* argue that the parameters should be restricted to equal zero, since this ensures that the MNL model is a special case of the ML model, which facilitates the use of nested hypothesis tests (such as the LR test). When it comes to the choice of parameter to restrict, Walker *et al.* show that this is not necessarily arbitrary, and suggests ways of identifying the parameter(s) that should be restricted to ensure that the equality condition holds. Since this issue must be evaluated on a case for case basis depending on the model structure, it will be addressed in the sections of the thesis where it becomes relevant.

3.4 Goodness of fit measures and hypothesis testing

3.4.1 Goodness of fit measures

The likelihood ratio index, $\rho_0^2 = 1 - \frac{LL(\beta)}{LL(0)}$, where $LL(\beta)$ is the value of the log likelihood function at the estimated parameters and $LL(0)$ is its value when all the parameters are set equal to zero, is a common summary measure of the goodness of fit of a discrete choice model.[10] The value of the likelihood ratio index will always be between 0 and 1, where a value higher than zero indicates that the estimated model fits the data better than the model where all the parameters equal zero. It is important to note that although ρ_0^2 will lie between 0 and 1, in contrast to the R^2 statistic a perfect fit would give a value of about 0.7 while a value higher than 0.2 can be considered a good fit (see Hensher, 1979). An alternative, but very similar goodness of fit measure is given by $\rho_c^2 = 1 - \frac{LL(\beta)}{LL(c)}$ where $LL(c)$ is the value of the log likelihood function with alternative specific constants only.[11] It should be noted that while the R^2 statistic used in regression analysis is a measure of the explained variation in the dependent variable, the log likelihood has no such intuitive interpretation. It is simply a measure of the percentage increase in the log likelihood function above the value taken at zero parameters (or with alternative specific constants only). As a consequence, the likelihood ratio index cannot be used to compare the fit of models estimated using different samples. It is, however, a useful statistic in comparing the fit of different models estimated on the same sample.

The rho-bar squared statistic, $\bar{\rho}^2 = 1 - K\left[\frac{LL(\beta)}{LL(0)}\right]$, is an analogue to the $\bar{R}^2$ statistic used in regression analysis. $K = \frac{\sum_n J_n - N}{\sum_n J_n - N - k}$, where J_n is the number of alternatives in individual n's choice set, N is the sample size and k is the number of coefficients in the model. Analogous to the $\bar{R}^2$ statistic, the $\bar{\rho}^2$ statistic penalises the fall in degrees of freedom when adding explanatory variables to the model. Hence, if a variable is added that does not increase the model's explanatory power $\bar{\rho}^2$ falls.

[10] This measure is also called the rho-squared statistic

[11] In the case of the multinomial logit model, when alternative specific constants are included in the model specification the average probability that an alternative is chosen equals the observed share of that alternative in the sample. If there are no coefficients in the utility functions the probability that an alternative is chosen equals 1 divided by the number of alternatives available.

3.4.2 Hypothesis testing

The likelihood ratio (LR) test is often used to test restrictions on the coefficients in a model. The LR test statistic is given by $2\left[LL^U - LL^R\right]$, where LL^U and LL^R are the values of the log likelihood functions of the unrestricted and restricted models respectively. The LR test statistic can be shown to be asymptotically chi-squared distributed with r degrees of freedom, where r is the number of restrictions imposed. The null hypothesis of "accepting" the restrictions imposed is rejected if the LR statistic exceeds the critical value from the chi-squared distribution with r degrees of freedom at a selected level of significance.

3.5 Forecasting and aggregation

The models described in the previous sections estimate the probability that an individual (or a particular group of individuals sharing the same characteristics) will choose a particular alternative from his or her choice set. Predicting the behaviour of a specific individual, however, is usually of little use in helping make investment or planning decisions. The interest of the researcher is normally (the present thesis included) to predict changes in *aggregate* demand following a change in one or more policy variables (for instance how an increase in the frequency of buses on a given route influences bus ridership) or to predict the demand for a new mode (such as a park and ride service). In this section the most common aggregation method is described; the method of sample enumeration. For a more complete review of aggregation methods for discrete choice models see Ben-Akiva and Lerman (1985).

The market share, or the share of the population choosing a given alternative, is given by averaging the sum of the individual probabilities such that:

$$\hat{S}_i = \frac{1}{N}\sum_{n=1}^{N} P_{ni} \tag{3.26}$$

where $\hat{S}_i$ is the estimated market share of mode i. It is a well-known result (see Ben-Akiva and Lerman, 1985 for a proof) that the MNL model will reproduce the market shares in the estimation sample such that:

$$\hat{S}_i = \frac{1}{N}\sum_{n=1}^{N} y_{ni} \tag{3.27}$$

where y_{ni} equals 1 if individual n is observed to choose alternative i and 0 otherwise.

The method of sample enumeration can also be applied to calculate aggregate elasticities. Greene (2003b) argues, however, that this may lead to implausibly high elasticity estimates if there for some reason exists one or more observation in the sample with an extreme configuration of attributes. An alternative to sample enumeration is to weight the elasticity by the probability of the alternative being chosen such that:

$$E_{iz_i} = \frac{P_{ni}}{\sum_n P_{ni}} E_{iz_{ni}} \tag{3.27}$$

where E_{iz_i} is the probability weighted aggregate elasticity. The weighting scheme will offset the extreme effect given that the implausibly high elasticity estimate has a low probability. Greene (2003b) argues that this aggregation method produces reasonable elasticity estimates in almost all cases.

3.6 Heteroscedasticity

Following Munizaga *et al.* (2000), the issue of heteroscedasticity in discrete choice models can be divided into two categories:

1) Heteroscedasticity between alternatives

2) Heteroscedasticity between observations

Heteroscedasticity between alternatives arises when variations in the representative utility function explain the variations in utility of some alternatives better than others. In this case the utility of the latter alternatives will have a larger degree of 'randomness', which is represented by a higher error variance in the utility function of those alternatives. This may be the case, for instance, when individuals have less information about some alternatives than others. Heteroscedasticity between observations may arise when multiple data sources are used to calibrate the model, where both data sources contain the same options but the error variance of one data source is higher. It will also arise if the representative utility functions explain better the variations in utility for some socio-economic groups (for example, blue-collar workers may be more responsive to changes in the observable alternative attributes than white-collar workers, while white-collar workers may be more concerned with immeasurable attributes like status). This section will only deal with the issue of heteroscedasticity between alternatives. The special case of heteroscedasticity arising when combining Revealed and Stated Preference data is discussed in chapter 4.

Among the models presented so far in this chapter only the Mixed Logit model can represent heteroscedasticity, since both the MNL model and NL models are homoscedastic models by definition (the error variances are assumed to be the same for all alternatives and observations). In the ML model, however, heteroscedasticity between alternatives can be represented in various ways, either by making the coefficients for alternative specific attributes random or by entering dummy variables for the alternatives as error components in the model. Perhaps the more straightforward way to specify a heteroscedastic ML model is to enter a dummy variable for each alternative as an error component as suggested by Walker *et al.* (2003). The authors point out, however, that this model is not identified unless the coefficient for one of the error components is constrained to equal zero[12]. Furthermore, in the case of the heteroscedastic ML model it is not arbitrary which coefficient is normalised to zero, as different normalisations result in different estimation results. Walker *et al.*, (2003) show that the correct normalisation can be identified by either estimating the full (unidentified) model in order to identify the smallest element of the variance covariance matrix and subsequently re-estimating the model constraining this element to zero or by estimating J models (where J is the number of alternatives), setting the coefficient for each alternative to zero in turn and choose the model with the highest log-likelihood. If the first approach is feasible it is obviously the least time consuming method. It is possible, however, that problems of convergence can arise when estimating the unrestricted model, in which case the second approach is the only feasible alternative.

It should be pointed out that there are other discrete choice models than the Mixed Logit model that can represent heteroscedasticity, of which the Heteroscedastic

[12] In principle it can be restricted to equal any constant, but normalising to zero ensures that the MNL

Extreme Value model (Bhat, 1995) and the Multinoial Probit model (Daganzo, 1979) are the most prominent. Since the Mixed Logit model can essentially reproduce the error pattern of these models, however, (the heteroscedastic ML model described above is conceptually equal to the Heteroscedastic Extreme Value model for instance) these models will not be discussed further here.

3.7 Value of time estimation

As described in chapter 2 the value of time subjective value of time (*SVOT*) is given by the rate of substitution between the time and cost of travelling by a given mode. It is straightforward to show that in the case in which the representative utility is specified to be linear in the attributes, the subjective value of time is given by the ratio of the time and cost coefficients in the model:

$$SVOT = \frac{\beta_T}{\beta_C} \tag{3.28}$$

Armstrong *et al.* (2001) point out that since β_T and β_C are estimators of the true time and cost coefficients, the computed *SVOT* is also an estimator with a certain probability distribution, which is different from the distribution of β_T and β_C. Also, since β_T and β_C can be shown to be distributed asymptotically normal (Ben-Akiva and Lerman, 1985), the ratio of the two coefficients follow a probability distribution which is unknown *a priori*. It can be shown that in the case where the correlation between the two coefficients equals zero the value of time is Cauchy distributed, but

model is nested within the ML model (see also section 3.3.4).

this distribution is unstable since it has an indefinite variance and its mean does not have an analytical expression. If the correlation is non-zero the distribution is also unstable and even more complex. Therefore Armstrong *et al.* (2001) argue that a procedure for making statistical inference on the ratio of the time and cost coefficients should not resort to direct use of the PDF of the ratio of the coefficients, but rather the probability distribution of the coefficients themselves.

Based on the findings in Garrido and Ortuzar (1993), Armstrong *et al.* (2001) show that when the indirect utility function is linear in the time and cost variables the upper and lower bounds of the confidence interval for *SVOT* can be calculated as:

$$V_{S,I} = \left(\frac{\beta_t}{\beta_c}\frac{t_c}{t_t}\right)\left(\frac{t_t t_c - \rho t^2}{(t_c^2 - t^2)}\right) \pm \left(\frac{\beta_t}{\beta_c}\frac{t_c}{t_t}\right)\frac{\sqrt{(\rho t^2 - t_t t_c)^2 - (t_t^2 - t^2)(t_c^2 - t^2)}}{(t_c^2 - t^2)} \quad (3.29)$$

where t_c and t_t are the t-statistics for the time and cost coefficients respectively, t is the critical value given the required confidence limit and ρ is the correlation between the time and cost coefficients. Equation (3.29) has some interesting properties: firstly it should be noted that the confidence interval is not symmetrical with respect to the point estimate of *SVOT*. Secondly, it can be seen that the higher the correlation between the coefficients, ρ, the tighter the confidence interval. In addition it can be seen that the more significant the coefficients are (as represented by higher values of t_c and t_t) the tighter the confidence interval is. Armstrong *et al.* (2001) also point out that when N and $t_c t_t$ approach infinity, the confidence interval approaches the point estimate of *SVOT*, indicating that a larger sample size leads to a narrower confidence interval.

Armstrong *et al.* (2001) present a number of other ways of calculating the confidence interval for the subjective value of time, some of which are also applicable when the indirect utility function is non-linear. Here only one of those will be presented; the approach Armstrong *et al.* call the method of 'simulation of multivariate normal variates' (MVNS). This method involves taking a large number of draws (1000, say) of the time and cost coefficients given their joint distribution (which is asymptotically normal with the variance and co-variance given by the estimated variance-covariance matrix), and then calculating the value of time for each of these draws. The generated sample can then be used to calculate various statistics such as the mean and variance of the value of time. The upper and lower bounds of a 95% confidence interval can be obtained by calculating the 2.5 and 97.5 percentile points respectively. This approach is essentially the same as the one suggested by Krinsky and Robb (1986) for calculating confidence intervals for elasticities in non-linear models. Hensher and Greene (2003) present a practical way of making draws from any multivariate normal distribution based on draws from the standard normal distribution produced by a random number generator.

Appendix 3.1 Proof of the theorem in section 3.2.2

Proof of the theorem that the cross elasticity $E = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}}$ is higher when alternative *i* and *j* belong to the same nest (E^1) than when *i* and *j* belong to different nests (E^2), given that the global condition for utility maximisation holds.

Theorem:

$E^1 \geq E^2$ for all $0 \leq \lambda_k \leq 1$

where,

$$E^1 = -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} \left[P_{nj|B_k} (\frac{1}{\lambda_k} - 1) + P_{ni} \right] = -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj|B_k} \left[\frac{1}{\lambda_k} - 1 + P_{nB_k} \right],$$

$$E^2 = -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj} = -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj|B_k} P_{nB_k}$$

Proof:

For $E^1 \geq E^2$ it is needed that $\frac{1}{\lambda_k} - 1 + P_{nB_k} \geq P_{nB_k}$

This can be written as $\frac{1}{\lambda_k} - 1 \geq 0$, which holds for all $\lambda_k \leq 1$

Chapter 4

Data and Estimation Procedures

This chapter outlines the differences between the two main types of data used in discrete choice modelling and presents of some of the issues that arise when using data derived from choice experiments, usually called stated preference data, to calibrate a model. Section 4.1 describes the differences between revealed preference (RP) and stated preference (SP) data, section 4.2 outlines the design of a choice experiment, while sections 4.3 – 4.6 give an account of several important issues relating to SP modelling: the "scale" problem, data fusion, testing for fatigue and learning effects and the "repeated measurements" problem.

4.1 Revealed and stated preference data

The data used in discrete choice models can be divided into two main categories: revealed preference (RP) and stated preference (SP) data. Revealed preference data are observations of alternatives actually chosen in the market, and the attributes of the alternatives available to each individual. Stated preference data, on the other hand, are results from a hypothetical *choice experiment* where each individual chooses between alternatives with attributes specified by the researcher.

Although the earlier models of disaggregate travel demand were estimated using revealed preference data, stated preference methods have become increasingly popular in transportation research over the past two decades (see Hensher, 1994 or Ortuzar, 1999 for good introductions to the SP methodology). This is mainly due to the flexibility of the SP experiment to introduce new alternatives and attributes and to incorporate a wider range of attribute levels than what is observed in the market. It can also overcome problems often encountered with RP data such as little variance and/ or multicollinearity in the independent variables and measurement errors. The use of SP data has, however, also been met with much scepticism because of the hypothetical nature of the data. The question is simply how reliable data elicited from a hypothetical choice situation are. It is argued by several practitioners that SP data seem to be reliable given that the experiment is well designed and clearly explained to the respondents (see Louviere *et al.*, 2000). There is also a growing body of evidence of successful use of SP models in forecasting (Beaton *et al.*, 1998; Fowkes and Tweddle, 1999). We give a more detailed account of the strengths and weaknesses of the two data types below.

4.1.1 Revealed preference data

Revealed preference data reflect the choices actually made in the market and therefore have the obvious strength of depicting the current market equilibrium. They will also embody the technological, personal and market constraints that each individual faces. Because of these features RP data are generally regarded as a reliable source of information. Apart from these benefits, however, there are several potential problems with RP data:

1) The observed attributes of the alternatives may have little variance, which makes estimation of their coefficients difficult or impossible.

2) The attributes may be highly collinear, such that it is difficult to estimate their separate effects on the choice variable. This is likely to be the case between modal attributes such as time and cost, since expensive alternatives are likely to be faster than less expensive alternatives (consider the choice between going by car and cycling, for instance). As a result, the estimators of the coefficients may be insignificant (have low t-statistics) even though their total effect might indicate their importance (high rho-bar squared).

3) RP data may suffer from measurement error, especially if the researcher does not directly observe the individuals' choices and alternative attributes (i.e. the data is based on the individuals' self-report).

4) Defining the individuals' choice sets is a difficult task. It has been shown (Williams and Ortuzar, 1982) that a discrete choice model that ignores the problem of choice set generation by assuming that all alternatives are available to every individual in the sample may be seriously misspecified. The common practise in the discrete choice literature is to assume that the individuals' choice sets can be defined deterministically. This approach does not take into account, however, that people will differ in their perceptions of which alternatives are available to them. It is likely, for instance, that some people will be willing to walk further than others. Some papers have explored how such heterogeneities in perception can be accounted for (Swait, 2001; Ben-Akiva and Boccara, 1995 and Swait and Ben-Akiva, 1987), but choice set generation is still a largely unresolved issue in the literature.

5) Because RP data depicts the world as it is, it is not well suited to measure the response to new products and attributes.

4.1.2 Stated preference data

Since Stated preference data is the result of a controlled experiment it can overcome most of the difficulties related to RP data. In particular:

1) The ranges of the alternative attributes can be extended to values not observed in the market.

2) Since the experimental design is controlled by the researcher multicollinearity can be avoided. This will improve the precision of the parameter estimates.

3) There is no measurement error in the data as the attributes are specified by the researcher (there may however be differences in perception of the attribute values).

4) The choice-set is pre-specified by the researcher.

5) The experiment can include attributes and alternatives that do not exist in the market at the present.

SP data also have some benefits that are not directly related to the problems with RP data outlined above. An advantage of the SP methodology is that it is both feasible and common to present several choice tasks to the respondents in the SP survey. As a result each sampled individual provides more information about his or her preferences compared to RP data, which typically consist of one observation per respondent (with the exception of costly travel diaries that follow respondents over a period of time). Thus collecting SP data is in general more efficient than collecting RP data.[1]

A further difference between RP and SP methods is that in the SP methodology there are several ways of eliciting the respondents' preferences. The method that most closely resembles the choice process observed in the market is to instruct the respondent to choose her preferred alternative. This approach is referred to as the stated choice (SC) method in the stated preference literature. Alternative approaches are the rank and rate methods in which the respondents rank the available alternatives or rate the alternatives following a given semantic scale (a typical question would be "On a scale from 1 to 10 how do you rank alternative A?"). The rank and rate methods collect more information about the individuals' preferences compared to the SC method. There are, however, also some drawbacks to the rank and rate approaches (see Willumsen and Ortuzar, 2001 for a discussion). In the SP application in chapter 7 the individuals were asked to choose their preferred option and the following discussion will therefore concentrate on this approach.

As previously mentioned the main concern when it comes to the use of SP data in modelling choice behaviour is that the choices observed are hypothetical. As a consequence, SP data does not in general depict the market equilibrium and cannot easily reflect changes in personal constraints (e.g. work location, income and information availability). Some critics have gone as far as claiming that SP data have no value, since "hypothetical questions result in hypothetical answers". This is clearly an exaggeration given that a growing number of studies focusing on the external validation of SP models (see chapter 5 for a review) suggest that a well-designed SP design can elicit preferences similar to those observed in the market. On the other hand the fact that SP data are a result of a hypothetical choice situation should not be ignored, and the SP questionnaire should be carefully designed in order to reduce the likelihood of bias in the responses. Much effort has been devoted to identify the sources of bias in the choice variable that may be present as a result of the

[1] It should be pointed out, however, that some recent studies have collected RP data using new data collection techniques such as GPS, which makes it possible to collect several observations per respondent given access to GPS technology.

hypothetical choice situation (see, for example, Fowkes and Preston, 1991). The sources of bias include:

1) Policy bias. Respondents may be inclined to answer strategically in order to achieve their desired policy response. The goal of the researcher is to make the experiment sufficiently complex to make it difficult for the respondents to bias their answers in order to influence the results of the study (and hence the policy recommendations derived from the study) in a straightforward manner.

2) Justification bias. Respondents may choose a particular alternative in order to justify their current behaviour. Justification bias is difficult to identify, especially since it resembles choice inertia, which is congruent with actual choice behaviour.

3) Self selectivity bias. It is possible that the characteristics of survey respondents differ from those of the overall sample. This is especially likely to be a problem if the response rate of the survey is low. It should be noted that this type of bias may be present in surveys of all types, not just SP surveys. In the case of an SP experiment designed to forecast the demand for a new service there is clearly more incentive for likely users of the service to respond.

4) Non-commitment bias. The respondents to the survey are not committed to behave in the way that they have responded. This is related to the policy and justification bias discussed above.

In addition the SP experiment might suffer from factors such as learning (learning effects) and boredom (fatigue effects) (McFadden, 1986). In the presence of such effects preferences are unstable over the sequences of choices performed by the individual, which may lead to biased parameter estimates. It has been shown that the likelihood of learning and fatigue effects increases with the complexity of the experiment (Sælensminde, 2001). Consequently there is a trade-off between reducing the likelihood of response bias by making the experiment sufficiently complex, and reducing the potential for learning and fatigue effects by making the design relatively easy to complete. It should be noted that the potential bias due to learning and fatigue effects can be reduced by presenting the choice scenarios to respondents in a randomised order.

In spite of these difficulties the use of stated preference data should not be readily dismissed given its strength in forecasting *changes* in behaviour by incorporating a wider range of attribute levels as well as having the flexibility to introduce new alternatives and attributes. Since it is evident that both RP and SP data have their advantages there has been a growing interest in combining the different types of data to provide more robust parameters for the choice model. We will describe two methods for combining the two data types in section 4.4.

Table 4.1. Comparison of RP and SP data

	RP data	SP data
Preference	Choice behaviour in actual market. Cognitively congruent with actual behaviour	Preference statement for hypothetical scenarios. May be cognitively incongruent with actual behaviour
Alternative s	Actual alternatives. Response to non-existing alternatives are not observable	Generated alternatives. Can elicit preference for new (non-existing) alternatives
Attributes	May include measurement errors Correlated attributes Ranges of attributes are limited	No measurement errors Multicollinearity can be avoided by design Ranges of attributes can be extended
Choice set	Ambiguous in many cases	Pre-specified
Number of responses	Difficult to obtain multiple responses from an individual	Repetitive questioning is easily implemented
Response format	Preference information available is "choice"	Various response formats (e.g., choose one, ranking, rating) are possible

Source: Morikawa, 1994

4.2 The design of a choice experiment

As seen in the previous section, stated preference data is the outcome of a hypothetical choice experiment. In the experiment the researcher defines the attributes and attribute levels of the alternatives and the respondents are asked to choose the alternative they prefer. The observed choices, together with the attributes/ levels in the design are then used to elicit the marginal utilities of the attributes using the discrete choice methodology described in chapters 2 and 3.

In general, if in a choice situation there are A alternative attributes that vary over L levels, the *full factorial design*, which is a matrix of all possible combinations of attribute levels, is given by L^A. The full factorial design grows exponentially with the number of attributes and levels, and in many cases it is not practical for one individual to choose between all possible combinations. In order to reduce the number of alternatives available to the individual the researcher can create a *fractional factorial design*, which consists of a subset of the alternatives in the full factorial. The aim of the researcher is to create a fractional design that satisfies some statistical properties while allowing for estimation of the effects of interest.

Ideally, the fractional design matrix should be *orthogonal*, or in other words the design should not exhibit any degree of collinearity.[2] In addition the design should be *balanced*, meaning that the levels of each attribute appear with equal frequency in the matrix. In practice it is difficult to satisfy both these principles exactly, and the researcher chooses the design that most tends toward orthogonality and balance. This is called the optimal design or the most efficient design.

The size of the fractional factorial design depends on the number of *effects* the researcher wishes to estimate. The simplest fractional design allows for the recovery of all *main effects*, or the effect on the dependent variable following a marginal change in a single attribute holding all other variables constant. A more complex design also allows for the recovery of *interaction effects*, or the effect on the dependent variable following a marginal change in a single attribute given different values of another attribute. The more effects the researcher wishes to estimate the larger the fractional factorial design must be. For linear models it has been shown that the main effects typically account for 70 to 90 percent of explained variance, two-way interactions account for 5 to 15 percent while higher order interactions (including more than two variables) account for the remaining explained variance (Dawes and Corrigan, 1974). Thus, a "rule of thumb" when creating a fractional design is to allow for estimation of main effects and two-way interactions since they account for virtually all of the explained variance (Louviere *et al.*, 2000).

In a fractional factorial designed to estimate a subset of the effects of the full factorial, the included effects will be *aliased* with one or more omitted effects (Louviere *et al.*, 2000). For instance, if the full design has three attributes with two levels and the fractional factorial is designed to estimate the main effects only, the main effect of attribute A will be aliased with the BC interaction. Aliasing implies that it is impossible to disentangle the two effects, and the main effect of A is estimated if and only if the BC interaction is insignificant. Otherwise the estimate is a combination of the main effect of A and the BC interaction. In order to reduce the problem of aliasing as many effects as considered practically feasible should be incorporated in the fractional factorial.

---

[2] A matrix X is orthogonal if $X'X = I$, where I is the identity matrix with ones along the main diagonal and zeros elsewhere. In an orthogonal matrix all column vectors are orthogonal (their scalar product is 0) (See e.g. Sydsaeter and Hammond, 1995).

4.3 Forecasting using stated preference models. The "scale" problem.

Recall from chapter 3 that in the multinomial logit model the probability that individual n chooses alternative i from a set of J alternatives is given by:

$$P_{ni} = \frac{e^{\frac{1}{\mu}V_i}}{\sum_j e^{\frac{1}{\mu}V_j}} \tag{4.1}$$

where $\frac{1}{\mu}$ is a positive scale parameter and V_j is the representative utility function described in chapter 2. As mentioned in chapter 3 it is customary to normalize the scale parameter to unity since it cannot be identified in estimation. As a consequence of this normalisation the true scale parameter will be confounded with the α_i and β parameters in the representative utility function. This leads to the so-called "scale" problem (Bates, 1988) that needs to be taken into account when using stated preference models for forecasting. We will describe the "scale" problem below.

It is a well-known result (see Ben-Akiva and Lerman, 1985) that the multinomial logit model will reproduce the market shares in the estimation sample such that:

$$\frac{1}{N}\sum_{n=1}^{N} y_{ni} = \frac{1}{N}\sum_{n=1}^{N} P_{ni} \tag{4.2}$$

where y_{ni} equals 1 if individual n is observed to choose alternative i and 0 otherwise. Because of this there are no serious implications of confounding the scale parameter with the coefficients in the representative utility function when using RP data for estimation, since the RP model nevertheless reproduces the market equilibrium embodied in the sample. SP data, however, do *not* in general embody information about the market equilibrium, and SP models will not reproduce the market equilibrium in simulation *unless* the error variance in the SP model equals the error variance in the RP model. This is easy to demonstrate if it is recalled from chapter 3 that the scale parameter is inversely related to the error variance. Even if the true coefficients of the representative utility function are the same in the two models, $\alpha_i^{RP} = \alpha_i^{SC}$ and $\beta^{RP} = \beta^{SC}$, the forecasts from the two models will be different unless $\mu^{RP} = \mu^{SC}$ which will only be the case if $\sigma_{RP}^2 = \sigma_{SC}^2$.[3] Furthermore it can be shown that if $\sigma_{RP}^2 < \sigma_{SC}^2$ the SP model will overpredict the minor mode or the mode with the lower share (see appendix 4.1 for a numerical example).

This begs the question of whether or not the error variances from the RP and SP models are likely to be equal. The answer is unfortunately that they are not because the sources of the random terms in the two models will be different. The main sources of error in the RP model will be measurement error in the explanatory variables, taste differences (assuming β is equal for all n when in fact it is not), and model specification error such as wrong functional form and missing variables (see Train, 2003). While the latter two will clearly apply also in the SP model, measurement error is not likely to be a problem since the values of the attributes are given in the experiment. However, there is another important source of error in the SP model, namely that individuals might behave differently when making choices in an

[3] Note that it is also possible, of course, that $\alpha_i^{RP} \neq \alpha_i^{SC}$ and $\beta^{RP} \neq \beta^{SP}$. Wardman (1988) examines the equality of coefficients of several SP and RP models and concludes that there is evidence of

experimental setting compared to making choices in the market (see section 4.1.2). As a consequence of the differences in the source of error in the different types of data Bates (1988) concludes that "it seems unlikely that a utility function as derived from SP analysis will be correctly scaled relative to the random effects which we hypothesise to be active in *real* choices". There is thus a need for rescaling the estimated coefficients in the SP model. We will describe two of the most common methods below.

4.4 Solutions to the "scale" problem: simple rescaling and data fusion.

Since it is likely that the RP and SP scales will differ it is necessary to use additional RP data to rescale the coefficients in the SP model. One straightforward way to do this is to rescale the coefficients to reproduce one or more coefficients from an RP model. This is the method which will be employed in chapter 7. An approach that has become increasingly common the literature is to combine RP and SP data in a process called *data fusion*. The rationale behind data fusion is that combining RP and SP data has the potential to yield more robust parameter estimates given the relative benefits of the two types of data. Furthermore, the scale differences between RP and SP alternatives can be estimated simultaneously with the parameters of the model. The joint estimation approach would be feasible in chapter 7 if the SP experiment included users of other existing modes such as bus. Since it was chosen to focus on the

equality given that heterogeneities in the sample are accounted for.

switching behaviour of car drivers, however, this approach cannot be adopted here. In the following the technical aspects of the data fusion process will be described nevertheless, as it looks to become the standard approach for combining RP and SP data in the future.

The data fusion process, which was originally proposed by Morikawa (1989), assumes that there is at least one common variable in the RP and SP data and that the coefficients for all common variables are equal. Also, because of the different sources of error between the RP and SP alternatives, the scale of the RP and SP alternatives are allowed to differ. Ben-Akiva and Morikawa (1990) describes a simultaneous estimation approach for the multinomial logit model, in which the RP scale parameter is normalised to unity for normalisation purposes, $\frac{1}{\mu_{RP}} = 1$, while the SP scale parameter, $\frac{1}{\mu_{SP}}$, is freely estimated along with the parameters of the model (all SP alternatives are restricted to have the same scale, however).[4] This model cannot be estimated in standard econometrics packages but is relatively straightforward to implement in a package like GAUSS.[5]

Although the majority of the applications of the data fusion methodology to date have used the multinomial logit model (Morikawa, 1989; Ben-Akiva and Morikawa, 1990; Hensher and Bradley, 1993; Swait *et al.*, 1994) more recent work has employed more flexible models, such as the heteroscedastic extreme value model

[4] Since $\frac{1}{\mu_{SP}} = \frac{\pi}{\sqrt{6\sigma^2_{SP}}}$ and $\frac{1}{\mu_{RP}} = \frac{\pi}{\sqrt{6\sigma^2_{RP}}}$ (see chapter 3), it follows that $\frac{\mu^2_{RP}}{\mu^2_{SP}} = \frac{\sigma^2_{RP}}{\sigma^2_{SP}}$ where σ^2_{RP} and σ^2_{SP} are the error variances in the RP and SP models. This ratio gives the error variance of the SP data as a percentage of the variance of the RP data.

[5] Hensher and Bradley (1993) and Bradley and Daley (1997) describes how this model can be estimated using a nested logit structure.

(Hensher, 1997; Hensher *et al.*, 1999), the nested logit model (Cherchi and Ortuzar, 2002) and the mixed logit model (Brownstone *et al.*, 2000; Bhat and Castellar, 2002). The mixed logit model offers the most general modelling framework as the benefits of the other models (heteroscedasticity and correlation between alternatives) can be incorporated along with taste heterogeneity and correlation between choices made by the same respondent (see section 4.6).

The views of the authors using the data fusion methodology differ somewhat when it comes to the origin of the data. Louviere *et al.* (2000) argue that the RP data and SP data may come from different sources while Morikawa (1994) emphasises that the RP and SP data should come from the same individuals. Since both preferences and sources of error may differ between individuals in different locations it is likely that combining data from different sources will not always be possible (see Atherton and Ben-Akiva, 1976 for a discussion of model transferability).

A formal test of preference homogeneity in the RP and SP data is given in Swait and Louviere (1993). The test statistic for the hypothesis that the common utility parameters are equal is given by $-2\left[(L^{RP}+L^{SP})-L^{J}\right]$, where L^{RP}, L^{SP} and L^{J} are the log-likelihoods of the models estimated on the RP data, SP data and the joint RP/SP data respectively. The statistic can be shown to be asymptotically chi-squared distributed with $|\beta|-1$ degrees of freedom, where $|\beta|$ is the number of parameters common to the RP and SP models. Although this test is designed to test the null hypothesis of preference homogeneity in the RP and SP data there are several reasons other than preference heterogeneity that can lead to rejection of the null. The design, layout, framing, context etc. of the SP experiment are all crucial elements to combining the data successfully. If the task is to forecast the real market accurately, the SP experiment should closely reflect the choices made in the market with regards to the process of defining the task, the attributes, attribute levels etc. In addition, ill-conditioned RP data may affect the outcome of the test, as may omitted variables such as interaction effects (see Louviere *et al.*, 2000 for a discussion).

Note that it is up to the researcher to specify the number of parameters the two models have in common. If full data enrichment (all common parameters except alternative specific constants are equal) is rejected, it is possible to re-specify the joint model to allow for more parameters to be "data specific" in order to test the hypothesis of *partial* data enrichment. Partial data enrichment, however, leaves the researcher with the question of which parameters should be used for prediction. This question is not yet fully resolved in the literature, but it's been suggested (Louviere *et al.*, 2000; Morikawa, 1994) that the prediction model should contain the RP alternative specific constants and all the parameters that were jointly estimated.

4.5 Testing for fatigue/ learning effects in stated preference models

The joint estimation procedure outlined above can also be used to test for fatigue and learning effects in the SP experiment (see section 4.1.2) using the approach outlined in Bradly and Daly (1994). The testing procedure utilizes the method of estimating separate scale factors for different alternatives (which in the previous section are the "RP alternatives" and "SP alternatives") to estimate separate scale factors for each choice task in the SP choice sequence *within a single SP model*. The scale parameter for one of the choice tasks in the sequence is normalized to unity, typically the scale parameter for the first or last choice performed. If the scale parameter is found to be increasing (the variance decreasing) in the number of choices performed, this can be interpreted as evidence of a learning effect since the decrease in the error variance indicates that the respondents behave more consistently towards the end of the experiment. If the opposite is observed this is evidence of a fatigue effect. The likelihood ratio test described in chapter 3 can be used to test whether the explanatory power of the unrestricted model (with "free" scale parameters) is significantly better than the restricted model (with a common scale parameter for all choice tasks). If the null hypothesis of the restrictions being valid is rejected this is evidence of a learning or fatigue effect.

4.6 The "repeated measurements" problem

One of the advantages of SP data is that each respondent typically performs several choice tasks, thus providing more information about his or her preferences than in an RP survey. Many authors have argued, however, that these responses are not likely to be independent since there may be unobserved individual characteristics influencing the choices made in all choice tasks (Ouwersloot and Rietveld, 1996; Abdel-Aty *et al.*, 1997).[6] Since this implies that the random components are not independently distributed, the IID assumption underlying the multinomial logit model is violated. This is often referred to as the problem of "repeated measurements" (Bates and

[6] Consider for instance an individual who has a particular dislike for travelling with others.

Terzis, 1999). The problem arises since the additional information provided by having the same individual perform more than one choice task is not as great as if the choice tasks were performed by different individuals. It has been shown that the correlation in the random terms leads to an upward bias of the *t*-statistics in the model (Cirillo *et al.*, 2000). The *coefficient estimates* of the model, however, are biased only if the random terms are correlated with the explanatory variables (Morikawa, 1994). This may be the case if 1) the respondents' actual (RP) choice is included as an explanatory variable in the model (to investigate whether there is evidence for choice inertia) or 2) the design attributes are based around the attributes of the actual choice. A solution to the repeated measurements problem is to use the mixed logit model with an error structure that takes the correlation between choices into account (see chapters 3 and 7).

Appendix 4.1 Numerical example of a forecast derived from a wrongly scaled SC model.

Let us assume that the representative utility of mode 1 and 2 are given by:

$$V_{n1} = -0.2 \times TIME_{n1} - 0.1 \times COST_{n1}$$

$$V_{n2} = -0.2 \times TIME_{n2} - 0.1 \times COST_{n2}$$

The travel times and cost of the two modes for a hypothetical individual are given in the table below. It is easy to see that in this case the SC model will over-predict the demand for the minor mode by 78% given that the SC scale is half the size of the RP scale ($\mu^{SC} / \mu^{RP} = 0.5$).

Table A1. Travel time and cost of two alternatives.

Alternative	Time	Cost
1	5	10
2	15	5

Table A2. Comparison of the forecasts derived from the RP and SC models assuming the SC scale is half the size of the RP scale.

V^{RP}	V^{SC}	P^{RP}	P^{SC}
-2	-1	0.82	0.68
-3.5	-1.75	0.18	0.32

Chapter 5

Discrete Choice Modelling of Work Trip Mode Choice

There have been numerous applications of the discrete choice methodology in studies of commuters' mode choice since McFadden's groundbreaking work on commuting in the San Francisco Bay Area (McFadden, 1974; 1978). Because of the number of contributions it would be a near impossible task to offer a complete review of the literature on the subject. The present chapter summarizes the main findings of some studies that show the breadth of topics investigated in the literature (the value of travel time savings, forecasting the demand for a new mode, forecasting the response to policy measures such as road pricing etc.) as well as highlighting the methodological developments outlined in chapters 3 and 4.

5.1 Early disaggregate mode choice studies

An early example of using a random utility framework to model travel demand is given in McFadden (1978). McFadden estimates a multinomial logit model of work-trip mode choice, using data on a sample of commuters in the San Fransisco Bay Area before the inauguration of BART (Bay Area Rapid Transit), a new light rail service.[1] The model is subsequently used to predict the share of users of the new service, and the predictions compared to the sample modal-split after the implementation of BART. The estimated model is described in table 5.1 below.

Both the coefficient on cost divided by post-tax wage and the coefficients on travel time have negative signs and are statistically different from zero (the marginal utilities of cost divided by post-tax wage and travel time are negative). Travel cost is divided by post-tax wage to reflect that a highly paid individual is less concerned about the cost of travel than one with a lower income. This also facilitates calculating the subjective value of time (*SVOT*) as a percentage of the wage. The time spent while travelling is decomposed into in-vehicle time, walk time and wait time, and the coefficient on in-vehicle time is allowed to vary between auto and transit modes. This decomposition allows for analysis of policies trading off these components, an example being a policy that places more buses on fewer bus lines and thereby decreasing wait time and increasing walk time.[2] It also allows the estimation of *SVOT* for the different time components. The headway of the first bus is the number of minutes between bus arrivals at the first bus stop (initial wait time is often calculated as half of the headway).

[1] The four transport modes available before BART were car alone, bus with walk access (Bus W), bus with car access (Bus C) and carpool.
[2] See Train (1980)

Table 5.1 Multinomial logit mode choice model in McFadden (1978)

Variable	Alternative	Coeff.	t-stat.
Constant	Car alone	-5.260	5.93
Constant	Bus C	-5.490	5.33
Constant	Carpool	-3.840	6.36
Family income (thousands of $ per year)			
Effect up to $7.5K	Car alone	-0.005	0.05
Effect between $7.5K and $10.5K	Car alone	-0.057	0.43
Effect above $10.5K	Car alone	-0.054	0.91
Number of drivers in household	Car alone	-0.102	4.81
Number of drivers in household	Bus C	-0.990	3.29
Number of drivers in household	Carpool	-0.872	4.25
Dummy if commuter is head of household	Car alone	-0.627	3.37
Employment density at work location	Car alone	-0.002	2.27
Home location in (2) or near (1) CBD	Car alone	-0.502	4.18
Number of cars per driver	Car alone	5.000	9.65
Number of cars per driver	Bus C	2.330	2.74
Number of cars per driver	Carpool	2.380	5.28
Cost/post-tax wage (cents/cents per min.)	All	-0.028	4.31
Auto in-vehicle time (min.)	Car alone, Bus C, Carpool	-0.064	5.65
Transit in-vehicle time (min)	Bus W, Bus C	-0.026	2.94
Walk time (min)	Bus W, Bus C	-0.069	5.28
Transfer wait time (min)	Bus W, Bus C	-0.054	2.3
Number of transfers	Bus W, Bus C	-0.105	0.78
Headway of first bus (min)	Bus W, Bus C	-0.032	3.18
Log-likelihood at zero L(0)		-1069.0	
Log-likelihood: final value L(β)		-595.8	
Rho-squared (with L(0))		0.443	

The coefficients on family income are allowed to vary for different income groups to facilitate a non-linear relationship between income and representative utility. However, the coefficients are small and insignificant. McFadden finds this unsurprising given that the number of autos per driver enters as a separate (and highly significant) variable in the model and travel cost is divided by wage. Thus, the channels through which income is likely to influence mode choice are controlled for independently of the income variable.

The alternative specific dummies ensure that the mean of the unobserved component of utility is zero, and can be interpreted as the average effect of the unobservable components on the utility of choosing an alternative. Since only differences in utility matter and not the overall scale, one of the constants (Bus with walk access) is normalised to zero. This is also the case for the socio-demographic variables, and is necessary to facilitate the estimation of the parameters (see chapter 3).

The value of travel time as a percentage of post-tax wage can be calculated as the ratio of the time and cost coefficients times 100 (see chapter 2). The value of auto in vehicle time is found to be substantially higher than the value of transit in vehicle time, indicating that the time spent in the car is regarded as more onerous than the time spent on the bus or train. This does not include the other aspects of travelling by transit such as walk and wait times, and McFadden suggests it might be a result of the positive aspects of transit travel such as being able to read and work while travelling.

The estimated model is used to predict the share of BART users in the sample. This is done simply by including BART with auto access and BART with walk access as alternatives in the logit formula. However, since there are no alternative specific constants relating to BART in the model, the alternative specific constant for bus with car access was used as an approximation. McFadden acknowledges this as a weakness of the forecasting process, which could have been overcome by incorporating stated preference data in the model (see chapter 4 and section 5.6). Nevertheless, the model was found to predict the share of BART users exceptionally well, with a forecasted share of 6.4% compared to the actual share of 6.2% (table 5.2). McFadden reports, however, that the accuracy of prediction is better than one would expect given the size of the standard errors of the forecasts. The model underpredicts the share of the auto alone mode and overpredicts the shares of the bus modes. This is likely to be a result of the IIA property: the users of the new transit mode come proportionally more from users of the existing transit modes than those commuting alone by car and this is not taken account of in the MNL model. This could have been overcome by estimating either a nested or a mixed multinomial logit model that allowed for this pattern of substitution (chapter 3).

5.2 Modal split after the introduction of BART. Predicted vs. actual shares.

	Car alone	Bus with walk access	Bus with car access	BART with bus access	BART with car access	Carpool
Predicted share	55.8%	12.5%	2.4%	1.1%	5.3%	22.9%
Actual share	59.9%	10.8%	1.4%	0.95%	5.2%	21.7%

A somewhat later application of the multinomial logit to model commuters' mode-choice is given in Dunne (1984), who uses data from Livingston, Scotland to calibrate his model. Livingston is one of the so-called "New Towns" in Scotland, designated in 1962. Its planners paid special attention to the mobility of the town's residents with regard to minimising traffic flow delays, providing an extensive segregated footpath system and low walk times to bus stops from dwellings. Dunne finds, as the only study in this review, that the time and cost of the modes are insignificant determinants of mode choice. On the other hand car-ownership, relative to household size and number of workers, seems to be an important determinant of car use, along with gender and status within the household. This is in line with the findings in McFadden (1978) and other studies in the review.

5.2 Joint car ownership/ mode choice models

Some authors have argued that since the commuters' choice of mode and the households' level of car ownership level is likely to be made simultaneously, a mode-choice model should not treat car ownership as being exogenous to the mode-choice decision. This is important since the parameters in the model may be biased if car ownership is indeed an endogenous variable. There are some early examples of joint car-ownership and mode choice models in the literature (Ben-Akiva and Atherton, 1977; Train 1980; Thobani, 1984), but this section will focus on a more recent application given in de Palma and Rochat (2000).

The model in de Palma and Rochat (2000) is a nested logit model of joint car-ownership and mode choice estimated using a sample of commuters in Geneva, Switzerland. Since the level of car ownership in this region is very high (98% of the households in the sample own at least one car) the car-ownership choice set is defined as the choice between owning one or two or more cars. In this application of the nested logit model the inclusive value term described in chapter 3 can be interpreted as the expected utility the commuter derives from a specific car ownership level. As always the dissimilarity parameter must lie in the 0 to 1 interval to ensure that the model is globally consistent with utility maximisation (see chapter 3). If the parameter is equal to one the nested logit collapses to the multinomial logit model and simultaneity is rejected. De Palma and Rochat find that the parameter in their model lies in the 0 to 1 interval and is significantly different from one, and therefore conclude that there is evidence for simultaneity of the car-ownership/ mode-choice decision. This is consistent with the findings in previous studies. Train (1980) compares the nested logit model to the simpler multinomial logit where car ownership is treated as exogenous and finds that there is little bias in the parameter estimates in the MNL model. This is reassuring since it suggests that the simpler (MNL) model may be adequate for modelling short-term travel decisions where the level of car-ownership is assumed to be constant.

An innovative feature of de Palma and Rochat's model is the inclusion of comfort and availability as alternative specific attributes. While it is recognised in the literature that attributes other than time and cost influence individuals' choice of mode (Nerhagen, 2001), these attributes are rarely included in practical applications since they are difficult to quantify. De Palma and Rochat get around this problem by asking individuals to rank the availability and comfort of their chosen mode. The average ranking for each mode enters as the level of the attributes of the alternatives in the individual's choice set. The authors argue that this ranking procedure, as opposed to having all individuals rank the alternatives available to them, reduces the likelihood of justification bias (ranking the chosen mode higher/ alternative modes lower). The authors find that availability is an important determinant of mode choice while the parameter on comfort has the expected sign but insignificant.

De Palma and Rochat include other innovative variables in their model. In particular network experience (number of years on principal route) and congestion seem to be important determinants of mode choice. The estimated attribute elasticities are similar in magnitude to other models of urban commuting, the time and cost elasticities are -0.27 and -0.29 for the car mode and -0.61 and -0.43 for the public transport modes respectively. The authors argue that the relatively low cost elasticities suggests that monetary incentives may not be effective in reducing car use while policies directed to improve the efficiency of the public transport network may be more successful.

The car ownership model primarily includes household variables as the decision on how many cars to purchase is assumed to be made on the level of the household rather than the individual. Unsurprisingly (household) income seems to be a significant determinant of the households' car ownership level, confirming the result from other studies (Train, 1980; Thobani, 1984; Hensher *et al.*, 1989; Pendyala *et al.*, 1995). In addition the size of the household and the occupation of the sampled individual seem to be important. The authors explain the significance of the "occupation" variable by noting that white-collar workers are often provided with a car from their company.

5.3 The bicycle as an alternative to the private car

Noland and Kunreuther (1995) is the only paper in the review focusing specifically on how to increase the share of commuters travelling by bicycle. This reflects the fact that the focus in the literature has until quite recently been on public transport as the main alternative to the private car. As a parallel to the increasing focus on cycling among policymakers in the UK and elsewhere (DoT, 1996), however, some recent studies have focused on the bicycle as an alternative to car use, especially for commuters who live relatively close to their workplace (Cleary and McClintock, 2000; Kingham *et al.*, 2001).

Noland and Kunreuther hypothesise that safety concerns are the main barrier to bicycle use. In order to investigate this hypothesis the survey respondents were asked to rank the probability of having an accident and, given that an accident had taken place, the expected severity of the accident. This information was used to create an individual risk coefficient for each mode (car, transit, walk and bicycle), which enters as an explanatory variable in the model. As expected it was found on average that bicycle was perceived as being the riskier mode (also by bicycle users), while transit was perceived as being the safest mode. In addition to the generic risk coefficient the respondents were asked to rank how they felt the risk of riding a bicycle relates to weather conditions (rain, snow, ice etc.) as well as road conditions (potholes on surface, no hard shoulder etc.). The authors find that the coefficent on the generic risk coefficent had the right sign but was of low significance, while the coefficent on road conditions was strongly significant in the expected direction. The coefficent on weather conditions was also insignificant, which is perhaps reassuring since policymakers have no influence over weather conditions.

In addition to the risk variables the individuals were asked to rank the perceived comfort and convenience of each mode. Both the coefficient on comfort and convenience were found to be highly significant and of the expected sign. The coefficient on the time variable was found to be insignificant when the convenience variable was included in the model. This is, the authors argue, explained by the positive correlation between these two variables and therefore suggests that one of the variables should be dropped from the equation. The model including convenience is found to be superior to the one including time, suggesting that there are elements other than time influencing the convenience of the modes which are relevant in the mode choice decision. Whether there is bicycle parking available at the workplace is also found to be a significant determinant of mode choice along with cost, gender (males are found to be more likely to cycle and use transit) and car ownership. In conclusion the authors argue that a policy designed to increase the share of commuters travelling by bicycle should focus on providing convenient bicycle lanes and bicycle parking at the workplace, possibly combined with policies aimed at making the car less attractive such as restricting parking by turning car parks into bicycle parks and timing traffic lights such that they reflect the average speed of bicycles rather than automobiles.

The ranking procedure in Noland and Kunreuther differs from the one in de Palma and Rochat (2000) since each individual ranks all the alternatives in her choice set. This may increase the likelihood of justification bias (ranking the chosen mode higher/ alternative modes lower), which will lead to upward bias in the coefficient estimates. Another problem related to the inclusion of perceived attributes in the model is that the link between perceptions and objective values is ambiguous. Although, from a behavioural perspective, it is the perceived level of the attributes that matters for the individual's decision making process, a model estimated using the perceptions of the attributes has little predictive power unless one knows the link between perceptions and objective values (Small, 1992). Noland and Kunreuther recognise this argument and suggest that this is an important area for future research.

5.4 Parking and mode choice

There is a branch in the literature on commuters' mode choice focusing explicitly on the link between parking conditions at the workplace and mode choice (see Feeney, 1989 for a review). Apart from discrete choice models the most common methodology is "before and after" studies that investigate to what extent the modal split changes following a parking policy change at the workplace, usually the introduction of a parking charge (e.g. Shoup and Willson, 1990). In studies using the discrete choice methodology the models are usually estimated on a cross section of commuters with different parking conditions at the workplace, in order to investigate how the differences in conditions influence the choice of mode.

Willson (1992) investigates how employer-paid parking influences mode-choice for the work trip using a sample of commuters in the Los Angeles area. The sample consists of two groups: one consisting of individuals who are provided with free parking at or near the worksite and one of individuals who have to pay the market price to park. Following Gillen's (1977) argument that individuals may respond differently to changes in parking costs from changes in automobile running costs (usually defined as fuel and maintenance costs), Willson specifies separate coefficients for running costs and parking costs in his model (a similar argument is made by many authors in the literature on congestion charging, see section 4.7). He finds that the coefficients for both cost components are significant in the expected direction, and that the coefficient for running cost is larger (in absolute value) than the coefficient for parking costs. Hensher (2001a) hypothesizes (and finds evidence for) that the cost component which is greatest in magnitude will have the smaller coefficent (see section 4.7). Since the parking charge is likely to be higher than running costs for most of the commuters in the sample, Willson's finding supports Hensher's hypothesis.

In addition to specifying a separate coefficient on parking costs, Feeney (1989) argues that parking models should include separate coefficients for walking time (from parking to work-site) and the time spent searching for a free parking space. This is not a feature of Willson's model, and the author recognises this as a shortcoming of the study. However, since Willson's aim is to predict changes in the modal-split following the introduction of parking charges he argues that the inclusion of these variables in a generic "door-to-door" travel time variable is acceptable.

Willson estimates that between 25 and 34 percent fewer cars would be driven to work following the introduction of a ($4.15) parking fee. As a consequence he argues that the introduction of parking charges has significant potential for reducing the number of cars driven to work, while the current practice of subsidizing car parking seriously undermines policies designed to encourage the use of alternative modes such as car sharing and public transport.

5.5 Estimation of the value of travel time[3]

A great number of studies in the literature are concerned with the estimation of the subjective value of time (*SVOT*). Commuters' value of time is of importance both for forecasting and in assessing the benefits of improving the infrastructure and in many countries the authorities have commissioned studies estimating *SVOT* both for commuting and other types of trips (the UK, the Netherlands and the Scandinavian countries among others).

While *SVOT* can be derived from any mode choice model as the ratio of the travel time and cost coefficients in the model (see chapter 2), some authors have argued that mode choice models are not the most suitable approach for valuing travel time savings since they are likely to confound *SVOT* with other factors related to the difference between the modes. In particular, since the time coefficient in a mode

[3] This section has benefited substantially by comments made by Prof Otto Anker Nielsen.

choice model is usually specified to be generic, this coefficient will typically capture the relative comfort, convenience, privacy etc. of the different modes unless these attributes are controlled for in the model. While it is possible to specify separate time coefficients for the different modes in the model, car and bus say, it is still assumed that car and bus users have the same car time coefficient. As a consequence, Calfee and Winston (1998) argue that the car time coefficient is likely to be inflated "if bus users do not choose auto because they attach a higher disutility in driving in congestion than auto users". Hence, if the goal is to estimate the value of travel time savings for the car mode for individuals who currently travel by car, a mode choice model may be inappropriate.

Calfee and Winston (1998) estimate the benefits of a congestion charge (the value of the reduction in commuting time following the implementation of the charge) for commuters who currently travel by car and face some congestion. They argue that given the drawbacks of the mode choice model when it comes to estimating *SVOT* for a particular group of commuters, a better approach is to use a stated preference experiment where the respondents rank several scenarios involving different times and (toll) costs but where none of the scenarios involve switching to an alternative mode. They separate the time spent travelling into congested travel time and uncongested travel time based on the hypothesis that the disutility of congested travel time is lower than that of uncongested travel time. The reason for this is intuitive; uncongested travel time is perceived as less onerous than congested travel time as it enables commuters to "decompress" after work. This hypothesis is confirmed by the estimation results, as the coefficient for congested travel time is found to be roughly three times higher than that of uncongested travel time.

Calfee and Winston's main finding is that the value of congested travel time (or, alternatively, the willingness to pay to reduce the time spent travelling in congested traffic conditions) is estimated to be between 14 to 26 percent of the gross hourly wage, which is considerably lower than the *SVOT* derived from mode-choice studies (Small, 1992a, finds in a review of mode choice models that a reasonable average is 50% of the gross wage). Furthermore, the value of time is found to be insensitive to alternative uses of the revenues arising from the toll. This finding is in contrast to Small (1983), (1992b) and Mohring and Anderson (1994), who suggest that the key to political acceptance of a congestion charge lies in how the revenues from the toll are spent. Calfee and Winston conclude that their findings help explain why there has been little public support for tolls in the US and elsewhere, since it is doubtful that the net benefits from a toll are high. In spite of this, however, the authors support the widely held claim that other measures directed towards reducing congestion (expanding public transportation, implementing intelligent vehicle road systems that guide motorists onto the least congested routes) may be even less desirable in the long run since "commuters who have previously avoided congested roads by, for example, driving during off-peak hours, will be lured back onto the roads by the promise of uncongested travel".[4]

Hensher (2001a) investigates how different model specifications (multinomial versus mixed logit) influence the estimate of the mean value of travel time savings. He uses a stated choice approach to investigate how car drivers value the time spent travelling under different conditions. The experiment is similar to that of Calfee and Winston in that none of the scenarios involves switching to an alternative mode. The travel time is divided into three components: free flow time, slowed down time and

[4] This phenomenon is known as Downs' (1962) law.

start/ stop time. In addition there is an uncertainty allowance included in the experiment, which is defined as the extra time the commuter needs to allow herself to ensure that she arrives at work on time. Hensher hypothesises that the magnitude of the time coefficients will be increasing (the disutility of slowed down time is higher than that of free flow time and the disutility of start/ stop time is higher than that of slowed down time), which is confirmed by the empirical evidence.

Hensher also specifies different coefficients for running costs and toll costs in his model, since commuters may respond differently to toll costs from running costs. He hypothesizes that the cost attribute that is the greatest in magnitude, which in this case is the toll, will have the smaller coefficient (in absolute value). This hypothesis is confirmed by the empirical evidence for all model specifications. Hensher argues that the decomposition of travel time and travel costs would be difficult using revealed preference data, since there is normally too much confoundment in RP data to obtain precise parameter estimates at this level of disaggregation. In addition some attributes (such as toll costs) do not exist or are of limited variability which makes it impossible to establish their influence on mode choice.

A crucial question that faces the analyst when applying the mixed logit model is deciding which parameters should be allowed to vary, and which distribution to use for the random parameters. This question seems to be guided by practical issues and experience with which specifications yield behaviourally plausible values of *SVOT* rather than theory. Ruud (1996) has pointed out that a mixed logit model where all parameters are allowed to vary has a tendency to be unstable. Brownstone (2000), on the other hand, points out that if both the time and cost parameters are specified to be normally distributed (and uncorrelated) the distribution of the ratio of the coefficients (*SVOT*) will have a Cauchy distribution, which has no finite moments (see also section 3.7). This will also be the case if the cost coefficient is normally distributed and the time coefficient fixed, since the reason that the value of time has no finite moments is that the distribution for the cost coefficient crosses zero. It follows that any distribution for the cost coefficient which is strictly positive yields a distribution of value of time with finite moments, given that the time coefficient follows a distribution which has itself finite moments. The distribution of the value of time may also have finite moments when the distribution for the cost coefficient limit zero, but this is not the case for all distributions with this characteristic (it holds for the lognormal distribution, but not for the exponential distribution for example). Given these findings, in addition to the observation that the coefficient for cost should logically be negative, the majority of the applications of the ML model in the literature have specified the cost coefficient to be either fixed (Revelt and Train, 1998; Train, 1999; Hensher, 2001b; Carlsson, 2003; Alpizar and Carlsson, 2003) or lognormally distributed (Train, 1997; Brownstone and Train, 1999). Specifying the cost coefficient to be fixed is convenient since this implies that the value of time follows the distribution of the time coefficient (Revelt and Train, 1998; Carlsson, 2003). The lognormal distribution, on the other hand, is convenient when the time coefficient is specified to be lognormally distributed, since in this case the value of time will also be lognormally distributed.[5] In the majority of the applications the time parameter is specified to be normally or log-normally distributed, while in Hensher (2001a) the time parameters are specified to follow a triangular (tent shaped) distribution. It should be pointed out that the normal and triangular distributions may not be appropriate if the estimates imply that a substantial share of the population

[5] Train (1999) and Hensher (2001a) argue that because of the thick tails of the lognormal distribution a log-normally distributed cost parameter may result in *SVOT* estimates that are behaviourally implausible as parameter estimates very close to zero give very high estimates of *SVOT*. This claim is not supported by all authors in the literature, however.

have positive time parameters. Since this cannot be determined *a priori* this issue should be evaluated on a case for case basis and will be addressed in the empirical section of the thesis.

Hensher (2001a) finds that the value of time derived from the multinomial logit model is lower than the *SVOT* derived from the less restrictive mixed logit models which allow for more flexible (non IIA) substitution patterns. This is consistent with similar findings using more flexible choice models such as the heteroscedastic extreme value model, the covariance heterogeneity model and mixed logit to model the choice of mode for long distance travel (Bhat, 1995; Hensher 1997; 2001b; c) but not with the findings in Brownstone and Small (2003), Nielsen and Jovicic (2003) and Nielsen and Sørensen (2004).

5.6 External validation of Stated Preference models

As discussed in chapter 4 the main concern regarding the use of stated preference travel demand models is whether choices made in a hypothetical setting are congruent with actual choice behaviour. This question cannot be answered in the abstract, and a growing number of empirical applications have focused on comparing stated preference models to market data (external validation).[6]

Following Beaton *et al.* (1998) there are essentially two types of external validity tests available. The first type is based on the hypothesis that, given preference equality, there should be no significant difference between the parameter estimates of

[6] As opposed to internal validation, or tests for the consistency of SP responses (for example whether the responses satisfy the reflexivity axiom).

an SP model and those of an RP model estimated on a common sample of decision makers. Because of differences in scale, however, these estimates cannot be directly compared (see chapter 4). The ratio of two coefficients, on the other hand, can be compared since in this case the scale factors cancel out. Wardman (1988) examines the equality of the coefficient ratios of several SP and RP models and concludes that there is evidence of equality given that heterogeneities in the sample (due to differences in socio – demographic characteristics) are accounted for. The likelihood ratio test described in chapter 4 provides a formal way of testing for parameter equality in SP and RP models.

Beaton *et al.* (1998) point out that the tests for parameter equality do not test the *predictive* validity of the SP model, as the explanatory power of the model can be high (the model is well suited to explain current choices) while its ability to forecast switching behaviour may be low. The second type of external validity test focuses on prediction rather than explaining current behaviour. Forecasts derived from an SP model predicting the demand for a new mode or the response to a change in one or more policy variables is compared to the actual modal split (the truth set) after the new mode is made available or the changes have been implemented. As in the first type of test the sample from which the model was estimated should be the same sample for which forecasts and truth sets are derived. It should be noted that this type of test is also relevant for RP models (see, for example, section 5.1 in this review).

In Beaton *et al.* (1998) a multinomial logit model estimated on a sample of SOV (single occupancy vehicle) commuters is used to forecast the demand for a new shuttle bus service at the respondents' worksite. Each respondent was asked to complete a stated choice experiment with the alternative of going by car as before or switching to public transport. The public transport option is combined with a shuttle bus service taking the commuters from the nearest transit stop to the worksite (the respondents had the choice between the new shuttle bus service and an already existing shuttle bus driving a different route). The design variables include parking costs, parking availability at worksite, the starting time of the new shuttle bus, whether the shuttle bus has room to stand and the headway of the shuttle bus. In addition several other (non-design) attributes are included in the model such as whether the respondent has a designated parking space, car running costs, public transport fare and walking time to the nearest public transport stop. All the design attributes, as well as most of the non-design attributes are found to be significant in the expected direction. The coefficient for parking costs is found to be higher in magnitude than the coefficient for car running costs, indicating that the marginal disutility of an increase in parking costs is higher than that of an increase in running costs.[7]

The probabilistic and deterministic forecasts derived from the model are compared to the actual switching to public transport. The probabilistic forcast is given by averaging the estimated probabilities for the sample individuals. This forecasting method is consistent with random utility theory and will reproduce the market shares when the model is estimated using market data. For the reasons discussed in chapter 4, this is unlikely to hold for an SP model since the SP scale is likely to differ from the RP scale. The deterministic forecast is given by assuming that the mode with the higher representative utility is the chosen mode for all individuals in the sample. This approach is not consistent with random utility theory as the random component of the

[7] This is in contrast to the finding in Willson (1992). The explanation may be that the parking costs in Beaton *et al.* are lower than the running costs of the car mode, as indicated by Hensher's (2001a) hypothesis.

model is ignored. Fowkes and Preston (1991) hypothesise that the probabilistic and deterministic forecasts are likely to bound the true share, as the probabilistic forecast is likely to overpredict the demand for a minor mode while the deterministic forecast is likely to underpredict the demand (see chapters 4 and 7).

Beaton *et al.* present the forecasts for the total switching to shuttle bus (both the new and old service) and switching to the new shuttle bus only. Compared to the modal split two years after the initial survey the probabilistic method overpredicts the switching to shuttle bus by 15% while the deterministic method underpredicts by 60%. The probabilistic method underpredicts the switching to the new shuttle bus by 23% percent, while the deterministic method underpredicts by 100%. In the first case the Fowkes and Preston hypothesis is confirmed. Due to employee turnover the employees' preferences might have changed over the years, and thus the quality of the aggregate modal split as a truth set degrades. In spite of this, Beaton *et al.*'s results suggest that a carefully specified SP model can predict the demand for a new mode reasonably accurately. This conclusion is supported by Fowkes and Tweddle (1999) in the context of anglo-continental freight.

Ben-Akiva and Morikawa (1990) use the data fusion method described in Chapter 4 to model the choice between access modes for train commuters in Yokohama, Japan. The purpose of the study is to forecast the switching to a new subway line from the previously existing access modes (walk, bicycle, bus and car). The respondents were asked if they intended to use the new service, and if so he or she was considered to have chosen the subway mode over the currently used mode. This particular type of stated preference data is called stated intentions (SI) data. The modal attributes in the RP, SP and joint RP/SP models include in-vehicle time for the mode, walk time and the number of transfers required for the public transport modes (a cost variable was dropped from the model as it was found to be insignificant. Ben-Akiva and Morikawa suggest that the reason may be that the cost of commuting is usually provided by the employer in Japan). The coefficients for all the attributes were found to be significant in the expected direction. The successful pooling of the two data sources suggests that the underlying preferences are similar given that differences in scale are accounted for.

In the joint RP-SP model the SP scale was estimated to be 0.559 (with the RP scale normalized to one), suggesting that the SP data contain more noise (have a higher error variance) than the RP data. This is similar to the findings in many studies utilizing the data fusion method (see for example Hensher and Bradley, 1993 and Ortuzar and Willumsen, 2001) although there are exceptions where the RP error variance is found to be higher than the SP variance (Morikawa, 1989; Chapter 7). Morikawa's (1994) finds that the RP and SP error variances are more similar when serial correlation and choice inertia are explicitly taken into account in the SP model.

A sample of train commuters taken after the survey was opened was used to estimate an "after" model with the same attributes as the "before" model. The parameter estimates of the after model are similar to the before model with the exception of the subway constant which is insignificant in the after model, while positive and significant in the before model. The authors suggest that this is evidence that the subway constant in the before model captures the policy bias in the SI data.

The predictions from the before model were compared to the actual switching to the new subway service. A difficult question when using joint RP/SP models for prediction is which alternative-specific constants to include in the model. Hensher and

Bradley (1993) argue that the theoretically consistent approach is to include the RP alternative-specific constants as well as the SP alternative-specific constants which have no RP counterpart (such as the subway constant in Ben-Akiva and Morikawa, 1990). An alternative approach suggested by Hensher and Bradley (1993) is to include all the alternative specific constants in the model. Ben-Akiva and Morikawa compare the forecasts derived from the RP model, the SP model (with and without the subway constant) and the joint RP/SP model (with and without the subway constant). They find that the SP model and the joint RP/SP model without the subway constants are the best performing models, overpredicting the switching to the new service by about 9% (the actual share was 59.35%). The models without the "bias adjustment" (those which include the subway constant) overpredict the switching to the new service by 20%. The authors suggest that the overprediction may result from the models' inability to take into account that some commuters are captive to the mode used prior to the construction of the subway. Captive travellers may not use the subway for reasons such as disliking subways, unfamiliarity with the service and habitual usage of alternative modes. Despite this shortcoming Ben-Akiva and Morikawa conclude on the basis of their findings that SP models that can be corrected for potential bias can have good predictive validity.

5.7 Conclusions

As pointed out in the beginning of the chapter the aim of this review is to provide an overview of many of the issues that are relevant in applied discrete choice modelling of commuters' mode choice, rather than providing an exhaustive review of the literature. The findings documented here along with the topics presented the previous chapters form a background to the analysis in the remaining chapters of the thesis, which present the empirical work conducted by the author. It should be noted that not all the issues covered are relevant for all the chapters: the issues regarding value of time estimation, for instance, are particularly relevant for chapter 6, while the issues surrounding the external validity of SP models are an important part of the discussion of the results in chapter 7. Some issues such as model specification and prediction, on the other hand, will be discussed in the applications in all three remaining chapters.

Chapter 6

Commuters' Mode Choice in Small Towns in Rural Areas: The Case of St Andrews

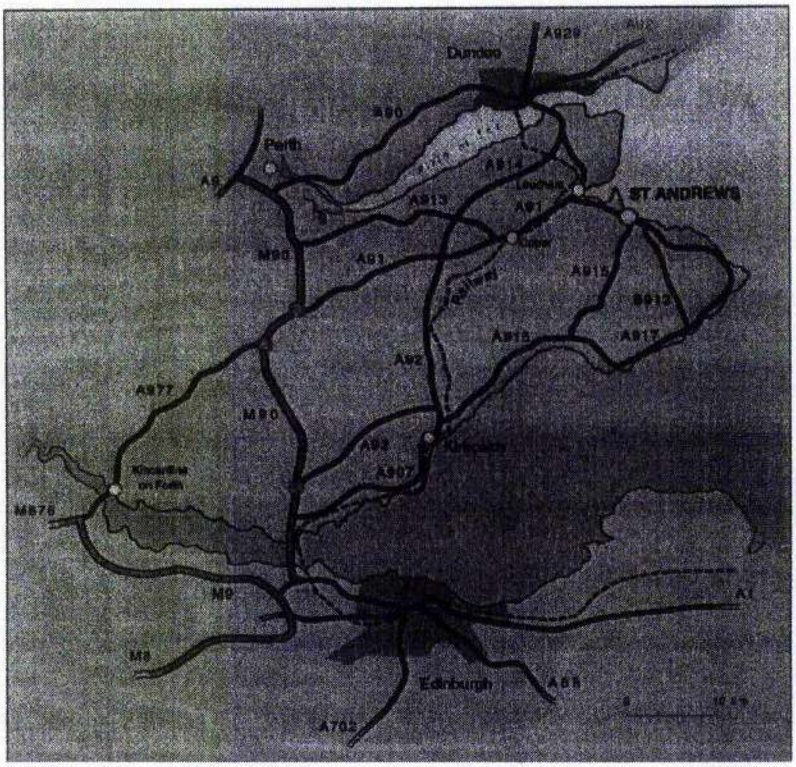
There have been many studies of commuting in urban areas in the UK, but relatively little research has been done on commuting in small towns in rural areas. Rural commuting differs from urban commuting in several important respects: there is little or no road congestion, a parking space is usually provided free by the employer and the supply of convenient public transport is often limited (Nutley, 1998). As a result a high share of rural commuters will depend on the private car to get to their workplace. Another consequence of these differences is that car reduction policies designed for large cities with ample public transport may be unsuitable for smaller towns. In particular pricing policies (such as congestion charges) may be less effective in reducing the share of drivers and encouraging public transport use in rural areas, as commuters with no convenient substitute to driving are unable to change mode. Since pricing policies will only be effective once a substitute is in place, improving public transport service quality is likely to be the most important policy tool to reduce driving in rural areas. It follows that in order to design effective policies to encourage use of public transport, policies must be based on evidence from studies focusing explicitly on rural commuters as one cannot *a priori* expect important policy parameters such as elasticities to be equal across geographical locations where commuting conditions differ markedly (Acutt and Dodgson, 1995).

St Andrews is a small town of about 18000 inhabitants[1] located in the rural North-Eastern part of Fife, Scotland (see figure 1). It is a typical Scottish small town in that it has rather limited public transport links, but somewhat untypical in being the location of Scotland's oldest University. The main mode of commuting is the private car followed by walking and cycling. Public transport has a relatively low market share, although some people commute by bus. Train is hardly used at all for commuting, as the nearest train station (Leuchars station) is about 5 miles away from the town with a relatively poor bus connection.

The current chapter develops multinomial, nested and mixed logit models of work-trip mode choice estimated using data from a survey of employees of the University of St Andrews, the town's main employer. The models are subsequently used to estimate aggregate direct and cross mode-choice elasticities and the value of travel time. The outline of the chapter is as follows: section 6.1 describes the data as well as providing some descriptive results from the survey, section 6.2 presents the modelling results and section 6.3 offers some policy recommendations and concluding remarks.

[1] Including students.

Figure 6.1 Map of Fife.



6.1 Data and descriptive statistics

6.1.1 Data characterization

As part of the development of a travel plan for the University of St Andrews a survey of employees' commuting behaviour was undertaken with questionnaires distributed to all members of St Andrews University staff. The survey collected information on the current mode used for commuting, socio-demographic variables such as occupation and car ownership as well as public transport availability at home and near the workplace (see appendix 6.1). Of the 1661 questionnaires that were distributed 642 were returned, giving a response rate of 38.7%. The sample is broadly representative of the staff population in the University, although there is an overrepresentation of females and individuals aged 40 or over in the sample (see table 6.1 below). Of the 642 questionnaires that were returned, 585 responses with complete information about the work trip and socio-demographic characteristics were used for model estimation. A list of the variables with some descriptive statistics is given in table 6.2 below.

Table 6.1 Characteristics of respondents compared to the population average

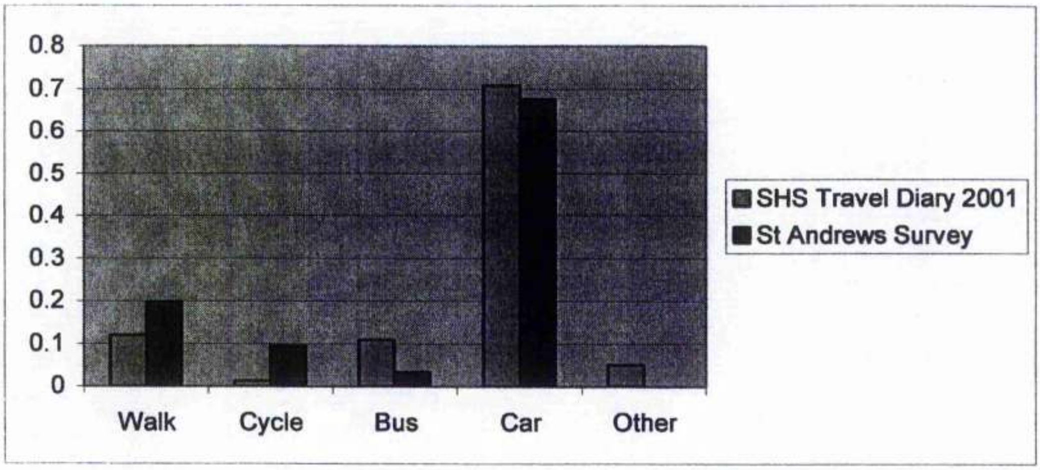
	Sample share	Population share
Female	54%	48%
Academic	39%	36%
Age		
Less than 30	12%	26%
30 – 40	26%	23%
40 – 50	28%	22%
Over 50	34%	29%

It can be seen from the table that the majority of commuters travel by car to work followed by walking and cycling, while only a small share of the commuters travel by bus. The relatively high shares of commuters who walk and cycle relative to the national average (see figure 6.2) reflects that a large proportion of the University staff live in the St Andrews area and that walking and cycling conditions are relatively favourable. The low share of commuters who travel by public transport is a result of the fairly poor bus service in the area. It can be seen from table 6.2 that 62% of the commuters in the sample do not have access to an hourly bus service going to and from their home to their workplace and that bus fares are relatively high with an average fare of £1.96 for a one-way ticket.

Table 6.2 Description of variables and data characteristics.

Mode	Sample Share
Walk	19.7%
Cycle	9.4%
Bus	3.4%
Car	67.5%
Choice set	
Walk available	25%
Cycle available	52%
Bus available	88%
Car available	91%
Alternative attributes	**Mean/ Share**
Door-to-door commuting time in minutes	
Walk	13.5
Cycle	12.1
Bus	36.83
Car	18.1
Walking time in minutes	
Walk	13.5
Cycle	1.2
Bus	14.0
Car	2.8
Travel cost in pence	
Bus	195.8
Car	122.7
Frequency of bus service to and from work	
At least one bus less frequent than 2 per hour	88%
At least one bus less frequent than 1 per hour	62%
Socio-economic variables	**Sample share**
High income	44%
Number of cars in household (mean)	1.4

Figure 6.2 Comparison to the modal split for commuting trips in the 2001 Scottish Household Survey Travel Diary.



It is well documented in the literature that there are differences between men and women's commuting behaviour, in particular in terms of bicycle use. In a recent British study, Dickinson *et al.* (2003) find that females are significantly less likely than males to cycle to work and equally car dependent in spite of having shorter commutes. The explanation may be that women have more complex trip characteristics than men due to tasks such as transporting children and shopping and/ or are more concerned with safety issues. In the models gender enters as a dummy explanatory variable (1= female, 0= male), which allows us to examine whether there is a similar difference between male and female commuting behaviour in the St Andrews area.

It is expected that the more cars a household owns, the more likely the individuals living in the household are to travel by car to work. Car ownership may be considered endogenous to the mode-choice decision as argued by Train (1980), who suggests a joint car-ownership/ mode-choice model using a nested logit structure (see chapter 5). Given that the data set contains few variables that are relevant to the households' car ownership decision, however, this approach cannot be followed here. Since the models estimate mode choice conditional on car ownership, they represent a short-run response to a change in the policy variables.

In addition to the socio-economic characteristics of the commuters, it is expected that the attributes of the modes are important determinants of mode choice. In particular the travel time and cost of the modes have been found to be significant explanatory variables in virtually all studies of commuting behaviour (see chapter 5). In addition, it is expected that the more frequent the bus service, the more likely the individual is to travel by public transport.[2] The frequency of the bus service enters the models as two dummy variables, indicating whether the individual has access to an hourly/ less frequent bus service (with a frequency of more than one bus per hour being the reference category). The reason that the bus frequency variable is specified in this way, rather than as a continuous variable, is that the majority of commuters have access to an hourly or bihourly bus service *or* a very infrequent service, e.g. a school bus which runs two times a day. The respondents self-reported the in-vehicle/ cycling time and walking times for their chosen mode. The travel time components for the alternative modes were calculated by regressing travel time on distance for each mode, using the estimated regression equations to calculate travel times for the non-chosen modes for all individuals in the sample.[3] It is hypothesized that an

[2] In a previous survey of staff commuting in the University of St Andrews (University of St Andrews, 2002) improving key elements of service quality such as the frequency and reliability of buses was found to be most important both to current public transport users and other commuters when asked what would encourage them to use public transport more often.

[3] Separate OLS regression equations was estimated for bus and car in-vehicle time, cycling time and walking time (see appendix 6.2 for modelling results). Walking time for the bus mode is calculated as the estimated walking time to the nearest bus stop, while walking times for the cycle and car modes are calculated as the average walking time for these modes.

increase in the travel time of an alternative will lower the probability of the alternative being chosen. Furthermore, a marginal increase in walking and cycling times is expected to lead to a higher decrease in the probability compared to a marginal increase in the time spent travelling in a motor vehicle.

It is expected that an increase in the cost of a mode will decrease the probability of the mode being chosen. The respondents self-reported the pecuniary cost of travelling by bus to work, while the cost of going by car was calculated as 15 pence per mile.[4] Car costs include variable costs such as petrol and servicing costs but not fixed costs such as road tax and insurance, and also neglecting depreciation.[5] Walking and cycling is assumed to be costless.

6.1.2 Choice set formation

When estimating a discrete choice model the available alternatives for each individual must be pre-determined by the researcher (see chapter 4). For each individual in the sample the available choice set is considered to be walk, cycle, bus and car with some exceptions. Going by car is considered unavailable to individuals without a driver's licence and to those living in a household without a car. Going by bus is considered unavailable to individuals who reported to have no bus service available, as well as to those living too close to work for bus to be a practical alternative.[6] Walking to work is

[4] In order to calculate the cost of the bus mode for those respondents who did not report it themselves bus fare was regressed on distance, using the estimated regression equation to calculate the fare (see appendix 6.2).

[5] The variable cost was calculated using a fuel price of 79p per litre, assuming a fuel consumption of 36 miles per gallon. The average costs of tyres, servicing and repairs per mile is calculated using figures given by the Automobile Association.

[6] Bus is not considered to be a practical alternative if the combined distance to and from bus stops exceeds the distance from the commuter's home to her workplace.

considered feasible for individuals commuting one mile or less, while going by bicycle is considered feasible for all respondents commuting three miles or less.[7]

It can be seen from table 6.3 that the majority of individuals who currently walk and cycle to work live within a one and three mile radius of the University respectively. It is also interesting to note that the majority of the respondents who live within a one mile radius of their workplace walk to work (72%) while only about 16% of the individuals who live within a three mile radius cycle. This finding implies that there is considerable scope for increasing the share of individuals cycling to work.

Table 6.3 Cross-tabulation of commuting distance and mode choice

	Dist <=1 miles	Dist <=3 miles	Dist >3 miles
Walk	72%	45%	0%
Cycle	11%	16%	4.5%
Bus	0%	2%	4.5%
Car	17%	37%	91%
Total	119	254	331

It should be pointed out that there are 29 individuals in the sample that walk longer than one mile and 15 individuals that cycle longer than 3 miles to get to work. It could therefore be argued that the definition of the choice set should be extended, since it does not include the choices made by all the sample respondents. This would imply increasing the cut-off points in the definition of the choice set to 2.5 miles for walking and 16 miles for cycling, since these are the longest distances travelled by the two modes. The obvious counter argument to this approach is that assuming that *every* commuter considers walking longer than 1 mile and cycling longer than 3 miles to get to work is unrealistic since only about a quarter of the respondents who actually walk

[7] The British Medical Association (1992) suggests that 3 miles is within cycling distance for most people.

and cycle to work travel farther than that, and that the definition of the choice set should somehow reflect what an average commuter is willing to do. Since the definition of the choice set is admittedly somewhat arbitrary, however, (which, it has to be stressed, is a weakness of virtually all applied travel demand analysis using RP data) it is interesting to ask to what extent changing the definition of the choice set influences the modelling results. This issue will be considered in section 6.2.1.

6.2 Estimation results

The estimation results for the multinomial, nested and mixed logit mode choice models are summarized in tables 6.4 – 6.7 below. In all the models gender, car ownership and the time and cost of the alternatives enter as explanatory variables. In the multinomial logit model presented in table 6.4 (model 1) the attributes of the alternatives (door-to-door travel time and cost) are entered in levels, implying that the marginal utility of a change in an alternative attribute is constant. The coefficients for the cycle, bus and car constants are negative and significant, while the walk constant is normalised to zero for identification purposes. The alternative specific constants represent the mean impact of all variables that are not included in the model that influence the choice of a mode.

The coefficient for car ownership is positive and significant as expected, indicating that the utility of going by car increases significantly with the number of cars the household owns. Note that since the car mode is only considered available for commuters living in a household with at least one car, the car ownership variable represents the increase in the utility of going by car following an increase in car ownership from one car to two cars or more. The coefficient for gender is negative and significant for the bus mode, which implies that females have a significantly higher disutility of going by bicycle to work. This confirms the finding in Dickinson *et al.* (2003).

Table 6.4 Multinomial logit mode choice model

Variable	Alternative	Model 1 (MNL – linear) Coeff.	t-stat.
Constant	Cycle	-2.051	-7.25
Constant	Bus	-2.579	-5.14
Constant	Car	-2.359	-5.62
Female	Cycle	-1.720	-4.54
Bus frequency – 1 or more per hour (ref)			
Bus frequency – less than 1 per hour	Bus	-1.913	-2.52
Number of cars in household	Car	0.603	2.55
Travel time (door-to-door, minutes)	All	-0.048	-2.90
Cost (pence)	All	-0.010	-2.44
Observations		585	
Log-likelihood: constant only L(c)		-241.543	
Log-likelihood: final value L(β)		-212.462	
Rho-squared (with L(c))		0.120	
Rho-squared adjusted (with L(c))		0.113	

As expected an increase in the bus frequency leads to an increase in the probability of choosing bus. Although the difference between having an hourly service or a more frequent service was not found to be significant (and hence this variable was dropped from the model), there is a significant difference between having and not having an hourly service. This implies that the provision of an hourly bus service is an important incentive in order to encourage more commuters to travel by public transport. The coefficients for (door-to-door) travel time and cost are negative and significant at the 5% and 10% level respectively.

The income[8] and age of the commuters were not found to be significant determinants of mode choice and therefore these variables are not included in the final model specifications reported in table 6.4. Some of the influence of income on mode choice will nevertheless be incorporated through the car ownership variable, as income is found to have a strong influence on households' car ownership level (see chapter 5).

As discussed in chapter 2 the marginal disutility of an increase in travel time/ cost may not be constant, but a function of travel times/ costs. Several studies have found that allowing for non-linearities in the utility specification improves the fit of the model (Gaudry and Wills, 1978; Gaudry *et al.*, 1989; Jara-Díaz and Videla, 1989). As outlined in chapter 2, the second order approximation to the expenditure rate model suggests including second order terms in the specification of the indirect utility function, to represent that the disutility of travel time/ cost increases as travel times/ costs increase. Re-specifying model 1 by subdividing door-to-door travel time into in-vehicle/ cycling time and walking time and including quadratic time-variables leads to a substantial improvement in model fit, but the signs of the coefficients for the quadratic terms are not consistent with the theoretical model, since in that case the coefficients should be negative to reflect the increasing marginal disutility of an increase in travel time (see table 6.5). The coefficients are all positive (and highly

[8] The survey data do not include direct information about income, partially due to concerns that including an income question in the survey would cause some individuals not to respond. The data includes information on occupation, however, and a proxy for income was derived by dividing respondents into high and low income groups on the basis of their occupational rank in the University. Because of the lack of income data it was decided to include income as a dummy variable in the model, rather than using the wage rate or expenditure rate specifications described in chapter 2. Another approach would be to segment the coefficients for the alternative attributes based on the income dummy, but this was decided against since the coefficient estimates in this case should be expected to be imprecise as a result of the relatively low sample size (since only 8 high-income individuals chose bus, for example, the coefficient for the 'high-income bus frequency' variable would be estimated on the basis of those 8 observations only).

significant), however, which implies that the marginal disutility of an increase in travel time is *decreasing* with travel time.

Table 6.5 Multinomial logit mode choice model with quadratic terms

		Model 2 (MNL - quadratic)	
Variable	Alternative	Coeff.	t-stat.
Constant	Cycle	0.398	0.40
Constant	Bus	-5.520	-6.76
Constant	Car	-6.249	-8.19
Female	Cycle	-1.623	-3.70
Bus frequency – 1 or more per hour (ref)			
Bus frequency – less than 1 per hour	Bus	-1.539	-1.74
Number of cars in household	Car	0.560	1.86
Walking time (minutes)	All	-0.608	-7.62
Cycling time (minutes)	Cycle	-1.249	-6.52
In-vehicle time (minutes)	Bus, Car	-0.068	-1.35
Walking time squared (minutes)	All	0.0140	6.29
Cycling time squared (minutes)	Cycle	0.0430	5.94
In-vehicle time squared (minutes)	Bus, Car	0.0024	2.42
Cost (pence)	All	-0.0065	-1.35
Observations			
Log-likelihood: constant only L(c)		-148.908	
Log-likelihood: final value L(β)		-241.543	
Rho-squared (with L(c))		0.384	
Rho-squared adjusted (with L(c))		0.375	

Table 6.6 Multinomial logit mode choice models – square-root and log specifications

		Model 3 (MNL – square-root)		Model 4 (MNL – log)	
Variable	Alternative	Coeff.	t-stat.	Coeff.	t-stat.
Constant	Cycle	-3.062	-3.38	-2.243	-2.39
Constant	Bus	-3.522	-4.40	-2.942	-3.61
Constant	Car	-4.945	-6.54	-4.405	-5.47
Female	Cycle	-1.879	-4.67	-2.150	-5.06
Bus frequency – 1 or more per hour (ref)					
Bus frequency – less than 1 per hour	Bus	-1.567	-2.01	-1.482	-1.90
Number of cars in household	Car	0.640	2.39	0.533	1.94
Square-root of walking time (minutes)	All	-1.408	-7.48		
Square-root of cycling time (minutes)	Cycle	-0.867	-3.11		
Square-root of in-vehicle time (minutes)	Bus, Car	-0.179	-0.88		
Log of walking time (minutes)	All			-1.794	-7.89
Log of cycling time (minutes)	Cycle			-1.837	-4.39
Log of in-vehicle time (minutes)	Bus, Car			-0.615	-1.90
Cost (pence)	All	-0.013	-2.81	-0.012	-2.81
Observations				585	
Log-likelihood: constant only L(c)		-241.54		-241.543	
Log-likelihood: final value L(β)		-177.72		-167.532	
Rho-squared (with L(c))		0.264		0.306	
Rho-squared adjusted (with L(c))		0.256		0.299	

Although incompatible with the utility maximising model, this result is acceptable from a behavioural point of view (and in line with some of the results in Gaudry *et al.*, 1989). It should be noted, however, that a problem with this quadratic specification is that as travel times increase utility will eventually be increasing with travel time because of the positive quadratic term, which is illogical. It is therefore necessary to re-specify the model by ensuring that an increase in travel time always leads to a decrease in the utility of a mode, while allowing for a decreasing marginal utility of travel time. There are various functional forms with this property, including the log, square-root and Box-Cox transformations[9]. The square-root and log specifications were estimated, and it was found that the specification with the time variables in logs yielded the superior data fit (see table 6.6).[10] In addition, all the travel time components in the log specification have the expected sign and are significant at the 5% level, except the coefficient for in-vehicle time, which is significant at the 10% level. It can be seen that this specification leads to a considerable increase in the rho-bar squared compared to model 1.

A number of nested logit models were fitted based on the log specification of the MNL model to allow for a more flexible substitution pattern between the modes. On the basis of the score on goodness-of -fit measures (rho-squared, rho-bar squared) as well as compliance with the utility maximising condition (see chapter 3), the superior nesting structure was found to be car and walk in a common nest and cycle and bus in separate nests.[11] This model structure implies that car and walk are closer substitutes than car and bicycle/ public transport.[12] The results are presented in table 6.7, columns 7 and 8 (model 5).

It can be seen that the inclusive value (IV) parameter is lower than one, which implies that the model is consistent with utility maximising behaviour. The inclusive value parameter is significantly different from unity, indicating that the walk and car alternatives are correlated and that as a result the IIA property is rejected. Furthermore there is a marked increase in the rho-bar squared compared to model 4, indicating that the data fit of the nested logit model is superior to the MNL model. Apart from the coefficient for car ownership, which is now insignificant, the sign and significance of

[9] The Box-Cox transformation leads to a decreasing marginal disutility of travel time given that $\lambda<1$, where the Box-Cox transformation of the time variable T is given by $T^{(\lambda)} = \ln T$ when $\lambda=0$ and $T^{(\lambda)} = (T^{\lambda} - 1)/\lambda$ otherwise (see also chapter 7).

[10] It was also attempted to estimate the Box-Cox model, but this did not converge.

[11] The estimation results for some alternative nesting structures are reported in appendix 6.3.

[12] This holds only, of course, when walking is an available alternative to driving.

the remaining coefficients are not substantially different to those in models 1-4, indicating that the findings are robust.

Table 6.7 Nested and mixed logit mode choice models

Variable	Alternative		Model 5 (NL – log) Coeff.	Model 5 (NL – log) t-stat.	Model 6 (ML - log) Coeff.	Model 6 (ML - log) t-stat.
Constant	Cycle	Mean	-2.379	-2.54	-3.474	-3.51
Constant	Bus	Mean	-2.408	-2.84	-2.254	-2.04
Constant	Car	Mean	-3.482	-4.07	-5.797	-5.90
Female	Cycle	Mean	-1.976	-5.20	-2.979	-4.36
Bus frequency – 1 or more per hour (ref)						
Bus frequency – less than 1 per hour	Bus	Mean	-1.572	-1.96	-1.301	-1.48
Number of cars in household		Mean	0.369	1.24	0.717	2.11
Log of walking time (minutes)	All	Mean	-1.638	-6.75	-2.550	-6.12
Log of cycling time (minutes)	Cycle	Mean	-1.509	-4.30	-3.150	-4.16
		Std. Dev.			1.161	3.88
Log of in-vehicle time (minutes)	Bus, Car	Mean	-0.550	-1.83	-0.966	-2.04
Cost (pence)	All	Mean	-0.013	-3.18	-0.013	-2.04
IV parameter (t-stat w.r.t. 1)	Walk, Car		0.533	-2.03		
Observations			585		585	
Log-likelihood: constant only L(c)			-241.543		-241.543	
Log-likelihood: final value L(β)			-165.286		-162.05	
Rho-squared (with L(c))			0.316		0.329	
Rho-squared adjusted (with L(c))			0.307		0.321	

As mentioned in chapter 4 a crucial question that faces the analyst when applying the mixed logit model is which parameters that should be allowed to vary and which distribution to use for the random parameters. As in Hensher (2001b), Carlsson (2003) and Alpizar and Carlsson (2003) the cost variable is specified to be fixed, while the time parameters are specified to follow a normal distribution in the model.[13] Fixing the cost coefficient is convenient for several reasons: it ensures that the value of time has finite moments[14] and that the sign of the cost variable is negative for all respon

[13] It was also attempted to specify the time coefficients to follow a triangular distribution as in Hensher (2001a), but this resulted in a model with a lower rho-bar squared.

[14] When the time coefficient is random and the cost coefficient fixed the distribution of the value of time is distributed in the same way as the time coefficient (Revelt and Train, 1999; Carlsson, 2003).

dents.[15] The standard deviations of the coefficients of the walking and in-vehicle time variables were found to be insignificant, however, and constraining the standard deviations of those coefficients to equal zero did not lead to a decrease in the rho-bar squared. Table 6.7 reports the estimation results of the more parsimonious model (model 6) with fixed walking and in-vehicle time coefficients and normally distributed cycling time coefficient (the full model is reported in appendix 6.4).[16] This model structure implies that the error variance of the cycle mode is higher than that of the other alternatives. The alternatives remain uncorrelated, however, since the cycling time variable only enters the utility function of the cycle mode.

Some other model specifications based on the ML model were also attempted: 1) A mixed nested logit model, which can be seen as a combination of models 5 and 6. This specification does not lead to an increase in the rho-bar squared compared to model 6, however, and the inclusive value parameter is insignificant (see appendix 6.4). 2) An ML model with a normally distributed cycling time coefficient and heteroscedastic error components. As shown by Walker *et al.* (2003) one of the error components need to be constrained to equal zero for this model to identified. Furthermore, the choice of normalisation is not arbitrary, since different normalisations may lead to different modelling results/ goodness of fit. Hence 4 versions of this model were attempted, normalising each error component to zero in turn. The only model specification that converged, however, was the model with the error component for the car mode normalised to zero. It can be seen that this model

[15] Alternatively, to ensure that its sign is positive the coefficient for the cost variable could be specified to be log-normally distributed. Various models with log-normally distributed coefficients were attempted, but these models did not converge.
[16] Both models were estimated using Kenneth Train's GAUSS code with 500 Halton draws which can be freely downloaded at http://elsa.berkeley.edu/~train/software.html. The models were run in OxGauss which can be freely downloaded at www.doornik.com.

leads to a slight increase in the rho-bar squared compared to model 6, but the coefficients for the error components were all found to be insignificant (see appendix 6.5). As a result the more parsimonious model 6 was decided to be the preferred specification.

The sign and significance of the coefficients in model 6 are similar to those in models 1-5. All the time coefficients are significant at the 5% level and have the expected sign along with the coefficients for cost, gender and car ownership. The coefficient for bus frequency, however, has the expected sign but is insignificant. This is likely to be a result of the relatively small number of individuals in the sample choosing bus, which makes it harder to obtain precise estimates of the bus-specific coefficients.

6.2.1 Elasticities

Aggregate elasticities provide a summary measure of the likely response to a change in an alternative attribute and are therefore valuable tools that can assist in developing efficient car-reduction policies. The aggregate elasticities derived using models 4 – 6 are reported in table 6.10 – 6.12 below. The elasticities are calculated by simulating the change in the modal shares following a 1% increase in a given alternative attribute using the method of sample enumeration (Ben-Akiva and Lerman, 1985). Since the models do not allow for traffic generation, these elasticities should be interpreted as mode-choice elasticities.

When comparing the elasticity estimates derived from the three models some patterns emerge. The cross elasticities of the walk mode with respect to the attributes of the car mode derived from the nested logit model are higher than those derived from the other models. This is a result of the correlation between these two modes facilitated by the chosen nesting structure. Also, the direct elasticities of the cycle mode derived from the mixed logit model are lower than the multinomial and nested logit elasticities. This can be explained by the higher error variance of the cycling mode relative to the other modes in the mixed logit model, resulting in less weight being placed on the deterministic elements of the utility function. Furthermore, it can be seen that the mixed logit direct cost elasticities for the bus and car modes are substantially lower than those derived from the other models.

It can be seen from the tables that contrary to expectations the highest elasticity (in absolute value) is the bus fare elasticity (1.156 – 1.496). Indeed this is higher than what is found in most studies of urban commuting. Dargay and Hanly (2002), find that the short-run bus fare elasticity for England as a whole is around –0.4 and that elasticities at the county level vary widely (between 0 and –1.6), although the authors suggest that the county specific elasticities should be interpreted with caution due to the small number of observations. In a comprehensive review, Dargay and Hanly (1999) find that the average short-run bus fare elasticity is -0.3.[17] The high elasticity estimate in the present study is likely to be related to the fact that bus fares in the St Andrews area have doubled over the last decade, as there is evidence that the demand for public transport is more price sensitive at higher fare levels (Dargay and Hanly, 2002). Since the elasticity measures the percentage change in the modal share from the base share, however, the increase in the share of bus users is not as substantial as the elasticity estimate might imply. Nevertheless, the estimate suggests that subsidising bus fares would be an important factor to incentivise more commuters

[17] It should be noted that the elasticity estimates reported in Dargay and Hanley are regular elasticities as they also take traffic generation into account. Oum *et al.* (1992) argue that mode-choice elasticities may serve as lower bounds for regular elasticities in terms of absolute values.

to use public transport. The walking time elasticity for the bus mode is also higher than what is found in most studies, indicating that decreasing walking times by increasing the number of bus stops will substantially increase the share of commuters travelling by bus. The bus in-vehicle time elasticity is markedly lower than the walking time elasticity, which implies that commuters are less sensitive to changes in the time spent travelling by bus than to changes in access and egress times.

Table 6.10 Aggregate elasticities. MNL model.

	Percentage change in the probability of choosing			
Due to a 1% increase in	Walk	Cycle	Bus	Car
Cycling time	0.142	-0.947	0.146	0.084
In-vehicle time (Bus)	0.000	0.011	-0.351	0.016
In-vehicle time (Car)	0.051	0.191	0.292	-0.056
Walking time (Walk)	-0.290	0.287	0.000	0.043
Walking time (Cycle)	0.132	-0.840	0.117	0.071
Walking time (Bus)	0.005	0.043	-1.024	0.044
Walking time (Car)	0.148	0.574	0.877	-0.167
Bus costs	0.000	0.032	-1.404	0.065
Car costs	0.015	0.191	1.053	-0.084

Table 6.11 Aggregate elasticities. NL model.

Due to a 1% increase in	Percentage change in the probability of choosing Walk	Cycle	Bus	Car
Cycling time	0.130	-0.830	0.117	0.071
In-vehicle time (Bus)	0.005	0.022	-0.323	0.015
In-vehicle time (Car)	0.065	0.186	0.264	-0.054
Walking time (Walk)	-0.318	0.284	0.029	0.049
Walking time (Cycle)	0.135	-0.819	0.117	0.068
Walking time (Bus)	0.005	0.055	-0.968	0.041
Walking time (Car)	0.183	0.579	0.821	-0.164
Bus costs	0.005	0.044	-1.496	0.068
Car costs	0.022	0.218	1.144	-0.091

Table 6.12 Aggregate elasticities. ML model.

Due to a 1% increase in	Percentage change in the probability of choosing Walk	Cycle	Bus	Car
Cycling time	0.140	-0.802	0.151	0.064
In-vehicle time (Bus)	0.001	0.016	-0.441	0.019
In-vehicle time (Car)	0.060	0.175	0.385	-0.060
Walking time (Walk)	-0.320	0.326	0.013	0.046
Walking time (Cycle)	0.140	-0.721	0.114	0.054
Walking time (Bus)	0.002	0.043	-1.160	0.049
Walking time (Car)	0.158	0.465	1.044	-0.160
Bus costs	0.001	0.022	-1.156	0.052
Car costs	0.013	0.105	0.875	-0.060

The direct car cost elasticity is found to lie in the range 0.060 – 0.091, which is comparable in size but somewhat lower than the car cost elasticity reported in most studies of urban commuting (Oum *et al.*, 1992, provide a review of car cost elasticities derived from discrete choice models). This confirms the prior expectation that increasing the cost of driving is not likely to be an effective deterrent to car use unless a convenient alternative mode of transport is provided. The walking time and in-vehicle time elasticities for the car mode are also found to be relatively low, indicating that an increase in travel time will not lead to a substantial decrease in car use. Bus is found to be the closest substitute to car, as the cross elasticities with respect to a change in a car attribute is higher for bus than for the other modes. Given that walking and cycling are only considered available for relatively short commutes this result is expected. The direct walking and cycling time elasticities are found to lie in the region 0.290 – 0.320 and 0.802 – 0.947 for the walk and cycle modes respectively. Given that the time spent walking and cycling is closely related to commuting distance, these elasticity estimates reflect how the probability of walking and cycling to work changes as a result of increasing/ decreasing the distance from the home to the workplace.

Since bus frequency is represented as a dummy variable in the model it is not possible to calculate the elasticity with respect to an increase in bus frequency. Instead the method of sample enumeration is used to simulate the effect of increasing the bus frequency such that all the commuters in the sample have access to an hourly service. The results from the simulation exercise are presented in table 6.13 below.

Table 6.13 Change in modal shares following an increase in bus frequency

	Walk	Cycle	Bus	Car
MNL	-0.31	-1.27	23.67	-1.32
NL	-0.48	-1.47	25.79	-1.42
ML	-0.15	-0.76	18.36	-0.95

It can be seen that increasing the bus frequency is predicted to lead to a substantial increase in the share of commuters who travel by bus, although the predicted share derived from the mixed logit model is markedly lower than that of the multinomial and nested logit models. Nevertheless, this finding together with the estimated elasticities suggests that a policy directed towards increasing the use of public transport for commuting should focus on subsidising bus fares as well as providing an hourly bus service for as many commuters as considered possible given the dispersed nature of the St Andrews area. This policy is likely to be particularly effective if combined with parking charges and/ or increases in the petrol tax in order to deter driving.

6.2.2 Weighted Elasticities

The elasticities in section 6.2.2 are caluculated using the sample data and will only be valid for the population of commuters in the University of St Andrews if the sample is representative of the population. In order to investigate whether this is in fact the case, the elasticities are re-estimated using the population shares of gender, occupation type (academic vs. non-academic) and age as weights (see table 6.1). It can be seen from tables 6.14 – 6.16 that the elasticity estimates derived using the re-weighted data are similar to the elasticities derived using the un-weighted sample data and that no clear relationship between the two can be detected (e.g. the estimates are not consistently smaller or larger, for instance).

Table 6.14 Aggregate elasticities – rescaled data. MNL model.

Due to a 1% increase in	Percentage change in the probability of choosing Walk	Cycle	Bus	Car
Cycling time	0.138	-0.959	0.156	0.083
In-vehicle time (Bus)	0.000	0.032	-0.402	0.023
In-vehicle time (Car)	0.046	0.194	0.313	-0.063
Walking time (Walk)	-0.296	0.291	0.045	0.044
Walking time (Cycle)	0.133	-0.851	0.134	0.071
Walking time (Bus)	0.005	0.075	-1.139	0.065
Walking time (Car)	0.143	0.571	0.960	-0.186
Bus costs	0.005	0.054	-1.630	0.101
Car costs	0.015	0.194	1.206	-0.113

Table 6.15 Aggregate elasticities – rescaled data. NL model.

Due to a 1% increase in	Percentage change in the probability of choosing Walk	Cycle	Bus	Car
Cycling time	0.125	-0.842	0.131	0.071
In-vehicle time (Bus)	0.005	0.033	-0.370	0.021
In-vehicle time (Car)	0.060	0.188	0.283	-0.060
Walking time (Walk)	-0.325	0.277	0.044	0.050
Walking time (Cycle)	0.136	-0.831	0.131	0.066
Walking time (Bus)	0.011	0.078	-1.067	0.060
Walking time (Car)	0.179	0.565	0.892	-0.183
Bus costs	0.011	0.067	-1.763	0.109
Car costs	0.022	0.222	1.328	-0.124

Table 6.16 Aggregate elasticities – rescaled data. ML model.

Due to a 1% increase in	Percentage change in the probability of choosing Walk	Cycle	Bus	Car
Cycling time	0.139	-0.797	0.153	0.065
In-vehicle time (Bus)	0.000	0.021	-0.428	0.019
In-vehicle time (Car)	0.062	0.170	0.398	-0.059
Walking time (Walk)	-0.320	0.329	0.031	0.046
Walking time (Cycle)	0.155	-0.776	0.122	0.059
Walking time (Bus)	0.005	0.043	-1.162	0.050
Walking time (Car)	0.160	0.468	1.040	-0.159
Bus costs	0.000	0.021	-1.131	0.053
Car costs	0.015	0.106	0.887	-0.059

6.2.3 The value of travel time

Prior to undertaking investments in transport infrastructure it is important to assess the benefits of the investment. It is generally held in the literature that a significant proportion of the benefits of infrastructure improvements is due to road users' travel time savings. In a recent study, Mackie *et al.* (2001) suggest that the value of travel time savings accounts for 80% of the monetised benefits within the cost benefit analysis of major road schemes in the UK. It follows that in order to make well-informed investment decisions it is crucial to obtain as precise estimates of the subjective value of time (*SVOT*) as possible, and in many countries the authorities have commissioned studies estimating *SVOT* both for commuting and other types of trips (the UK, the Netherlands and the Scandinavian countries among others). Since the multinomial, nested and mixed logit models are rooted in microeconomic theory, the value of time can be shown to be given by the ratio of the travel time and cost coefficients when the alternative attributes enter in levels in the model (see chapter 2). When travel time enters in the log form (as in models 4 – 6), *SVOT* is a decreasing function of travel time:

$$SVOT = \frac{\beta_T}{\beta_C}\frac{1}{T} \tag{6.1}$$

where β_T and β_C are the time and cost coefficients for a given mode and T is the travel time for that mode. The estimated values of time evaluated at the average time for each travel time component, using models 4 – 6, are given in table 6.17 below.

Table 6.17 Values of time (in pence per minute)

	Walking time	Cycling time	In-vehicle time (Bus, Car)
MNL	20.28	13.99	2.69
NL	17.09	10.60	2.22
ML – Mean	26.61	22.14	3.90
ML – Std. Dev.		8.16	

It can be seen that the commuters are on average willing to pay more for a decrease in the time spent walking compared to a decrease in cycling time, which indicates that walking is considered more onerous than cycling. Furthermore, a marginal decrease in cycling time is valued higher than a marginal decrease in in-vehicle time, indicating that cycling is considered more onerous than travelling in a motor vehicle. The significant standard deviation of the cycling time coefficient in the mixed logit model implies that some commuters have a comparatively low value of cycling time[18], while others have comparatively high values of cycling time (29% of the commuters in the

[18] 0.33% of the commuters in the sample are found to have a positive cycling time coefficient. It is not unlikely that for some cycling enthusiasts the time spent cycling is a good rather than a bad.

sample find cycling more onerous than walking). It is interesting to note that the value of time estimates derived from the ML model are substantially higher than those derived from the MNL and NL models, which is consistent with the finding in Hensher (2001a), but not the findings in Brownstone and Small (2003), Nielsen and Jovicic (2003) and Nielsen and Sørensen (2004).

In a review of British studies reporting the value of in-vehicle travel time, Wardman (1998) finds an average value of 5.64 pence per minute which is considerably higher than the average value of in-vehicle time found in the present study[19]. It is likely that the low *SVOT* estimate reflects the fact that roads in the St Andrews area are relatively uncongested. As mentioned in Chapter 5 Calfee and Winston (1998) and Hensher (2001a) find, using data from the USA and New Zealand respectively, that the value of time spent travelling under congested conditions is substantially higher than time spent travelling in free-flow traffic.[20] Since the UK average value of in-vehicle time is calculated using data from urban as well as rural areas and therefore partially reflects substantially more congested commuting conditions than those in the St Andrews area, the national average *SVOT* should be expected to be higher than that in the present study.

The average value of walking time is found to be about 7-8 times higher than the estimated value of in-vehicle time, and about 3-5 times higher than the UK average in-vehicle *SVOT*. This is comparable to the findings of studies of commuting in urban areas. The average value of cycling time is about 5-6 times higher than the estimated value of in-vehicle time and about 2-4 times higher than the national

[19] Given that most of the studies in the review are likely to have used the MNL model to derive the estimate of *SVOT*, the most representative estimate for comparison with the review is perhaps that derived from the MNL model.
[20] In Calfee and Winston (1998) the value of congested travel time is found to be 3 times higher than that of uncongested/ free-flow travel time. A similar result is obtained by Hensher (2001).

average in-vehicle *SVOT*. The author knows of no other studies reporting the value of cycling time for commuting trips in the UK. Given the relatively favourable cycling conditions in St Andrews, the value of cycling time found in the present study is likely to be lower than that in urban areas where cycling by many is perceived to be dangerous due to heavy traffic, particularly in the absence of segregated cycle lanes which are more common in continental cities.[21] As there are few studies reporting the value of cycling time to date, more research is needed to investigate how the value of cycling time varies between geographical locations and according to the facilities provided. Given that the value of cycling time is also shown to vary with unobserved personal characteristics, incorporating random taste variation in the modelling framework will help disentangling the effects of changes in cycling conditions to individual specific preferences towards cycling.

It was also attempted to simulate the mean, standard deviation and median of the value of time using the method of simulation of multivariate normal variates (MVNS) described in section 3.7. It was found, however, that the mean estimate of *SVOT* was extremely sensitive to a relatively small number of draws of the cost coefficient which were very close to zero, resulting in a very high value of time (this was reflected in a very high standard deviation, many times the size of the mean *SVOT*). This problem did not go away by increasing the number of draws in the simulation. It was found that removing the 1% of the sample with the highest value of time led to more stable results, but since this approach is rather *ad hoc* the results are not reported here. Interestingly, however, the median estimates of *SVOT* are very similar to the point estimates, confirming the finding in Hensher and Greene (2003). The median estimates calculated using 10000 draws of the coefficients are presented

[21] Noland and Kunreuther (1995) and Ortúzar *et al.* (2000) investigate how changes in travel conditions influence individuals' choice of travelling by bicycle.

in table 6.18 below. It should be pointed out that the estimates varied very little with the number of draws when the number was above 1000.

Table 6.18 Median values of time in pence per minute calculated using MVNS (10000 draws)

	Walking time	Cycling time	In-vehicle time (Bus, Car)
MNL	19.57	13.58	3.15
NL	16.95	10.51	2.19
ML – Mean	26.99	22.48	3.93
ML – Std. Dev.		8.28	

It is also possible to use the modelling results to calculate the sample respondents' average willingness to pay (WTP) to have access to an hourly bus service. This is given by:

$$WTP_{FREQ} = \frac{\beta_{FREQ}}{\beta_C} \qquad (6.2)$$

where β_{FREQ} is the coefficient for the dummy indicating that the respondent does not have access to an hourly bus service. The estimated WTP for access to an hourly bus service derived using models 4 – 6 is given in table 6.19 below.

Table 6.19 Willingness to pay for an hourly bus service

	MNL	NL	MNL
Coefficient for bus frequency	-1.482	-1.572	-1.301
Coefficient for cost	-0.012	-0.013	-0.013
Willingness to pay (in pence)	124	121	100

It can be seen from the table that the commuters' are on average willing to pay 100 - 124 pence per trip to have access to an hourly bus service, which can be compared to the average bus fare of 196 pence. This estimate can be used as input to a cost-benefit analysis evaluating the desirability of improving the supply of public transport in those areas around St Andrews which do not have an hourly bus connection with the town centre. From a methodological point of view it is interesting to point out that in this case the higher WTP estimates are given by the MNL and NL models, supporting the statement by Train (1997) that the relative size of WTP estimates in MNL and ML models cannot be generalized, but must be evaluated on a case for case basis.

6.2.4 Sensitivity analysis

As described in section 6.1.2 walking to work is considered feasible only for individuals commuting one mile or less in all the models, which means that individuals with longer commutes are assumed not to consider walking to work as an alternative to travelling by car, bus or bicycle. As mentioned previously, however, 29 individuals in the sample walk a longer distance to get to work and it is therefore interesting to ask what impact increasing the upper limit on the walking distance has on the modelling results. Tables 6.20 and 6.21 presents the results from re-specifying models 1, 4, 5 and 6 by expanding the choice set to include walk for those 45 individuals who live between 1 and 1.5 miles away from work. The choice set definition now includes 93% (107 out of 115) of all the individuals who walk to work.

It can be seen from the tables that the walking time coefficients in the new models (4 - 6) are consistently higher in proportion to the coefficients for the other time components, indicating that the average marginal disutility of an increase in walking time relative to cycling and in-vehicle time increases when the choice set is expanded. Further, since the cost coefficient remains the same in all the models, this also implies that the value of walking time increases relative to the value of cycling and in-vehicle time. In absolute terms, however, the findings are mixed. The walking time coefficient increases in model 4 but decreases slightly in model 5 – 6, while the coefficients for in-vehicle and cycling time decreases in all the models.

It is clear from the previous discussion that the definition of the choice set has an impact on the estimates of the coefficients in the models, which in turn affect policy parameters such as elasticity and value of time estimates. It seems difficult from the findings documented here, however, to determine *a priori* in which direction this effect will work (whether demand will be more or less responsive to changes in alternative attributes, and whether the value of the various components of travel time will increase or decrease). It will not be attempted to resolve the issue of choice set specification in the present thesis, other than concluding that since the definition of the choice set clearly has an impact on the coefficient estimates, and therefore also the policy parameters that can be derived from the model, more attention to this issue should be paid in future research.

Table 6.20 Multinomial logit mode choice models – extended choice set

Variable	Alternative	Model 1 (MNL - linear) Coeff.	Model 1 t-stat.	Model 4 (MNL – log) Coeff.	Model 4 t-stat.
Constant	Cycle	-1.809	-6.66	-2.592	-2.89
Constant	Bus	-1.828	-3.83	-2.593	-3.57
Constant	Car	-2.111	-5.20	-4.202	-5.86
Female	Cycle	-1.528	-4.19	-1.850	-4.62
Bus frequency – 1 or more per hour (ref)					
Bus frequency – less than 1 per hour	Bus	-1.963	-2.57	-1.451	-1.86
Number of cars in household	Car	0.657	2.79	0.651	2.47
Travel time (door-to-door, minutes)	All	-0.068	-4.05		
Log of walking time (minutes)	All			-1.855	-8.35
Log of cycling time (minutes)	Cycle			-1.536	-3.91
Log of in-vehicle time (minutes)	Bus, Car			-0.500	-1.83
Cost (pence)	All	-0.011	-2.77	-0.012	-2.73
Observations		585		585	
Log-likelihood: constant only L(c)		-234.887		-190.232	
Log-likelihood: final value L(β)		-268.616		-268.616	
Rho-squared (with L(c))		0.126		0.292	
Rho-squared adjusted (with L(c))		0.118		0.284	

Table 6.21 Nested and mixed logit mode choice models – extended choice set

Variable	Alternative		Model 5 (NL) Coeff.	Model 5 t-stat.	Model 6 (ML) Coeff.	Model 6 t-stat.
Constant	Cycle	Mean	-2.626	-2.91	-2.580	-2.09
Constant	Bus	Mean	-2.422	-3.33	-3.313	-3.51
Constant	Car	Mean	-3.517	-4.67	-5.691	-5.53
Female	Cycle	Mean	-1.754	-4.86	-2.652	-3.49
Bus frequency – 1 or more per hour (ref)						
Bus frequency – less than 1 per hour	Bus	Mean	-1.548	-1.94	-1.207	-1.46
Number of cars in household			0.503	2.02	0.897	2.45
Log of walking time (minutes)	All	Mean	-1.615	-6.51	-2.534	-6.96
Log of cycling time (minutes)	Cycle	Mean	-1.212	-3.72	-2.726	-4.00
		Std. Dev.			1.169	3.56
Log of in-vehicle time (minutes)	Bus, Car	Mean	-0.336	-1.46	-0.650	-1.91
Cost (pence)	All	Mean	-0.013	-3.08	-0.013	-2.45
IV parameter (t-stat w.r.t. 1)	Walk, Car		0.551	-1.81		
Observations			585		585	
Log-likelihood: constant only L(c)			-187.394		-184.040	
Log-likelihood: final value L(β)			-268.616		-268.616	
Rho-squared (with L(c))			0.302		0.315	
Rho-squared adjusted (with L(c))			0.294		0.307	

6.3 Concluding remarks

This chapter has developed multinomial, nested and mixed logit mode choice models using data on commuters in the University of St Andrews. As St Andrews is located in a rural area with limited public transport supply it was expected that key policy variables such as elasticities and values of time would differ from those reported in studies of commuting in larger urban areas. It was found that the direct elasticities of the car mode were comparable to the estimates of studies reported in studies of urban commuting, while the demand for public transport was found to be considerably more elastic. Although this is partially a result of the fact that bus has a substantially lower market share in St Andrews compared to larger towns and cities, the finding nevertheless indicates that there is scope for increased use of public transport for commuting in St Andrews and other small towns in rural locations. The values of in-vehicle travel time were found to be lower than in most studies of urban commuting, reflecting that the roads in the St Andrews area are relatively uncongested. The value of walking time is found to be about 7-8 times higher than the value of in-vehicle time, while the value of cycling time is, on average, about 60% - 80% of the value of walking time. More research is needed to investigate how the value of cycling time varies across geographical locations and according to the facilities provided.

Appendix 6.1 Travel survey questionnaire - RP questions.

PART 1 Please be assured that all the information in the survey is strictly confidential

A About Today's Trip

1a What mode of transport did you use to get to work today? (if you used more than one mode tick the one that involved the longest distance)

Car (alone) ☐
Car (with others) ☐
Bus ☐
Bicycle ☐
Walk ☐

Other (please specify):

____

1b What other modes of transport (if any) did you use to get to work?

____

2 How long did the trip take?
(door-to-door, one way)

_______________ minutes

3 How far did you travel?

_______________ miles

4 If you came by car, bus or cycle, how much time did you spend walking (from your parking place to your workplace / to and from bus stops)?

_______________ minutes

5 What time did you arrive at your workplace?

____

B About Public Transport

6 How far is the closest bus stop to your home? (if it is closer than a mile give your answer in parts of miles, for example 1/4)

_______________ miles

Don't know ☐

7 What is the peak time (7:00 - 8:30) frequency of buses going in the direction of your workplace at this bus stop?

A bus leaves every _______________ minutes

Don't know ☐

8 How far is the closest bus stop to your workplace?

_______________ miles

Don't know ☐

9 What is the peak time (16:00 - 18:00) frequency of buses going in the direction of your home at this bus stop?

A bus leaves every _______________ minutes

Don't know ☐

10 Do you / would you need to change buses in order to travel by bus to work?

Yes ☐
No ☐
Don't know ☐

If yes, how many times?

____

11 How much does a bus ticket to work cost?

________________ pence

Don't know ☐

12 When did you last use the bus to get to work?

Today ☐
1-6 days ago ☐
1-4 weeks ago ☐
5+ weeks ago ☐
Never ☐

C About You And Your Household

13 Are you?

Male ☐
Female ☐

14 What is your age group?

Under 30 ☐
30 to 39 ☐
40 to 49 ☐
50 or over ☐

15 What is your home postcode?

____

16 What is your usual place of work? (name of University building)

____

17 What is your occupation / job title?

____

18 How many cars does your household currently own?

____

19 Do you have a driver's licence?

____

20 How many workers aged 16 or over (full tim and part time) are there in your household?

____

PLEASE COMPLETE SECTION D IF YOU CAME CAR TO WORK TODAY. IF YOU DID NOT COME BY CAR, PLEASE GO STRAIGHT TO SECTION E

D About Parking

21 Where did you park today?

At one of the University car parks ☐
Free parking in nearby street / car park ☐
Paid parking in nearby street / car park ☐

22 If the University was charging 50 pence per day for parking at University car parks, how likely is it that you would choose an alternative mode of transport to go to work today?

Not likely ☐
Very likely ☐
Not sure ☐

23 What if the charge was 1 pound per day?

Not likely ☐
Very likely ☐
Not sure ☐

E Additional Comments

24 If you have any additional comments about your travel to work, please use the space at the back of the last sheet of this questionnaire.

Appendix 6.2 Estimated OLS regression equations for calculating travel times/ cost for non-chosen alternatives (t-statistics in brackets).

$$\underset{}{WALKT} = \underset{(37.49)}{16.32}*SQRT(DIST), \qquad R^2 = 0.70$$

$$CYCLET = \underset{(17.54)}{8.54}*SQRT(DIST), \qquad R^2 = 0.54$$

$$INVT_{BUS} = \underset{(12.82)}{8.96}*SQRT(DIST), \qquad R^2 = 0.57$$

$$INVT_{CAR} = \underset{(63.39)}{1.58}*DIST + \underset{(6.74)}{3.69}*STAD, \qquad R^2 = 0.76$$

$$BFARE = \underset{(4.79)}{87.95} + \underset{(8.49)}{13.91}*DIST, \qquad R^2 = 0.48$$

Note: The walking time for the (non-chosen) bus alternative is derived by inserting the sum of the distance to and from bus stops in the equation for walking time. Walking times for the (non-chosen) cycle and car alternatives were calculated at their average values (1.18 and 2.77 minutes respectively).

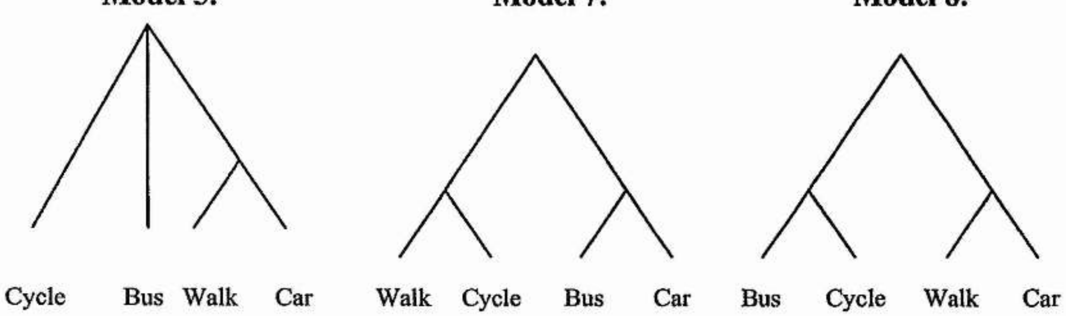
Variable definitions:

WALKT = Walking time in minutes
CYCLET = Cycling time in minutes
INVT = In-vehicle time in minutes
DIST = Door-to-door commuting distance in miles
STAD = Dummy variable equalling 1 when the individual lives in St Andrews and 0 otherwise
BFARE = Bus fare in pence

Appendix 6.3 Alternative nesting structures for the nested logit model

Variable	Alternative	Model 7 (NL)		Model 8 (NL)	
		Coeff.	t-stat.	Coeff.	t-stat.
Constant	Cycle	-1.931	-2.08	-2.371	-2.52
Constant	Bus	-2.685	-2.62	-2.452	-2.74
Constant	Car	-4.192	-4.11	-3.530	-3.98
Female	Cycle	-2.019	-5.28	-2.001	-4.65
Bus frequency – 1 or more per hour (ref)					
Bus frequency – less than 1 per hour	Bus	-1.110	-1.36	-1.609	-1.95
Number of cars in household		0.535	1.67	0.375	1.25
Log of walking time (minutes)	All	-1.649	-6.81	-1.658	-6.51
Log of cycling time (minutes)	Cycle	-1.755	-4.95	-1.539	-4.21
Log of in-vehicle time (minutes)	Bus, Car	-0.576	-1.59	-0.558	-1.84
Cost (pence)	All	-0.010	-2.17	-0.013	-3.16
IV parameter (t-stat w.r.t. 1)	Walk, Cycle	0.898	-0.53		
IV parameter (t-stat w.r.t. 1)	Bus, Car	0.787	-0.62		
IV parameter (t-stat w.r.t. 1)	Bus, Cycle			1.095	0.24
IV parameter (t-stat w.r.t. 1)	Walk, Car			0.542	-1.97
Observations		585		585	
Log-likelihood: constant only L(c)		-241.543		-241.543	
Log-likelihood: final value L(β)		-167.043		-165.250	
Rho-squared (with L(c))		0.308		0.316	
Rho-squared adjusted (with L(c))		0.299		0.307	

"Tree" diagrams for the nested logit models:



Appendix 6.4 Mixed logit and mixed nested logit models with normally distributed time coefficients

Variable	Alternative		Model 9 (ML) Coeff.	Model 9 (ML) t-stat.	Model 10 (M-NL) Coeff.	Model 10 (M-NL) t-stat.
Constant	Cycle	Mean	-1.932	-1.46	-2.271	-1.84
Constant	Bus	Mean	-3.505	-3.00	-3.271	-2.73
Constant	Car	Mean	-6.300	-3.95	-5.423	-3.64
Female	Cycle	Mean	-3.326	-4.07	-2.853	-3.48
Bus frequency – 1 or more per hour (ref)						
Bus frequency – less than 1 per hour	Bus	Mean	-1.395	-1.44	-1.329	-1.60
Number of cars in household			0.910	1.70	0.651	1.54
Log of walking time (minutes)	All	Mean	-2.788	-5.84	-2.450	-5.22
		Std. Dev.	0.006	0.23		
Log of cycling time (minutes)	Cycle	Mean	-3.740	-3.37	-2.964	-3.46
		Std. Dev.	1.233	4.06	1.095	3.05
Log of in-vehicle time (minutes)	Bus, Car	Mean	-1.341	-2.09	-0.920	-1.98
		Std. Dev.	0.863	0.93		
Cost (pence)	All	Mean	-0.012	-1.74	-0.013	-2.40
IV parameter (t-stat w.r.t. 1)	Walk, Car				0.876	0.41
Observations			585		585	
Log-likelihood: constant only L(c)			-241.543		-161.973	
Log-likelihood: final value L(β)			-161.619		-241.543	
Rho-squared (with L(c))			0.331		0.329	
Rho-squared adjusted (with L(c))			0.321		0.320	

Appendix 6.5 Mixed logit model with normally distributed cycle time coefficient and heteroscedastic error components

Variable	Alternative		Model 11 (ML) Coeff.	t-stat.
Constant	Cycle	Mean	-2.620	-1.68
Constant	Bus	Mean	-4.840	-2.10
Constant	Car	Mean	-7.334	-2.48
Female	Cycle	Mean	-3.769	-2.08
Bus frequency – 1 or more per hour (ref)				
Bus frequency – less than 1 per hour	Bus	Mean	-1.983	-1.55
Number of cars in household			0.876	1.26
Log of walking time (minutes)	All	Mean	-3.322	-2.68
Log of cycling time (minutes)	Cycle	Mean	-4.164	-2.49
		Std. Dev.	1.496	2.55
Log of in-vehicle time (minutes)	Bus, Car	Mean	-1.167	-1.67
Cost (pence)	All	Mean	-0.018	-1.73
Error components:				
Constant	Walk	Std. Dev.	0.685	0.31
Constant	Cycle	Std. Dev.	0.059	0.09
Constant	Bus	Std. Dev.	2.275	1.47
Constant	Car	Std. Dev.	0	
Observations			585	
Log-likelihood: constant only L(c)			-241.543	
Log-likelihood: final value L(β)			-160.728	
Rho-squared (with L(c))			0.335	
Rho-squared adjusted (with L(c))			0.324	

Chapter 7

Forecasting the Demand for an Employee Park and Ride Service

Encouraging employers to adopt travel plans is an important element of the UK Government's integrated transport strategy (DETR, 1998). The objective of a travel plan is to reduce the number of employees commuting alone by car to work and to encourage the use of more environmentally friendly modes such as public transport, cycling and walking. In recent years travel plans have become widely adopted in the UK, and have been proven to make a contribution to modal shift at the site level (Rye, 2002).

One of the measures that can be taken by the employer in order to reduce the number of commuters taking their car to the workplace is to introduce a Park and Ride service, i.e. a large off-site parking space with a shuttle-bus serving the workplace. This can be particularly effective in reducing car use if the workplace has poor public transport links and/ or limited parking space on-site. The University of St Andrews, which is the subject of the current paper, qualifies in having relatively poor public transport links for a majority of employees and partly in having insufficient parking space relative to car users on-site, particularly for those employees working in the centre of town. It was therefore decided by the University that the possibility of introducing a Park and Ride service should be investigated further.

Since the Park and Ride service is yet to be implemented there does not exist any revealed preference (RP) data that can be used for model estimation. A feasible alternative approach is to carry out a stated choice experiment. As pointed out in chapter 4, stated preference methods have become increasingly popular in transportation research over the past two decades due to their flexibility to introduce new alternatives and attributes and to incorporate a wider range of attribute levels than what is observed in the market. SP data can also overcome problems often encountered with RP data such as little variance and/ or multicollinearity in the independent variables and measurement errors. The use of SP data has, however, also been met with much scepticism because of the hypothetical nature of the data. The question is simply how reliable data elicited from a hypothetical choice situation are. It is argued by several practitioners that SP data seem to be reliable given that the experiment is well designed and clearly explained to the respondents (e.g. Louviere *et al.*, 2000). There is also a growing body of evidence of successful use of SP models in forecasting (Beaton *et al.*, 1998; Fowkes and Tweddle, 1999).

This chapter aims to forecast the share of car drivers that would switch to using Park and Ride given that such a service was provided. The structure of the chapter is as follows: section 7.1 describes the stated choice experiment, section 7.2

describes the data while section 7.3 and 7.4 presents the modelling and forecasting results respectively. Section 7.4 concludes.

7.1 The Stated Choice experiment

All members of University of St Andrews staff who drove a car to work on the day of the survey were asked to take part in the stated choice (SC) experiment. The commuters were asked whether they would choose to travel to work as usual or use Park and Ride if such a service was provided by the University (see appendix 7.1). The SC experiment contained two attributes: Park and Ride door-to-door travel time and cost, which both varied over three levels relative to the individuals' current commute. The experiment was deliberately kept as simple as possible, i.e. with a low number of attributes and levels, since studies have shown that people give the most reliable answers when assessing changes in only two or three factors simultaneously (Bradley, 1988). More complex choice tasks may lead people to use so-called lexicographic choice rules, where only one attribute is considered at the time (Johnson and Meyer, 1984). Also, given that the survey was distributed by mail, a simple survey was considered more likely to achieve a high response rate.

To increase the realism of the experiment the attributes of the Park and Ride option were based around the individuals' actual travel time and cost when going by car and parking on-site. As a consequence the design is not orthogonal, i.e there will be some collinearity in the independent variables (see chapter 4). It has been argued, however, that some (preferably low) degree of collinearity is acceptable if the realism of the experiment is enhanced (Fowkes and Wardman, 1988; Louviere, 1988). The full factorial design with two attributes varying over three levels provides 9 possible combinations of attribute levels ($3^2 = 9$). Nine choice scenarios were considered to be a manageable task for the respondents who were all presented with the full set of choices (table 7.1).

Table 7.1: The full SC design. The attributes are those of Park and Ride relative to the individual's current commute.

Question	Park & Ride	
	Cost (in pence)	Time (in minutes)
1	0	+5
2	0	+10
3	0	+15
4	-50	+5
5	-50	+10
6	-50	+15
7	-100	+5
8	-100	+10
9	-100	+15

The respondents were given three options: 1) Choose park on-site, 2) Choose park and ride and 3) Don't know (see appendix 7.1). The "Don't know" responses were left out when estimating the model.

7.2 Data and descriptive statistics

7.2.1 Data characterization

Questionnaires were distributed to all members of St Andrews University Staff via the internal mail. Of the 1661 questionnaires that were distributed 642 were returned, giving a response rate of 38.7%. All car drivers were asked to complete the stated choice experiment. This yielded 255 responses with complete information about the work trip and socio-demographic characteristics that were used for model estimation. Prior to the main survey a pilot survey was carried out with members of the department of Economics, where several flaws in the original questionnaire were detected and subsequently corrected.

Table 7.2 Description of variables and data characteristics.

Dummy Variables	**Sample Share**
Academic – High income	24%
Academic – Low income	14%
Non-Academic – High income	22%
Non-Academic – Low income	40%
Female	54%
Currently park in university parking	82%
Arrive at work later than 9am	15%
Work in a building with limited on-site parking	55%
Continous Variables	**Mean value**
Door-to-door commuting time in minutes	20.5
Walking time in minutes	2.7
Travel cost in pence (calculated as 15 pence pr mile)	163
Number of cars owned by household	1.7

The individuals in the sample were categorized as academics or non-academics and divided into high and low income groups on the basis of their occupation as described in Chapter 6. It is hypothesized that the low-income groups will be more willing to use the park and ride service as their opportunity cost of an increase in travel time may be lower. Furthermore, academics may be more aware of environmental issues than non-academics and hence more willing to switch to the "greener" mode.

It is possible that females are more dependent on the car than males since they are often responsible for tasks such as picking up children from school. The number of cars in a household may be a proxy for attitudes towards driving, in the sense that an individual living in a household with many cars may be less inclined to use other modes of transport compared to an individual who lives in a household with fewer cars.

A person who works in a building with limited parking space nearby is likely to be more willing to switch to Park and Ride than a person who works in a building with ample parking space. If he/ she arrives late to work this effect is expected to be stronger since finding a parking space will be even more difficult. It is expected that an individual who parks in a University car park is less likely to switch to Park and Ride, assuming that this is the individual's preferred parking option. Also, it is hypothesized that an increase in the travel time and cost of an alternative will lower the probability of this alternative being chosen. Finally, a marginal increase in walking time is likely to lead to a higher decrease in the probability compared to a marginal increase in the time spent travelling in the vehicle.

7.2.2 Lexicographic responses

It is interesting to ask how many of the car drivers that completed the survey are prepared to switch to Park and Ride. Table 7.3 below shows the percentage of car drivers that chose car in all the scenarios (41.2%), along with those who chose Park and Ride in all the scenarios (6.7%).

Table 7.3 Lexicographic responses

	Always choose car	**Always choose P&R**
Number	105	17
Percentage share	41.2%	6.7%

It should be recalled that lexicographic responses imply that the respondent simplifies the completion of the survey by concentrating on one design attribute only, ignoring the other information presented in the experiment. In the present setting those respondents who always chose car could therefore be accused of using the lexicographic choice rule 'choose the mode with the lowest travel time'. It is far from obvious, however, that this is the reason for why a substantial share of the respondents never chose Park and Ride as there are many other plausible explanations. It is likely that some drivers are captive to the car mode, for example those who need their car in their work and those who are responsible for picking children up at school and performing other tasks away from work during the working day. Others may simply have a strong dislike for public transportation. These individuals are not likely to choose Park and Ride given any reasonable combination of travel times and costs. Others may feel that the tradeoffs presented in the experiment are not sufficiently favourable for them to choose Park and Ride. It should be pointed out that the upper limit on the value of time which can be identified from the survey is 20 pence per minute, since a respondent who did not choose to go by Park and Ride in the scenario where the time difference is 5 minutes and the cost difference is 100 pence must have a value of time of 20 pence per minute or higher. This is almost four times as high as the estimated average value of in-vehicle time in Britian (see the discussion in chapter 6), however, and it is therefore not likely that this is a major reason for why the number of drivers who never chose Park and Ride is so high. Indeed, this demonstrates the reason why many authors (Calfee and Winston, 1998 and Hensher 2001a) have advocated that experiments designed to estimate the value of time should focus on one mode only, since this avoids confounding unobserved mode-specific attributes (flexibility, privacy) with the trade-off of times and costs. The aim of the analysis in the present chapter, however, is *not* value of time estimation, but forecasting the demand for a new transport mode. In this case it is not a problem that a high share of the respondents chose a particular mode, as long as this accurately represents what they would do if the new mode was introduced. Therefore, if the answer to the high number of individuals choosing to park on site is that the respondents have used lexicographic choice rules this is a problem since this behaviour is related to the experimental setting. This is not regarded as very likely, however, since the experiment only involves trade-offs between two attributes and should therefore be relatively easy to complete. On the other hand it is likely that the main explanation is that several car drivers feel captive to the car mode for various reasons. In terms of modelling this is not a problem, since this behaviour is consistent with behaviour in the 'real world'. In terms of transport policy, however, this is clearly an important impediment to making transport more sustainable.

7.3 Estimation results

Tables 7.4a-d below summarize the estimation results of various binary logit models. The simplest model (model 1) is linear in the time and cost attributes, and income and gender enter as explanatory variables. The Park and Ride constant is positive and significant. This variable represents the mean impact of all variables that influence the choice of mode that are not included in the model. The coefficient for the female dummy is negative as expected but not significant at the 5% level. It is interesting to note that when the model was re-estimated omitting the respondents that chose the same mode in all scenarios the coefficient was significant in the opposite direction. This indicates that when the females and males who find that going by car is the only option for them are omitted from the sample the remaining females are more likely to switch to Park and Ride than males.

Low-income academics are found to be significantly more likely to switch to Park and Ride than individuals in the other income categories. There are no significant differences between high-income academics and non-academics (with high and low income). As expected the likelihood of switching to Park and Ride decreases significantly when the number of cars in the household increases. The coefficients for travel time and cost are also strongly significant in the expected direction.

Table 7.4a Estimation results for the binary logit models.

	Model 1 (BL -linear)		Model 2 (BL -linear)	
Variable	Coeff.	t-stat.	Coeff.	t-stat.
Constant for Park and Ride	0.761	3.51	-0.157	-0.59
Female	-0.193	-1.68	-0.231	-1.91
Academic – High income	0.136	0.93	0.203	1.34
Academic – Low income	0.335	2.08	0.631	3.63
Non-Academic – High income	-0.032	-0.21	-0.036	-0.23
Number of cars in household	-0.187	-2.60	-0.144	-1.94
Limited on-site parking			0.658	5.29
Arrive at work later than 9am			-0.555	-1.94
Interaction (late*limited parking)			0.557	1.65
Park in University parking			0.626	4.12
Cost (pence)	-0.010	-7.17	-0.010	-7.24
Time (door-to-door, minutes)	-0.208	-14.65	-0.215	-14.81
Number of respondents in sample	255		255	
Number of responses	2105		2105	
Log-likelihood:				
Constant only L(c)	-1224.41		-1224.41	
Final value L(β)	-1065.08		-1036.46	
Rho-squared (with L(c))	0.130		0.154	
Rho-squared adjusted (with L(c))	0.127		0.149	

In model 2 the variables that relate to the individuals' current parking situation are also included. As expected the individuals who work in buildings with relatively poor on-site parking are significantly more likely to use Park and Ride than those who have good parking facilities nearby. The ones who arrive late at work *and* work in a building with poor on-site parking are even more likely to switch to park and ride as hypothesized. The ones who arrive late and work in a building with good on-site parking are the least likely to switch. Individuals who currently park in University parking are found to be significantly more likely to switch to Park and Ride. The explanation for this somewhat surprising result may be that University parking is not necessarily the employees' preferred parking option. The signs and significance of the variables already included in model 1 do not change markedly, apart from the Park and Ride constant which is no longer significant. The rho-bar squared increases from 0.127 in model 1 to 0.149 in model 2.

It is possible that people find travelling by car less onerous than travelling by shuttle bus. Using the Park and Ride will also entail some waiting time, which is usually regarded as more onerous than travelling in the vehicle. This is taken into account in model 3 by estimating a separate time coefficient for car and Park and Ride. Contrary to the prior expectations, however, the car mode has a slightly higher coefficient than that of Park and Ride. It is expected that people who currently have to park relatively far away from their workplace will be more likely to switch to Park and Ride. This is also accommodated in model 4 by separating the travel time for the car mode into walking time (from parking to workplace) and in-vehicle travel time. The coefficient for walking time is significant in the expected direction. The magnitude of the coefficient is slightly lower than the coefficient for in-vehicle time, however, which is again contrary to the prior expectations. The rho-bar squared increases very slightly from 0.149 in model 3 to 0.150 in model 4.

Table 7.4b Estimation results for the binary logit models.

	Model 3 (BL - linear)		Model 4 (BL – quadratic)	
Variable	Coeff.	t-stat.	Coeff.	t-stat.
Constant for Park and Ride	-0.231	-0.82	-1.625	-3.46
Female	-0.227	-1.87	-0.201	-1.62
Academic – High income	0.171	1.12	0.242	1.55
Academic – Low income	0.625	3.59	0.670	3.75
Non-Academic – High income	-0.056	-0.35	-0.073	-0.45
Number of cars in household	-0.078	-2.06	-0.086	-2.25
Limited on-site parking	0.677	5.33	0.479	3.54
Arrive at work later than 9am	-0.586	-2.05	-0.441	-1.49
Interaction (late*limited parking)	0.631	1.86	0.554	1.59
Park in University parking	0.626	4.09	0.647	4.08
Cost (pence)	-0.010	-7.25	-0.010	-7.29
Time (car) (minutes)	-0.223	-14.58	-0.320	-9.69
Time (P&R) (minutes)	-0.215	-14.82	-0.197	-4.80
Walking time (minutes)	-0.192	-7.49	-0.367	-5.68
Time (car) squared (minutes)			0.0016	2.38
Time (P&R) squared (minutes)			-0.0004	-0.63
Walking time squared (minutes)			0.0134	2.33
Number of respondents in sample	255		255	
Number of responses	2105		2105	
Log-likelihood:				
Constant only L(c)	-1224.41		-1224.41	
Final value L(β)	-1033.89		-1000.86	
Rho-squared (with L(c))	0.156		0.183	
Rho-squared adjusted (with L(c))	0.150		0.176	

Models 1 - 3 are linear in the alternative attributes, and therefore implicitly assume that the marginal disutility of an increase in travel time/ cost is constant. As discussed in chapters 2 and 6 it is possible that the marginal disutility of an increase in travel time/ cost is not constant, but a function of travel time/ cost. As shown in chapter 6 this hypothesis can be tested by entering quadratic terms in the representative utility function. Model 4 re-estimates model 3 including quadratic terms for the time attributes, which leads to a substantial increase in rho-bar squared from 0.150 to

0.176. It can be seen from table 7.4b that the coefficients for the quadratic terms are positive and significant, with the exception of Park and Ride time which has a negative but insignificant coefficient. This implies that the marginal disutility of an increase in car in-vehicle time and walking time is decreasing with travel time, which is not consistent with the utility maximising framework presented in chapter 2, but in line with the findings in chapter 6. As pointed out in that chapter, however, the quadratic specification with positive quadratic terms may lead to illogical results since the change in utility following an increase in travel time will eventually become positive as travel times increase. It is therefore necessary to re-specify the model by ensuring that an increase in travel time always leads to a decrease in the utility of a mode, while allowing for a decreasing marginal disutility of travel time. As in chapter 6 the square root and log transformations are adopted for this purpose. The estimation results for the square root and log models are presented in table 7.4c below.[1]

[1] For the sake of consistency it was chosen to enter the Park and Ride time variable in the log/ square-root form along with walking time and in-vehicle time for the car mode, although the models in which Park and Ride time entered linearly resulted in a slightly better fit.

Table 7.4c Estimation results for the binary logit models.

	Model 5 (log)		Model 6 (square-root)	
Variable	Coeff.	t-stat.	Coeff.	t-stat.
Constant for Park and Ride	5.208	8.17	1.267	3.27
Female	-0.167	-1.39	-0.188	-1.54
Academic – High income	0.217	1.42	0.206	1.33
Academic – Low income	0.637	3.67	0.596	3.38
Non-Academic – High income	-0.053	-0.34	-0.075	-0.47
Number of cars in household	-0.083	-2.20	-0.091	-2.36
Limited on-site parking	0.513	3.90	0.515	3.86
Arrive at work later than 9am	-0.456	-1.58	-0.558	-1.93
Interaction (late*limited parking)	0.626	1.85	0.709	2.07
Park in University parking	0.657	4.30	0.692	4.44
Cost (pence)	-0.010	-7.18	-0.010	-7.27
Log of time (car) (minutes)	-2.908	-12.57		
Log of time (P&R) (minutes)	-4.661	-13.41		
Log of walking time (minutes)	-0.475	-5.95		
Square-root of time (car) (minutes)			-1.931	-14.79
Square-root of time (P&R) (minutes)			-2.317	-15.34
Square root of walking time (minutes)			-0.837	-7.91
Number of respondents in sample	255		255	
Number of responses	2105		2105	
Log-likelihood:				
Constant only L(c)	-1224.410		-1224.41	
Final value L(β)	-1043.110		-1013.12	
Rho-squared (with L(c))	0.148		0.173	
Rho-squared adjusted (with L(c))	0.142		0.167	

It can be seen that in this case the square root transformation (model 6) yields the superior data fit, in contrast to the finding in chapter 6. Model 6 also has some other nice properties. Firstly, the time coefficient for P&R is higher than that of car, which is consistent with the expectation that people find travelling by shuttle bus more onerous than travelling by car. Secondly, while the coefficient for walking time is still lower than the coefficient for car in-vehicle time, the marginal disutility of an increase in walking time may still be higher since walking times are in general much lower than the time spent travelling in the vehicle and the marginal disutility of an increase in travel time is decreasing with travel time.

As pointed out in chapter 6 the Box-Cox transformation can also represent a decreasing marginal disutility of travel time given that $\lambda<1$, where the Box-Cox transformation of the time variable T is given by $T^{(\lambda)} = \ln T$ when $\lambda=0$ and $T^{(\lambda)} = (T^{\lambda} - 1)/\lambda$ otherwise (see e.g. Gaudry and Wills,1978 and Gaudry *et al.*, 1989). The benefit of the Box-Cox transformation is that λ is estimated along with the other parameters of the model, rather than assuming a particular transformation prior to model estimation. It can be seen, for instance, that when $\lambda=0.5$ the Box-Cox transformation is similar to the square root transformation, while when $\lambda=0$ it is equal to the log transformation. The drawback of the Box-Cox transformation, however, is that the standard error of the estimate of λ can only be obtained when T is strictly positive (see e.g. Greene, 2003b). In the present setting this means that some observations have to be eliminated because the walking time is reported to be zero. This results in a sample of 2003 observations which can be used for model estimation. The results for the binomial Box-Cox logit model estimated on this sample are reported in table 7.4d below (model 7). The model with square root transformations estimated on the same sample is included for the sake of comparison (model 8).

Table 7.4d Estimation results for the binary logit models (reduced sample).

	Model 7 (Box-Cox)		Model 8 (square-root)	
Variable	Coeff.	t-stat.	Coeff.	t-stat.
Constant for Park and Ride	1.404	2.89	1.493	3.76
Female	-0.213	-1.72	-0.213	-1.72
Academic – High income	0.238	1.51	0.245	1.55
Academic – Low income	0.599	3.38	0.605	3.41
Non-Academic – High income	-0.023	-0.14	-0.026	-0.16
Number of cars in household	-0.207	-2.62	-0.207	-2.62
Limited on-site parking	0.498	3.69	0.479	3.56
Arrive at work later than 9am	-0.610	-2.10	-0.590	-2.03
Interaction (late*limited parking)	0.573	1.65	0.581	1.68
Park in University parking	0.638	4.05	0.639	4.06
Cost (pence)	-0.010	-7.13	-0.010	-7.13
Time (car) (minutes)	-0.697	-4.11		
Time (P&R) (minutes)	-0.804	-3.57		
Walking time (minutes)	-0.380	-6.33		
Square-root of time (car) (minutes)			-1.890	-14.18
Square-root of time (P&R) (minutes)			-2.302	-14.99
Square root of walking time (minutes)			-0.839	-7.35
Lambda (t-statistic w.r.t. 1)	0.614	-4.54		
Number of responses	2003		2003	
Log-likelihood:				
Constant only L(c)	-1183.09		-1183.09	
Final value L(β)	-989.36		-990.27	
Rho-squared (with L(c))	0.164		0.163	
Rho-squared adjusted (with L(c))	0.157		0.157	

It can be seen from the table that λ is significantly different from 1, indicating that a non-linear specification is appropriate. The point estimate of λ is 0.6, which is similar to the square-root transformation and there is hardly any difference in the goodness of fit between the Box-Cox model and the model with square-root transformation. On the basis of this finding it was therefore concluded that the square-root transformation seems to be an accurate description of the non-linearities present in the representative utility function, and that therefore model 6 (which has the added benefit of being estimated on the full sample) will be used for forecasting purposes.

Model 6 is also used to test for learning and fatigue effects using the scaling method outlined in chapter 4. The null hypothesis of equal scale parameters cannot be rejected at the 5% significance level using the LR test (see appendix 7.2 for the estimation and test results). There is also no substantial difference in the coefficient estimates of the two models. This supports previous findings in the literature (Bradly and Daly, 1994; Sœlensminde, 2001), which conclude that strong fatigue effects are unlikely when offering no more than 10 choice comparisons within a single experiment.

As described in chapter 4, a drawback of the multinomial logit model is that the choices performed by the same individual are assumed to be independent. This is likely to lead to inflated *t*-statistics, and more seriously, biased parameter estimates if the random term is correlated with the explanatory variables. As described in chapter 3, the mixed logit model allows for correlation over choice tasks by adding an error term to the alternative utility functions that is constant over the choice tasks performed. The mixed logit estimation results are presented in tables 7.5a-c below.

Three mixed logit models were specified; one which can be interpreted as a heteroscedastic ML model, in which the coefficient for the P&R constant is normally ditributed (model 9), another in which the travel time coefficients are normally distributed (model 10) and another in which the travel time coefficients are log-normally distributed (model 11) to ensure that the coefficients for the travel time components are always negative. It should be pointed out, however, that the coefficient estimates in model 10 imply that all the individuals in the sample prefer lower in-vehicle travel times. The estimates imply that about 10% of the sample derive a positive utility from the time spent walking, but since there are arguments for why walking can be a good rather than a bad (exercise, fresh air etc.), this result does not seem completely implausible. It is important to note that the estimated coefficients in model 11 represent the mean (b) and variance (s) of the *log* of the coefficients, ($\ln\beta$). The median, mean and standard deviation of β is given by exp(b), $\exp(b+(s^2/2))$ and $\exp(b+(s^2/2))\times\sqrt{[\exp(s^2)-1]}$ respectively (see e.g. Train, 1997). Table 7.5c reports the median, mean and standard deviation of β derived from the estimates of b and s.

It can be seen from tables 7.5a-b that the standard deviations of the time coefficients in models 10 and 11 are highly significant, indicating that there are significant differences in the valuation of the different components of travel time in the sample. Since there is only a small difference in the rho-bar square between models 9 - 11, however, it seems that the substantial gain in goodness of fit compared to the binomial logit models is mainly a result of allowing for correlation across choice tasks performed by the same individual, rather than accommodating heterogeneity in tastes (the heteroscedastic model actually has a slightly higher rho-bar squared than the model with log-normally distributed time coefficients and the same rho-bar squared as the model with normally distributed time coefficients).

As expected the *t*-statistics in models 9-11 are substantially lower than those in model 6, with the exception of the coefficient for cost. The alternative attributes are still highly significant, while the socio-demographic characteristics were found to be insignificant with the exception of the coefficient for limited on-site parking and parking in University car parks. The remaining socio-demographic variables were therefore dropped from the models (the estimation results for model 10 including all socio-demographic characteristics are reported in appendix 7.3). The means of the coefficients in models 9-11 are substantially higher than those in model 6, which is consistent with the findings in Revelt and Train (1998) and Carlsson (2003). This result reflects that the error component of the mixed logit model is decomposed into two components, one which is specified by the researcher and one which is IID extreme value, and normalises the parameters on the basis of the second component (Revelt and Train, 1998).

Table 7.5a Estimation results for the Mixed Logit Models.

		Model 9		Model 10	
Variable		Coeff.	t-stat.	Coeff.	t-stat.
Constant for Park and Ride	Mean	2.855	1.20	4.311	2.25
	Std. dev.	7.199	9.31		
Limited on-site parking	Mean	2.567	2.39	2.979	3.79
Park in University parking	Mean	2.019	1.40	2.014	1.71
Cost (pence)	Mean	-0.038	-13.32	-0.039	-13.24
Square-root of time (car) (minutes)	Mean	-6.873	-9.08	-7.525	-9.35
	Std. dev.			0.301	4.00
Square-root of time (P&R) (minutes)	Mean	-8.281	-10.38	-9.307	-10.30
	Std. dev.			1.337	9.04
Square root of walking time (minutes)	Mean	-2.898	-3.59	-3.142	-5.96
	Std. dev.			2.439	7.89
Number of respondents in sample		255		255	
Number of responses		2105		2105	
Log-likelihood:					
Constant only L(c)		-1224.41		-1224.41	
Final value L(β)		-528.35		-527.81	
Rho-squared (with L(c))		0.569		0.569	
Rho-squared adjusted (with L(c))		0.567		0.567	

Table 7.5b Estimation results for the Mixed Logit Models.

Variable		Model 11 Coeff.	t-stat.
Constant for Park and Ride	Mean	4.662	2.36
Limited on-site parking	Mean	2.782	2.10
Park in University parking	Mean	1.968	1.92
Cost (pence)	Mean	-0.039	-13.26
Square-root of time (car) (minutes)	Mean of ln(coefficient)	1.965	18.88
	Std. dev. of ln(coefficient)	0.122	6.73
Square-root of time (P&R) (minutes)	Mean of ln(coefficient)	2.216	23.48
	Std. dev. of ln(coefficient)	0.154	9.50
Square root of walking time (minutes)	Mean of ln(coefficient)	1.181	5.68
	Std. dev. of ln(coefficient)	0.553	-7.39
Number of respondents in sample		255	
Number of responses		2105	
Log-likelihood:			
Constant only L(c)		-1224.41	
Final value L(β)		-530.29	
Rho-squared (with L(c))		0.567	
Rho-squared adjusted (with L(c))		0.565	

Table 7.5c Median, mean and standard deviation of the random coefficients in model 11.

	Median	Mean	St. Dev.
Square-root of time (car) (minutes)	7.134	7.188	0.883
Square-root of time (P&R) (minutes)	9.171	9.281	1.439
Square-root of walking time (minutes)	3.258	3.797	2.271

7.4 Forecasting results

For the reasons discussed in chapter 4 it may be necessary to rescale the estimated coefficients in the SC models before proceeding to forecast the modal split. An alternative forecasting method proposed by Fowkes and Preston (1991) is to average the probabilistic and the deterministic forecasts. The deterministic forecast is given by assuming that the mode with the higher representative utility is the chosen mode for all individuals in the sample. The random component of the model is thus ignored. The logic behind the Fowkes and Preston method is that the probabilistic forecast is likely to overpredict the minor mode while the deterministic forecast is likely to overpredict the major mode (Fowkes and Preston, 1991) (this holds when the error variance of the SC model is higher than that of the RP model). The correct forecast is therefore likely to be bounded by these forecasts. This hypothesis is supported empirically by Beaton *et al.* (1998). In the following the forecasts derived from the Fowkes and Preston method will be compared with the forecasts derived by the method of rescaling using a known RP coefficient.

As mentioned in chapter 4 the method of rescaling requires an RP estimate of one or more of the coefficients in the representative utility function.[2] A somewhat simplified version of the MNL and ML models in chapter 6 (without socio-demographic variables and time coefficients in square-root form to be consistent with the SP model) was chosen to compare the SP and RP estimates. The RP estimation results are given in table 7.6. All the alternative attributes are significant at the 5% level and have the expected sign, except the coefficient for in-vehicle time which has

[2] The joint RP/SP estimation approach outlined in chapter 4 would be feasible in the present study if the SP experiment included users of other existing modes such as bus. Since it was chosen to focus on the switching behaviour of car drivers, however, this approach cannot be adopted here.

the expected sign but is insignificant. In the ML model the coefficients for cycling time and walking time are specified to be normally distributed and have significant standard errors.[3]

Table 7.6 Estimation results for the RP Models.

		Model 12		Model 13	
Variable		Coeff.	t-stat.	Coeff.	t-stat.
Constant for walk	Mean	4.006	6.67	5.406	5.66
Constant for cycle	Mean	0.317	0.36	2.078	1.21
Constant for bus	Mean	0.204	0.43	0.704	1.06
Cost (pence)	Mean	-0.014	-3.48	-0.015	-2.81
Square-root of in-vehicle time (minutes)	Mean	-0.187	-1.00	-0.359	-1.36
Square-root of walking time (minutes)	Mean	-1.415	-7.69	-1.992	-5.44
	Std. dev.			0.405	2.41
Square-root of cycling time (minutes)	Mean	-0.927	-3.45	-1.990	-2.93
	Std. dev.			0.736	4.08
Number of respondents in sample		585		585	
Number of responses		585		585	
Log-likelihood:					
Constant only L(c)		-241.54		-241.54	
Final value L(β)		-196.94		-190.43	
Rho-squared (with L(c))		0.185		0.212	
Rho-squared adjusted (with L(c))		0.178		0.204	

It can be seen from table 7.6 that it would be necessary to rescale the coefficients in model 6 by a factor of 1.4 to reproduce the cost coefficient in model 12. As a consequence the forecasts derived from the SP MNL model without rescaling is likely to overpredict the share of Park and Ride users since rescaling by a factor higher than one implies that the error variance in the SP MNL model is higher than that of the RP MNL model (see chapter 4). Similarly, it can be seen that the

[3] It was also attempted to estimate a model with random coefficients for all the time variables, but this model did not converge.

coefficients in models 9 and 10 would have to be rescaled by a factor of 0.4 and 0.39 respectively to reproduce the cost coefficient in model 13. This implies that the variance of the SP ML models are lower than that of the RP ML model and that forecasts derived from the SP ML models without rescaling are likely to underpredict the share of Park and Ride users. It follows that the Fowkes and Preston hypothesis holds for MNL model but not for the ML models.

In order to produce the forecasts of the share of car drivers switching to Park and Ride it was necessary to estimate the Park and Ride travel time for all individuals in the sample. The travel times were calculated assuming that the Park and Ride site would be located at David Russell hall, which is just outside of the centre of town on the road to Strathkinness. The estimated travel time for each respondent depends on which area of town she works and her travel route into town (see appendix 7.4 for details). Needless to say the precision of the forecasts will depend on the accuracy of the estimated Park and Ride travel times.

The forecasts derived from models 6, 9 and 10 assuming that the cost of going by car and Park and Ride are the same are summarized in tables 7.7a-c below.

Table 7.7a Predictions of the modal shares derived from the MNL model assuming that travel costs are the same for the two modes.

Mode	Probabilistic- No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	81.5%	99.6%	87.9%	90.5%
Park & Ride	18.5%	0.4%	12.1%	9.5%

Table 7.7b Predictions of the modal shares derived from the ML model with normally distributed P&R constant assuming that travel costs are the same for the two modes.

Mode	Probabilistic- No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	81.7%	100%	78.9%	90.8%
Park & Ride	18.3%	0%	21.1%	9.2%

Table 7.7c Predictions of the modal shares derived from the ML with normally distributed time coefficients assuming that travel costs are the same for the two modes.

Mode	Probabilistic- No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	83.0%	100%	80.2%	91.5%
Park & Ride	17.0%	0%	19.8%	8.5%

It can be seen from tables that the MNL model without rescaling predicts that 18.5% of the car drivers will switch to Park and Ride using the probabilistic method while the deterministic forecast is that 0.4% will switch. The mean of these forecasts give the Fowkes and Preston prediction (9.5%). The forecast derived from the rescaled MNL model predicts that 12.1% of the car drivers will switch to Park and Ride. The heteroscedastic ML model (model 9) without rescaling predicts that 18.3% will switch to Park and Ride using the probabilistic method while the ML model with normally distributed time coefficients (model 10) predicts that 17% of the car drivers will switch. The deterministic forecast derived from both ML models is that none of the drivers will switch. The means of the probabilistic and deterministic forecasts imply that 9.2% and 8.5% of the drivers will switch to Park and Ride, while the forecasts derived from the rescaled models is that 21.1% and 19.8% of the drivers will switch using model 9 and 10 respectively. It should be noted that although the forecasts derived from the unscaled models are similar, the forecast from the rescaled ML models, which are perhaps the most reliable, are higher than that of the rescaled MNL model. This a consequence of the fact that the MNL model is rescaled by a factor higher than one, while the ML models are rescaled by a factor lower than one. It should also be noted that the forecasts derived from the ML models are consistently very similar.

Neither of the forecasts implies that a large percentage of car drivers will switch to Park and Ride, however. One of the measures that could be taken in order to encourage a larger take-up of the service is introducing on-site parking charges. In order for this strategy to be effective the charges would have to be coordinated with the local (Fife) Council so that car drivers do not merely switch from parking on-site to parking in the street.[4] The forecasts below are calculated assuming that the cost of parking on-site has increased by £1 following the introduction of parking charges.

Table 7.8a Predictions of the modal shares derived from the MNL model assuming that the cost of parking on-site is £1 higher than using Park and Ride.

Mode	Probabilistic- No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	63.4%	86.1%	67.3%	74.7%
Park & Ride	36.6%	13.9%	32.7%	25.3%

[4] The majority of parking in St Andrews has charges that are higher than the ones suggested here. There are, however, a small number of free parking spaces around town.

Table 7.8b Predictions of the modal shares derived from the ML model with normally distributed P&R constant assuming that travel costs are the same for the two modes.

Mode	Probabilistic-No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	66.3%	91.6%	64.6%	78.9%
Park & Ride	33.6%	8.4%	35.4%	21.1%

Table 7.8c Predictions of the modal shares derived from the ML with normally distributed time coefficients assuming that the cost of parking on-site is £1 higher than using Park and Ride.

Mode	Probabilistic-No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	68.4%	91.6%	66.5%	80.0%
Park & Ride	31.6%	8.4%	33.5%	20.0%

As expected the forecasts derived from all models using the various forecasting approaches suggest that the introduction of parking charges will increase the switching to Park and Ride. The MNL model without rescaling now predicts that 36.6% of the car drivers will switch using the probabilistic method and that 13.9% will switch using the deterministic method. The forecast derived from the rescaled model predicts that 32.7% of the car drivers will switch to Park and Ride, which is somewhat higher than the Fowkes and Preston forecast (25.3%). The heteroscedastic ML model without rescaling predicts that 33.6% will switch using the probabilistic forecast while the ML model with normally distributed time coefficients predicts that 31.6% will switch. Both models predict that that 8.4% will switch using the deterministic forecast. The rescaled models predict that 35.4% (model 9) and 33.5% (model 10) of the drivers will switch. The forecast from the MNL model without rescaling is somewhat higher than that of the ML models, suggesting that the MNL model overpredicts the impact of the parking charge on the demand for Park and Ride. The forecasts from the rescaled models are more similar, with the forecast from the rescaled ML models being slightly higher than their MNL counterpart. For the sake of completeness the forecasts derived from the ML model with log-normally distributed coefficients (model 11) are given in appendix 7.5a-b. These forecasts can be seen to be similar to those derived from the ML model with normally distributed time coefficients.

7.5 Conclusions

It can be seen from the previous analysis that the share of car drivers switching to Park and Ride will be relatively low unless supported by measures designed to make parking on-site less attractive such as introducing parking charges. This supports previous findings in the literature on travel plans (Rye, 2002) as well as the advice given in the UK government's travel plan guide (DETR, 1999) that a travel plan is most effective in reducing car use when it contains a combination of "sticks" and "carrots". In other words an effective travel plan should include measures aimed at discouraging car use as well as measures aimed at encouraging more environmentally friendly modes.

Parking charges seem to be justified as a means to deter driving as the current situation of providing free parking at the worksite actually subsidizes car use (Porter, 1999). Indeed Shoup (1997) finds that on average the cost of parking equals 75% of the variable cost of commuting by car. In this light the introduction of a parking charge is simply making the drivers pay a higher share of the variable cost of driving themselves.

An employee Park and Ride service seems to have the potential to be effective in reducing the demand for on-site parking when supported by measures to deter parking on-site. It is likely to be particularly effective at workplaces located in small towns (such as St Andrews) with poor public transport links and relatively limited parking facilities, although it could be considered at any workplace with little on-site parking or where the aim is to reduce the availability of on-site parking.

Appendix 7.1 Travel survey questionnaire - Stated Choice experiment

If you did not drive a car to work today you can skip this part.

PART 2

Park and Ride

Over the years a number of organisations have implemented Park and Ride facilities to reduce the need for their staff to bring their car to work. Please consider each of the nine hypothetical scenarios below and state whether you would use your current parking strategy or Park and Ride given that the University offered such a facility. The attributes are those of Park and Ride relative to your current commute.

Notes and assumptions:

- The travel time to your workplace will, of course, depend on the location of the Park and Ride. The variations in travel times in the hypothetical scenarios reflect this.

- Parking will be free for users of the Park and Ride facility and a parking space will be virtually guaranteed.

- The Park and Ride bus will be provided free of charge

SCENARIO 1	SCENARIO 2	SCENARIO 3
TIME OF TRIP (ONE WAY):	TIME OF TRIP (ONE WAY):	TIME OF TRIP (ONE WAY):
15 MINUTES LONGER THAN CURRENT COMMUTE	**10 MINUTES LONGER THAN CURRENT COMMUTE**	**5 MINUTES LONGER THAN CURRENT COMMUTE**
PRICE:	PRICE:	PRICE:
SAME AS CURRENT COMMUTE	**SAME AS CURRENT COMMUTE**	**SAME AS CURRENT COMMUTE**
GIVEN SCENARIO 1, I WOULD USE:	**GIVEN SCENARIO 2, I WOULD USE:**	**GIVEN SCENARIO 3, I WOULD USE:**
CURRENT PARKING SITE ☐	CURRENT PARKING SITE ☐	CURRENT PARKING SITE ☐
PARK & RIDE ☐	PARK & RIDE ☐	PARK & RIDE ☐
DON'T KNOW ☐	DON'T KNOW ☐	DON'T KNOW ☐

SCENARIO 4	SCENARIO 5	SCENARIO 6
TIME OF TRIP (ONE WAY):	TIME OF TRIP (ONE WAY):	TIME OF TRIP (ONE WAY):
15 MINUTES LONGER THAN CURRENT COMMUTE	**10 MINUTES LONGER THAN CURRENT COMMUTE**	**5 MINUTES LONGER THAN CURRENT COMMUTE**
PRICE:	PRICE:	PRICE:
50 PENCE CHEAPER THAN CURRENT COMMUTE	**50 PENCE CHEAPER THAN CURRENT COMMUTE**	**50 PENCE CHEAPER THAN CURRENT COMMUTE**
GIVEN SCENARIO 4, I WOULD USE:	**GIVEN SCENARIO 5, I WOULD USE:**	**GIVEN SCENARIO 6, I WOULD USE:**
CURRENT PARKING SITE ☐	CURRENT PARKING SITE ☐	CURRENT PARKING SITE ☐
PARK & RIDE ☐	PARK & RIDE ☐	PARK & RIDE ☐
DON'T KNOW ☐	DON'T KNOW ☐	DON'T KNOW ☐

SCENARIO 7	SCENARIO 8	SCENARIO 9
TIME OF TRIP (ONE WAY): **15 MINUTES LONGER THAN CURRENT COMMUTE** PRICE: **1 POUND CHEAPER THAN CURRENT COMMUTE**	TIME OF TRIP (ONE WAY): **10 MINUTES LONGER THAN CURRENT COMMUTE** PRICE: **1 POUND CHEAPER THAN CURRENT COMMUTE**	TIME OF TRIP (ONE WAY): **5 MINUTES LONGER THAN CURRENT COMMUTE** PRICE: **1 POUND CHEAPER THAN CURRENT COMMUTE**
GIVEN SCENARIO 7, I WOULD USE: CURRENT PARKING SITE ☐ PARK & RIDE ☐ DON'T KNOW ☐	**GIVEN SCENARIO 8, I WOULD USE:** CURRENT PARKING SITE ☐ PARK & RIDE ☐ DON'T KNOW ☐	**GIVEN SCENARIO 9, I WOULD USE:** CURRENT PARKING SITE ☐ PARK & RIDE ☐ DON'T KNOW ☐

Thank you for your cooperation!

Please return the questionnaire to:

Arne Hole
Department of Economics

Appendix 7.2 Estimation results for model 6 allowing for different scale parameters

	Model 6		Model 6*	
Variable	Coeff.	t-stat.	Coeff.	t-stat.
Constant for Park and Ride	1.267	3.27	1.399	2.45
Female	-0.188	-1.54	-0.288	-2.01
Academic – High income	0.206	1.33	0.176	1.04
Academic – Low income	0.596	3.38	0.719	3.50
Non-Academic – High income	-0.075	-0.47	-0.144	-0.75
Number of cars in household	-0.091	-2.36	-0.230	-2.19
Limited on-site parking	0.515	3.86	0.528	3.13
Arrive at work later than 9am	-0.558	-1.93	-0.704	-2.10
Interaction (late*limited parking)	0.709	2.07	0.863	2.17
Park in University parking	0.692	4.44	0.760	3.96
Cost (pence)	-0.010	-7.27	-0.011	-4.66
Square-root of time (car) (minutes)	-1.931	-14.79	-1.964	-7.62
Square-root of time (P&R) (minutes)	-2.317	-15.34	-2.362	-7.67
Square root of walking time (minutes)	-0.837	-7.91	-0.848	-5.58
Scale parameters				
(t-statistics w.r.t. 1)				
Choice 1 (base)			1.000	
Choice 2			1.064	0.39
Choice 3			0.691	-1.55
Choice 4			1.013	0.08
Choice 5			1.219	1.07
Choice 6			0.743	-1.02
Choice 7			0.955	-0.24
Choice 8			0.983	-0.07
Choice 9			0.678	-1.49
Number of respondents in sample	255		255	
Number of responses	2105		2105	
Log-likelihood:				
Final value L(β)	-1013.12		-1009.52	
LR – statistic	7.19		7.19	
Probability value	0.52		0.52	

Appendix 7.3 Estimation results for model 10 with all socio-demographic characteristics included

Variable		Model 10* Coeff.	t-stat.
Constant for Park and Ride	Mean	5.076	2.38
Female	Mean	0.725	0.78
Academic – High income	Mean	0.779	0.78
Academic – Low income	Mean	0.221	0.13
Non-Academic – High income	Mean	-0.954	-0.94
Number of cars in household	Mean	-0.839	-1.76
Limited on-site parking	Mean	2.625	3.02
Arrive at work later than 9am	Mean	-1.471	-0.81
Interaction (late*limited parking)	Mean	2.289	1.02
Park in University parking	Mean	1.905	2.04
Cost (pence)	Mean	-0.040	-12.95
Square-root of time (car) (minutes)	Mean	7.734	9.36
	Std. dev.	-0.306	-2.46
Square-root of time (P&R) (minutes)	Mean	9.292	10.18
	Std. dev.	1.340	8.25
Square root of walking time (minutes)	Mean	2.864	5.71
	Std. dev.	3.172	6.93

Number of respondents in sample	255
Number of responses	2105
Log-likelihood:	
Constant only L(c)	-1224.41
Final value L(β)	-526.40
Rho-squared (with L(c))	0.570
Rho-squared adjusted (with L(c))	0.567

Appendix 7.4 Estimated Park and Ride travel times relative to parking on-site (in minutes)

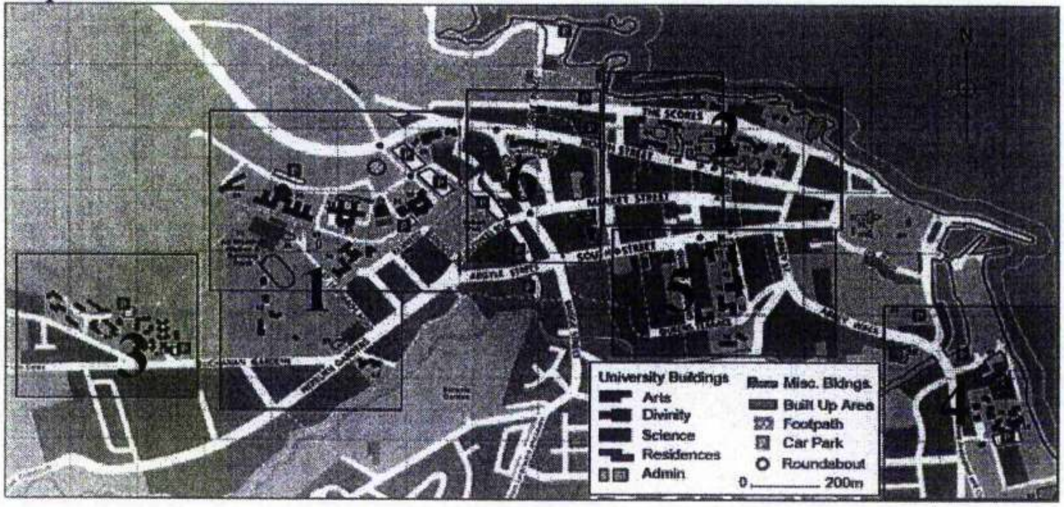
	Zone						
		1	2	3	4	5	6
	1	+4	+4	+0	+4	+4	+4
Route	2	+0	+0	+0	+0	+0	+0
	3	+3	+4	+0	+4	+4	+4
	4	+3	+5	+0	+8	+7	+5

Route and Zone definitions:

Route 1 – A91 (Dundee)
Route 2 – 8939 (Strathkinness)
Route 3 – A915 (Leven)
Route 4 – A917 (Crail)

Zone 1 – The North Haugh
Zone 2 – St Salvator's Quadrangle area (the Scores)
Zone 3 – David Russell Hall & Fife Park
Zone 4 – The Gatty Marine Lab area
Zone 5 – St Mary's College (South Street)
Zone 6 – McIntosh & Hamilton Halls & Student's Union

Map of zones:



Appendix 7.5a Predictions of the modal shares derived from the ML model with log-normally distributed time coefficients assuming that travel costs are the same for the two modes

Mode	Probabilistic-No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	81.5%	100%	79.1	90.7%
Park & Ride	18.5%	0%	20.9	9.3%

Appendix 7.5b Predictions of the modal shares derived from the ML with log-normally distributed time coefficients assuming that the cost of parking on-site is £1 higher than using Park and Ride

Mode	Probabilistic-No scaling	Deterministic	Probabilistic – Rescaled	Fowkes & Preston
Car	67.2%	91.6%	65.8%	79.4%
Park & Ride	32.8%	8.4%	34.2%	20.6%

Chapter 8

Can more compact cities contribute to reducing congestion on urban roads? Some evidence from Scotland

While chapters 6 and 7 focused on the analysis of the data from the survey of commuters in the University of St Andrews, this chapter will broaden the scope of the thesis somewhat by looking at mode choice among all Scottish commuters. In particular, the chapter investigates what impact making the cities in Scotland more densely populated would have on Scottish commuters' mode choice. Since making the cities more compact is also expected to have an impact on the propensity to link other activities to the commute, this will also be taken into account in the analysis.

8.1 Introduction

One of the often-cited benefits of the 'compact city' is that it offers the potential for developing an efficient public transport system (Burton, 2000).[1] There is much evidence that the relatively high level of public transport service quality typically provided in densely populated areas makes public transport a popular alternative to the private car among urban dwellers, and thus contributes to making urban transport more sustainable (Dieleman *et al.*, 2002).

It is well known in the transportation literature that commuters often link non-work activities to the work trip in order to reduce the time and cost spent travelling while fulfilling their travel needs. This role of the commuting trip was first emphasized by Oster (1977) and since his seminal contribution there have been many studies focusing on the determinants of individuals' propensity to form complex trip chains, both for commuting and other types of trips (Golob, 2000, 1986; Shiftan, 1998; Strathman *et al.*, 1994). It is widely accepted in the literature that since commuting trips usually takes place during peak hours, the tendency to link other activities to the commute exacerbates peak hour congestion.

Since the propensity to undertake complex trip chains is found to be higher among those living in areas with low accessibility to facilities (Williams, 1989), the trip-chaining literature is of relevance to the compact city debate. In particular, urban dwellers are expected to be less likely to link non-work activities to the commute than those who live outside the city and commute to the city to work, since the gain from trip chaining is lower for those living close to shops and other amenities. If this

[1] Although there are a number of definitions of the compact city, the focus of the current chapter (following Burton, 2000) is on population density, in the sense that cities with relatively high levels of population density are the more compact.

hypothesis is confirmed, the gain from the compact city will be twofold: 1) the increased use of public transport will contribute to making urban transport more sustainable and 2) the reduction in trip chaining propensity will alleviate peak hour congestion. This second potential benefit of the compact city has, as far as the author knows, been overlooked in the literature to date.

Following Bhat (1997) and Hensher and Reyes (2000) this chapter argues that trip-chaining and mode choice are inter-related decisions and should therefore be modelled jointly. Failure to do so may result in biased parameter estimates, and consequently erroneous forecasts. As in Hensher and Reyes (2000) the joint decision of making a particular type of trip chain and travelling with a certain mode is modelled using the mixed logit model described in chapter 3. The model is estimated using data from the 2001 Scottish Household Survey travel diary, which contains a representative sample of Scottish households' travel behaviour. The outline of the chapter is as follows: section 8.2 describes the data used for the analysis while section 8.3 and 8.4 presents the modelling and forecasting results respectively. Section 8.5 concludes.

8.2 Data and descriptive statistics

The Scottish Household Survey is a cross sectional survey commissioned by the Scottish Executive with the aim of providing representative information about the composition, characteristics and behaviour of Scottish households. A central part of the Survey is the Travel Diary, which is completed by a random adult (aged 16+) in each of the households surveyed. The 2001 Travel Diary contains 28519 trips made by 10163 individuals, of which 2954 reported to travel to work on the day of the survey. The estimation sample consists of 2472 work trip chains undertaken by 2368 individuals with complete information about the work trip and socio-demographic characteristics.[2] The observations are weighted to account for over-/ undersampling of certain socio-economic groups. The weights were employed at all stages of the analysis: deriving the descriptive statistics, estimating the models and calculating the commuters' response to the forecasting scenario described below. See Scottish Executive (2003) for detailed information about how the weights are derived.

The propensity to link non-work activities to the commute is somewhat lower in Scotland than that found in previous studies (see table 8.1), with 15.3% undertaking other activities between leaving for work and returning home as opposed to 20.4% in Strathman *et al.* (1994) and 21.5% in Golob (1986) who use data from Portland, USA and the Netherlands respectively. The most common trip chain type apart from the simple work – home – work chain is stopping at one or more non-work destinations on the way home from work (6.9%). This is followed by visiting a non-work destination in the middle of the day (3.7%) and stopping on the way to work (1.9%). A total of 2.8% made more complex trip chains, with combinations of stops going to and from work and/ or midday stops.

[2] A trip chain is defined as a sequence of trips originating and ending at the individuals' home. Therefore two work trips are recorded if the individual returned home in the course of the working day.

Table 8.1 Frequency of trip chains

Trip chain type	Share
Home – other – work – other – work – other – home	0.2%
Home – work – other – work – other – home	0.7%
Home – other – work – other – home	1.8%
Home – other – work – other – work – home	0.1%
Home – work – other – work – home	3.7%
Home – work – other – home	6.9%
Home – other – work – home	1.9%
Home – work – home	84.7%
Total complex	15.3%
Total simple	84.7%

The three main modes for commuting in Scotland are the private car, public transport (bus, train and underground) and walking, with car being the major mode (see table 8.2).[3] As expected the vast majority of the complex trip chains are made by car (81.3% of the complex trip chains are made by car as opposed to 72.2% of the simple trip chains), which reflects the greater flexibility of the car mode, both in terms of making stops and for carrying goods if the stop is made for shopping purposes. The share of public transport is roughly the same for simple and complex trip chains, while a relatively low share of the commuters who walk to work make a stop. This is likely to be related to the fact that those who walk live relatively close to the workplace, which reduces the potential benefit of linking non-work activities to the commute. We will investigate this hypothesis further in section 8.3.

[3] We consider the chosen mode to be the mode used for the longest trip in a trip chain. A relatively low share of the commuters travelled by alternative modes (bicycle and motorbike etc.). These observations were excluded from the sample for the modelling purposes.

Table 8.2 Modal split/ trip chaining by mode

	Share
Car	73.6%
Public transport	15.1%
Walk	11.3%
Simple trip chain	
Car	72.2%
Public transport	15.1%
Walk	12.7%
Complex trip chain	
Car	81.3%
Public transport	15.2%
Walk	3.5%

It can be seen from table 8.3 below that about 29% of the individuals in the (weighted) sample live in one of the four largest cities in Scotland (Glasgow, Edinburgh, Aberdeen and Dundee), while 38% work there. Of the 38% that work in one of the main cities about 13% live outside (in a different council area). As mentioned in the previous section it is well documented that urban form has an impact on the likelihood of commuting by public transport. Dieleman *et al.* (2002) find that commuters who live in one of the main Dutch cities are significantly more likely to travel by public transport to work, reflecting the comparatively high level of public transport service quality in the large urban areas. Similarly, Næss and Sandberg (1996) find that workplace location is a significant determinant of mode choice among individuals working in the Greater Oslo area, with commuters working in peripheral, low-density parts of the urban area being more car dependent that those working closer to the centre of town. In terms of linking non-work activities to the commute, Williams (1988) finds that accessibility to facilities is a key determinant in individuals' propensity to form complex trip chains. We therefore expect to find that those who live and work in one of the four main Scottish cities are more likely to commute by public transport as well as being less likely to form complex trip chains than those living and working elsewhere. In addition it is expected that those living close to their workplace will be less likely to form complex trip chains, since their gain from chaining non-work activities to the commute is lower. These individuals are also expected to be more likely to walk to work.

Table 8.3 Descriptive statistics

Discrete variables	Share	
Female	52%	
Single without children	11%	
Single with children	4%	
Couple without children	44%	
Couple with children	41%	
Work in an urban area – live outside	13%	
Live in an urban area – work outside	4%	
Live and work in an urban area	25%	
Continuous variables	**Mean**	**Std. Dev.**
Cars/ number of workers in household	0.72	0.44
Number of workers/ number of adults in household	0.86	0.22
Household income (in thousands)	23.42	11.64
Distance to work (in miles)	8.95	11.63
Work duration (hours)	7.53	2.87
Number of observations	2472	

The average household car ownership per worker ratio is relatively high in the sample (0.72), indicating that well over two thirds of the workers have access to a car for commuting purposes. This figure interestingly corresponds well to the share of commuters that travel by car to work. It is expected that the higher the car ownership per worker ratio the more likely the commuter is to go by car.

It is expected that household income will positively influence the likelihood of commuting by car, as well as the propensity to undertake complex trip chains. In a theoretical model of trip chaining behaviour, Adler and Ben-Akiva (1979) derive that high-income households are more likely to link other activities to the work trip since their time opportunity cost is higher. This result is confirmed empirically by Oster (1977) and Strathman *et al.* (1994). Furthermore, household composition has been found to be an important determinant of trip chaining. Clarke *et al.* (1981) find that young adults without children are more likely to link other activities to the work trip to satisfy their travel activity needs. Also, the ratio of workers per adult member of the household is also likely to be related to the propensity to form complex trip chains, as the higher the worker per adult ratio the tighter the time budget of the household. This hypothesis is confirmed by the findings in Oster (1977) and Strathman *et al.* (1994).

Unfortunately the data do not include information about the service characteristics (such as the time and cost) of the alternative modes available to the commuters. While it would be possible to estimate these characteristics using a similar approach to that described in chapter 6, the issue is further complicated here by the lack of data on public transport supply, which was included in the St Andrews survey. Given the size of the geographical area covered by the data it was considered too time consuming to obtain this data manually, using information on public transport service quality around the country. While the lack of level-of-service data is acknowledged as a weakness of the study, this approach is in line with other studies estimating mode choice models for large geographical areas (e.g. Dieleman *et al.*, 2002).

8.3 Estimation results

The estimation results for the multinomial logit mode choice/ trip chaining model are summarized in columns 3 and 4 in table 8.4. The marginal effects calculated at the (weighted) sample average of the independent variables are reported in appendix 8.1. The marginal effect of a dummy variable is calculated as the difference in the probabilities evaluated at the dummy set to zero and one respectively, evaluated at the sample average of the remaining variables in the model. The standard errors of the marginal effects are calculated using the Delta method (see e.g. Greene, 2003a). The sign and significance of the marginal effects are not found to be qualitatively different from the coefficients of the model, however, and therefore the latter will be focused on when interpreting the modelling results.

As expected the households' level of car ownership relative to the number of workers is found to be a significant determinant of the likelihood of going by car. Household income was not found to be significant determinant of the likelihood of going by car, which is somewhat unexpected but in line with the findings in chapter 6. Since car ownership is correlated with income, however, some of the influence of income on mode choice will be incorporated through the car ownership variable.

The individuals who live and/ or work in one of the four main cities are more likely to commute by public transport to work, in line with the findings in Dieleman *et al.* (2002) and Næss and Sandberg (1996). The ones who both live and work in a city are the most likely to commute by public transport. As expected the distance to work is found to be a highly significant determinant of whether the commuters walk to work.

In terms of trip chaining behaviour, there is not found to be a significant difference between men and women's propensity to undertake complex work trips. This is a somewhat unexpected finding, as conventional wisdom would suggest that females are more likely to undertake complex trips as they tend to be more responsible for household tasks such as shopping and picking up children from school/ kindergarten. It should be emphasized, however, that the present analysis includes stops for all purposes, including socialising and recreation. A more detailed analysis is needed to investigate whether females and males tend to link different kinds of activities to the commute. In general, however, there is no significant difference between the genders in terms of their propensity to link non-work activities to the commute.

Similar to Clarke (1981), it is found that single individuals with and without children are more likely to undertake complex work trips than households with two or more adults. This is likely to be related to the fact that single households have a tighter time budget, since shopping tasks etc. cannot be divided between several household members. Following the same logic it was expected that the workers divided by number of adults in the household variable would have a positive coefficient, indicating that individuals living in a household with 2 (or more) working adults are more likely to undertake complex trip chains that those living in households where one (or more) of the other adults in the household do not work. Surprisingly, this coefficient is found to be significant in the opposite direction.

Table 8.4 Multinomial and mixed logit mode choice models

		Model 1 (MNL)		Model 2 (ML)	
Variable	Alternative	Coeff.	t-stat.	Coeff.	t-stat.
Constant	Public transport	-0.809	-6.13	-0.810	-5.00
Constant	Walk	0.948	7.54	0.946	2.74
Cars/ number of workers	Car	2.751	18.27	2.750	14.30
Work in urban area – live outside	Public transport	1.548	8.72	1.548	7.44
Live in urban area – work outside	Public transport	1.206	4.03	1.206	3.20
Work and live in urban area	Public transport	1.800	13.10	1.800	10.96
Commuting distance	Walk	-0.303	-10.11	-0.302	-2.16
Constant	Complex	-0.062	-0.23	-0.073	-0.21
Female	Complex	-0.140	-1.19	-0.139	-0.97
Single without children	Complex	0.498	2.49	0.517	2.44
Single with children	Complex	0.534	1.96	0.546	1.97
Workers/ number of adults	Complex	-0.758	-2.72	-0.783	-2.18
Household income	Complex	0.026	5.26	0.026	3.92
Work duration	Complex	-0.275	-14.05	-0.282	-10.96
Commuting distance	Complex	0.010	2.06	0.010	1.91
Work in urban area – live outside	Complex	0.416	2.34	0.423	2.00
Live in urban area – work outside	Complex	1.034	3.83	1.060	3.04
Work and live in urban area	Complex	0.235	1.70	0.237	1.40
Error component	Car			0.011	0.71
Error component	Public Transport			0.008	0.41
Error component	Walk			0.022	1.06
Error component	Simple			0.000	
Error component	Complex			0.416	1.07
Observations		2472		2472	
Log-likelihood: constant only L(c)		-3197.28		-2568.47	
Log-likelihood: final value L(β)		-2568.65		-3197.28	
Rho-squared (with L(c))		0.197		0.197	
Rho-squared adjusted (with L(c))		0.195		0.195	

As in Strathman *et al.* (1994) it is found that individuals living in a high-income household are more likely to link non-work activities to the commute, reflecting the higher opportunity cost of time of these households. The most significant determinant of making a complex work trip is unsurprisingly found to be work duration; the longer the individual works the less likely she is to undertake a complex work trip. The coefficient for commuting distance is positive as expected and significant at the 5% level.

In terms of the location of the individuals' home and workplace it is found that those who work in one of the main cities and live outside are the most likely to link other activities to the work trip, along with the small proportion who live in the city and work outside. There is no significant difference between those who live and work in a city and those who live and work outside in terms of the propensity to undertake complex work trips. This finding confirms the prior expectation that commuters who both live and work in the city contribute less to peak hour congestion than those who live outside and commute into the city, which supports the hypothesis that a policy directed towards increasing the population density in urban areas will help alleviate road congestion. Section 8.4 simulates the effect of increasing the share of the commuters living and working in a city on trip chaining and modal split in Scotland.

In the multinomial logit models the errors are uncorrelated and thus the simultaneity of the mode choice/ trip chaining decision ignored. Bhat (1998) points out that the mixed logit model is well suited for multidimensional choice modelling since it accommodates correlation over both choice dimensions as opposed to the nested logit model (chapter 3) which only facilitates correlation across one choice dimension. He proposes a model which allows for correlation between alternatives that belong to each category in the two choice dimensions: in the present setting this means allowing for correlation between all alternatives that involve simple/ complex trip chains and all alternatives that involve going by car, public transport and walking. The variance-covariance matrix is thus given by:

Figure 8.1 Variance-covariance matrix for the Bhat – type multidimensional mixed logit model

$\sigma_1 + \sigma_4 + g$	.	.	.	.	.
σ_1	$\sigma_1 + \sigma_5 + g$	.	.	.	.
σ_4	0	$\sigma_2 + \sigma_4 + g$	.	.	.
0	σ_5	σ_2	$\sigma_2 + \sigma_5 + g$	.	.
σ_4	0	σ_4	0	$\sigma_3 + \sigma_4 + g$	.
0	σ_5	0	σ_5	σ_3	$\sigma_3 + \sigma_5 + g$

where g is the normalised extreme value variance, $\frac{\pi^2}{6}$. By applying the criteria for identification of the mixed logit model given in Walker (2002) and Walker *et al.* (2003) it can be seen that the Bhat model is not identified in this case, i.e. there are multiple combinations of the parameters in the utility functions/ variance-covariance matrix that maximise the log-likelihood.[4] The reason for this is straightforward: since the Jacobian of the variance-covariance matrix for the error differences of the mixed logit model (figure 8.2) has a rank of 5, the rank condition implies that only four of the parameters can be estimated (see section 3.3.5). One of the elements of the variance-covariance matrix must therefore be normalised (this can be seen directly from figure 8.2 by noting that σ_4 and σ_5 always appear together, which makes it impossible to determine their respective values). It was decided to impose the

[4] It can be shown that the Bhat model is identified when there are 3 or more alternatives per choice dimension. Since in the present case there are only two alternatives in the trip chaining dimension (simple and complex) the model is not identified.

restriction that σ_4 equals zero, although constraining σ_5 to equal zero is also a valid restriction in this case since both normalisations satisfy the equality condition because the estimated parameter is equal to ($\sigma_4 + \sigma_5$) regardless of which parameter is set to zero.

Figure 8.2 Variance-covariance matrix for the error differences of the Bhat – type multidimensional mixed logit model

$\sigma_4+\sigma_5+2g$	.	.	.	.
g	$\sigma_1+\sigma_2+2g$	.	.	.
$\sigma_4+\sigma_5+g$	$\sigma_1+\sigma_2+g$	$\sigma_1+\sigma_2+\sigma_4+\sigma_5+2g$	.	.
g	σ_1+g	σ_1+g	$\sigma_1+\sigma_3+2g$	.
$\sigma_4+\sigma_5+g$	σ_1+g	$\sigma_1+\sigma_4+\sigma_5+g$	$\sigma_1+\sigma_3+g$	$\sigma_1+\sigma_3+\sigma_4+\sigma_5+2g$

The estimation results for this model structure are reported in tables 8.4 and 8.5, columns 5 and 6.[5] It can be seen from the table that neither of the coefficients for the error components are found to be significant and that the 5% level, and that the rho-bar square of the MNL and ML models are identical. There is also very little difference between the models in terms of the sign and significance of the coefficients, although it can be seen that the *t*-statistics are in general lower in ML model. This indicates that the models will give similar predictions of changes in urban form on mode choice and trip chaining behaviour.

[5] The model was estimated using Kenneth Train's Gauss code with 125 Halton draws.

8.4 To what extent will more compact cities lead to a change in urban commuting conditions?

Given the findings in the previous section it is interesting to ask what impact a policy aimed at making cities more compact will have on modal split and trip chaining behaviour. Tables 8.5 and 8.6 below summarize the change in modal split and trip chaining propensity derived from the MNL and ML models respectively, given that a random 25% of the individuals who currently live outside and commute to the city to work moved into the city.[6] It can be seen that there is very little difference in the forecasts derived from the two models, reflecting the fact that the inclusion of error components in ML model resulted in only a very small change in the coefficient estimates of the model. As expected the share of commuters going by public transport is predicted to increase following this hypothetical change in demographics. The share of individuals walking to work is also predicted to increase, reflecting the fact that individuals who live and work in a town on average have shorter commuting distances than those who live outside and commute to the city to work (see the discussion below). In addition, it is found that the share of complex trips would fall. A policy directed at making the city more compact therefore has the potential to make commuting more sustainable by increasing the share of commuters who travel by public transport and walk as well as alleviating congestion by decreasing the propensity to link non-work activities to the work-trip.

[6] This is represented in the model through a change in the location dummies as well as a decrease in commuting distance for these individuals. The new commuting distance is calculated as the average commuting distance for workers who live and work in a city.

Table 8.5 Predicted percentage change in modal split/ trip chaining behaviour (MNL model)

	Change
Car	-0.340%
Public transport	0.679%
Walk	1.310%
Simple trip chain	0.163%
Complex trip chain	-0.905%

Table 8.6 Predicted percentage change in modal split/ trip chaining behaviour (ML model)

	Change
Car	-0.340%
Public transport	0.679%
Walk	1.310%
Simple trip chain	0.164%
Complex trip chain	-0.909%

It should be pointed out that the models do not predict a large change in the modal split and trip chaining behaviour following the hypothetical demographic change, indicating that a policy aiming at making the cities more compact should be coupled with other policies (such as a congestion charge) to bring about more substantial changes in commuting conditions. Also, although it is likely that a reduction in the propensity to form complex trip chains will help to alleviate congestion by reducing the number of trips undertaken during peak hours, the effect on energy use is uncertain. Since it is likely that the individuals will pursue the activities that used to be part of the commute at other (off-peak) times, a reduction in complex trip chains may increase the total distance travelled. This is mitigated by the fact that urban residents live closer to many amenities than non-urban residents, which reduces the distance travelled. When cross tabulating the average daily travel distance by place of residence and work a significant difference is found between the demand for travel across the different categories (see table 8.7). Individuals who live and work in a city are found to travel less than those living and working elsewhere, which suggests that the total effect of making the cities more compact would be a reduction in energy use.[7]

Table 8.7 Average daily travel distance by place of residence/ work.

Household category	Mean commuting distance (in miles)	Standard deviation
Live in a city – work in a city	12.9	22.3
Live outside – work in a city	41.4	34.1
Live in a city – work outside	36.7	27.2
Live outside – work outside	19.7	25.8

8.5 Conclusions

Joint models of commuters' mode choice and trip chaining behaviour have been estimated using a representative sample of Scottish commuters. It was found that urban residents are more likely to commute by public transport, as well as being less likely to form complex trip chains compared to those commuters who work in a city but live outside. Accordingly, a simulated change in demographics found that making the cities more compact would increase the share of commuters travelling by public transport and walking to work, as well as reducing the propensity to link non-work

[7] This finding is confirmed by Dieleman *et al.* (2002), who find that individuals living in the three largest Dutch cities travel less than those living elsewhere.

activities to the commute. As well as making urban transport more sustainable as a result of the increased use of public transport, this will contribute to lower levels of congestion in urban areas since the reduction in complex trip chains implies that fewer trips will be undertaken during peak hours.

Form a methodological point of view it is interesting to note that in the present case very little is gained from incorporating correlation between the different alternatives, as represented by the inclusion of error components in the mixed logit model. This finding cannot be generalised to other datasets, of course, but serves as a reminder that, due to its simplicity, the multinomial logit model is a natural starting point when estimating discrete choice models of travel demand, and in some cases provides an accurate representation of the substitution pattern observed from the data. This cannot be known *a priori*, however, and models that relax the IIA assumption (such as the nested or mixed logit models) should therefore be estimated and compared to the MNL model before a decision on the final model specification is made.

Appendix 8.1 Marginal effects from the multinomial logit mode choice/ trip chain model (calculated at the average of the independent variables).

Variable	Alternative	Marginal effect	t-stat.
Cars/ number of workers	Car	0.3080	15.28
Cars/ number of workers	Public transport	-0.2586	-14.66
Cars/ number of workers	Walk	-0.0495	-4.71
Work in urban area – live outside	Car	-0.1455	-8.76
Work in urban area – live outside	Public transport	0.1490	8.77
Work in urban area – live outside	Walk	-0.0034	-4.09
Live in urban area – work outside	Car	-0.1134	-3.91
Live in urban area – work outside	Public transport	0.1160	3.90
Live in urban area – work outside	Walk	-0.0027	-2.95
Work and live in urban area	Car	-0.1692	-13.18
Work and live in urban area	Public transport	0.1732	13.18
Work and live in urban area	Walk	-0.0040	-4.33
Commuting distance	Car	0.0055	8.42
Commuting distance	Public transport	0.0007	6.89
Commuting distance	Walk	-0.0061	-8.37
Constant	Complex	0.0069	-0.219
Female	Complex	-0.0154	-1.133
Single without children	Complex	0.0635	2.091
Single with children	Complex	0.0704	1.608
Workers/ number of adults	Complex	-0.0836	-2.607
Household income	Complex	0.0028	5.077
Work duration	Complex	-0.0303	-14.256
Commuting distance	Complex	0.0011	1.976
Work in urban area – live outside	Complex	0.0514	2.02
Live in urban area – work outside	Complex	0.1591	2.858
Work and live in urban area	Complex	0.0271	1.559

Chapter 9

Concluding remarks

This thesis contributes to the literature on the choice of transport mode for commuting trips, with special focus on the difference between urban and rural commuting in Scotland. Using data on mode choice for commuting trips from a survey of employees in the University of St Andrews it is found that the direct elasticities of the car mode are comparable to the estimates reported in studies of urban commuting, while the demand for public transport is found to be considerably more elastic. The values of in-vehicle travel time are lower than in most studies of urban commuting, reflecting that the roads in the St Andrews area are relatively uncongested. Further, introducing a park and ride service as an alternative to parking on-site is found to have a modest impact on the share of commuters parking on-site, unless the new service is accompanied by measures aimed at making parking on-site less attractive such as introducing parking charges.

The thesis also examines the impact of making cities more 'compact' on modal split and trip chaining behaviour. As well as making urban transport more sustainable as a result of an increase in the use of public transport, making cities more compact is found to contribute to lower levels of congestion in urban areas through a reduction in complex trip chains. The models do not predict a large change in the modal split and trip chaining behaviour following the hypothetical demographic change, however, which indicates that a policy aiming at making the cities more compact should be coupled with other policies to bring about more substantial changes in commuting conditions.

From a methodological point of view it is interesting to note that the simple multinomial logit (MNL) model with its restrictive independence from irrelevant alternatives property is found to perform surprisingly well in all the empirical applications presented in the thesis. The largest gain from using a model with a more complex error structure is found in chapter 7, where the mixed logit (ML) model leads to a substantial increase in data fit by overcoming the inability of the MNL model to account for the fact that the observations in the dataset are not independent. In the other chapters, however, the fit of the multinomial logit model is comparable to the more complex nested logit (NL) and ML models. More importantly it is found throughout the thesis that the models lead to similar predictions of changes in the modal split following the introduction of various car reduction policies. The value of time estimates are found to vary somewhat between the different models, however, with the ML model producing higher estimates than the MNL and NL models. Another interesting finding is that the marginal disutility of increasing the time spent travelling is found to be decreasing with travel time, implying that the value of time is a decreasing function of travel time. While this finding is also documented in other empirical studies it is not consistent with the standard utility maximising framework which is normally used as a basis for specifying mode choice models. Further, accounting for this particular functional form is found to substantially improve the fit of the models, highlighting that the specification of the functional form of the representative utility function is an important element of model specification along with accommodating flexible substitution patterns and controlling for unobserved heterogeneity through the introduction of more flexible error structures in the model.

To return to the policy implications that can be derived from the findings documented in the thesis it seems that, on balance, reducing the share of commuters travelling by car is a challenging, but not infeasible, task. On one hand, the findings seem to imply that policies aimed at improving the desirability of alternatives to the car are not likely to be successful unless coupled with measures aimed at deterring car use, such as parking or congestion charges. On the other hand pricing measures will not be effective in deterring car use unless a convenient alternative to driving is in place. This must be taken into account when designing policies to reduce the share of commuters going by car in less densely populated areas, where for many commuters no convenient alternative to the car currently exists.

It seems therefore, that the ideal policy should balance the use of 'sticks' and 'carrots' to bring about an increased use of 'green' modes for commuting trips. This finding supports the conclusions drawn in the UK Government's 1998 White Paper on transport (DETR, 1998). Unfortunately, however, the UK government's efforts have so far been focused on the carrots rather than the sticks. While increasing mobility, this policy is likely to have a negative impact on key policy parameters such as greenhouse gas emissions and congestion.

References

Abdel-Aty, M.A., R. Kitamura and P.P. Jovanis (1997) "Using Stated Preference Data for Studying the Effect of Advanced Traffic Information on Drivers' Route Choice", Transportation Research C 5(1), pp. 39-50.

Acutt, M.Z. and J.S. Dodgson (1996) "Cross Elasticities of Demand for Travel", Transport Policy 2(4), pp. 271-277.

Adler, T. and M. Ben-Akiva (1979) "A Theoretical and Empirical Model of Trip Chaining Behaviour", Transportation Research B 13(3), pp. 243-257.

Algers S., P. Bergstrom, M. Dahlberg and J.L. Dillén (1998) "Mixed Logit Estimation of the Value of Travel Time", Working Paper 15, Department of Economics, Uppsala University.

Alpizar, F. and F. Carlsson (2003) "Policy Implications and Analysis of the Determinants of Travel Mode Choice: An Application of Choice Experiments to Metropolitan Costa Rica", Environment and Development Economics 8(4), pp. 603-619.

Anas, A. and C. Chu (1984) "Discrete Choice Models and the Housing Price and Travel to Work Elasticities of Location Demand" Journal of Urban Economics 15(1), pp. 107-123.

Atherton, T. and M. Ben-Akiva (1976) "Transferability and Updating of Disaggregate Travel Demand Models" Transportation Research Record 610, pp. 12-19.

Armstrong, P., R. Rodrigo and J. de D. Ortúzar (2001) "Confidence Intervals to Bound the Value of Time, Transportation Research E 37 (2-3), pp. 143-161.

Bates J. (1997) "Measurement of Travel Time Values with a Discrete-Choice model: a Note", The Economic Journal 97, pp. 493-498.

Bates, J. (1988) "Econometric Issues in SP Analysis", Journal of Transport Economics and Policy 22(1), pp. 59-69.

Bates, J. J. and G. Terzis (1999) "Stated Preference and the Ecological Fallacy", in J. de D. Ortúzar (ed.) *Stated Preference Modelling Techniques*, PTRC Education and Research Services Limited, London.

Beaton, P., C. Chen and H. Meghdir (1998) "Stated Choice: A Study in Predictive Ability Using an Aggregate Truth Set", Transportation 25, pp. 55-75.

Becker, G. (1965) "A Theory of the Allocation of Time", The Economic Journal 75, pp. 493-517.

Begg, D. (2001) "Hit the Brakes", The Guardian, 6 March.

Ben-Akiva (1972) *The Structure of Travel Demand Models*, Ph.D thesis, Department of Civil Engineering, MIT.

Ben-Akiva M.E. and T. Morikawa (1990), "Estimation of Switching Models from Revealed Preferences and Stated Intentions", Transportation Research A 24(6), pp. 485-495.

Ben-Akiva, M. and D. Bolduc (1996) "Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure", Working Paper, Department of Civil Engineering, MIT.

Ben-Akiva, M. and S. Lerman (1979) "Disaggregate Travel and Mobility Choice Models and Measures of Accessibility" in D. Hensher and P. Stopher (eds.) *Behavioural Travel Modelling*, Croom Helm, London.

Ben-Akiva, M. and T. Atherton (1977) "Methodology for Short-Range Travel Demand Predictions: Analysis of Carpooling Incentives", Journal of Transport Economics and Policy 11(3), pp. 224-261.

Ben-Akiva, M.E. and B. Boccara (1995), "Discrete Choice Models with Latent Choice Sets", International Journal of Research in Marketing 12, pp. 9-24.

Ben-Akiva, M.E. and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge.

Bhat, C. (1995) "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice", Transportation Research B 29(6), pp.471-483.

Bhat, C. (1997) "Work Travel Mode Choice and Number of Non-work Commute Stops", Transportation Research B 31(1), pp. 41-54

Bhat, C. (1998) "Accommodating Flexible Substitution Patterns in Multi-dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice", Transportation Research B 32(7), pp. 455-466.

Börch-Supan, A. (1990) "On the Compatibility of Nested Logit Models with Utility Maximization", Journal of Econometrics 43, pp. 373-388.

Bradley M.A. and A. J. Daly (1994), "Use of the Logit Scaling Approach in Stated Preference Analysis", paper presented at the 6th World Conference on Transport Research, Lyon, July.

Bradley, M. (1988) "Realism and Adaptation in Designing Hypothetical Travel Choice Concepts", Journal of Transport Economics and Policy 22(1), pp. 121 -137

Bradley, M.A. and A.J. Daley (1997) "Estimation of Logit Choice Models Using Mixed Stated-Preference and Revealed-Preference Information", in P. Stopher and M. Lee-Gosselin (eds.) *Understanding Travel Behaviour in an Era of Change*, Pergamon, Oxford.

British Medical Association (1992) *Cycling: Towards Health and Safety*, Oxford University Press, Oxford.

Brownstone, D. (2000) "Discrete Choice Modelling for Transportation", Resource Paper, 9th International Association of Travel Behaviour Conference, Gold Coast, Australia, 1-7 July.

Brownstone, D. and K.A. Small (2003) "Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations", paper presented at the AEA meetings in Washington, DC, January.

Brownstone, D. and K.E. Train (1999) "Forecasting New Product Penetration with Flexible Substitution Patterns", Journal of Econometrics 89, pp. 109-129.

Bunch, D.S. (1991) "Estimability in the multinomial probit model", Transportation Research B 25(1), pp. 1-12.

Burton, E. (2000) "The Compact City: Just or Just Compact? A Preliminary Analysis", Urban Studies 37(11), pp. 1969-2001.

Calfee, J. and C. Winston (1998) "The Value of Automobile Travel Time: Implications for Congestion Policy", Journal of Public Economics 69, pp. 83 -102

Carlsson, F. (2003) "The Demand for Intercity Public Transport: The Case of Business Passengers", Applied Economics 35(1), pp. 41-50.

Cirillo, C., A. Daly and K. Lindveld (1999) "Eliminating Bias Due to the Repeated Measurements Problem in SP Data", In J. de D. Ortúzar (ed.), *Stated Preference Modelling Techniques*, PTRC Education and Research Services Limited, London.

Clarke, M.I., M.C. Dix, P.M. Jones and I.G. Heggie (1981) "Some recent developments in activity-travel analysis", Transportation Research Record 794, pp. 1-8.

Cleary, J. and H. McClintock (2000) "Evaluation of the Cycle Challenge Project: A Case Study of the Nottingham Cycle Friendly Employers' Project", Transport Policy 7(3), pp. 117-125.

Daganzo, C. (1979), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, Academic Press, New York.

Daly, A. (1987) "Estimating 'tree' logit models", Transportation Research B 21(4), pp.251-267.

Daly, A. and S. Zachary (1978) "Improved Multiple Choice Models", in D. Hensher and M. Dalvi (eds.) *Determinants of Travel Choice*, Saxon House, Sussex.

Dargay, J.M. and M. Hanly (1999) "Bus Fare Elasticities: Report to the Department of the Environment, Transport and the Regions", ESRC TSU, December.

Dargay, J.M. and M. Hanly (2002) "The Demand for Local Bus services in England", Journal of Transport Economics and Policy 36(1), pp. 73-91.

Dawes R. and B. Corrigan (1974) "Linear Models in Decision Making", Psychological Bulletin 81, pp. 95-106.

De Palma, A. and D. Rochat (2000) "Mode Choices for Trips to Work in Geneva: An Empirical Analysis", Journal of Transport Geography 8, pp. 43-51.

DeSerpa, A. (1971) "A Theory of the Economics of Time", The Economic Journal 81, pp. 828-846.

DETR (1998) *A New Deal for Transport: Better for Everyone*, Department of the Environment, Transport and the Regions, London.

DETR (1999) *A Travel Plan Resource Pack for Employers*, Department of the Environment, Transport and the Regions, London.

DfT (2004) *Transport Trends*, Department for Transport, London.

Dickinson, J.E., S. Kingham, S. Copsey and D.J.P. Hougie (2003) "Employer Travel Plans, Cycling and Gender: Will Travel Plan Measures Improve the Outlook for Cycling to Work in the UK?", Transportation Research D 8(1), pp. 53-67.

Dieleman, F.M., M. Dijst and G. Burghouwt (2002) "Urban Form and Travel Behaviour: Micro-level Household Attributes and Residential Context", Urban Studies 39(3), pp. 507-527.

DoT (1996) *National Cycling Strategy*, HMSO, Department of Transport, London.

Downs, A. (1962) "The Law of Peak-Hour Expressway Congestion", Traffic Quarterly 16, pp. 393.

Dunne, J.P. (1984) "Commuting Mode-Choice in New Towns: A Case Study of Livingston", Scottish Journal of Political Economy 31(1), pp. 60-71.

Evans, A. (1972) "On the Theory of the Valuation and Allocation of Time", Scottish Journal of Political Economy 19, pp. 1-17.

Feeney, B.P. (1989) "A Review of the Impact of Parking Policy Measures on Travel Demand", Transportation Planning and Technology 13, pp. 229-334.

Fowkes, A.S. and G. Tweddle (1999) "Validation of Stated Preference Forecasting: A Case Study Involving Anglo-Continental Freight", In J. de D. Ortúzar (ed.), *Stated Preference Modelling Techniques*, PTRC Education and Research Services Limited, London.

Fowkes, T. and J. Preston (1991) "Novel Approaches to Forecasting the Demand for New Local Rail Services", Transportation Research A 25(4), pp. 209-218.

Fowkes, T. and M. Wardman (1988) "Design of SP Travel Choice Experiments, with Special Reference to Taste Variations", Journal of Transport Economics and Policy 22(1), pp. 27-44.

Gaudry, M.J.I. and M.I. Wills (1978) "Estimating the Functional Form of Travel Demand Models", Transportation Research 12(4), pp. 257-289.

Gaudry, M.J.I., S.R. Jara-Díaz and J. de D. Ortúzar (1989) "Value of Time Sensitivity to Model Specification", Transportation Research B 23(2), pp. 151-158.

Garrido, R.A. and J. de D. Ortúzar (1993) "The Chilean Value of Time Study: Methodological Developments", In *Proceedings of the 21st PTRC Summer Annual Meeting*, University of Manchester Institute of Science and Technology, UK.

Gillen, D.W. (1977) "Estimation and Specification of the Effects of Parking Costs on Urban Transport Mode Choice", Journal of Urban Economics 4, pp. 186-199.

Gil-Moltó, M.J. and A.R. Hole (2004) "Tests for the Consistency of Three-level Nested Logit Models with Utility Maximization", Economics Letters 85(1), pp. 133-137.

Golob, T.F. (1986) "A Nonlinear Canonical Correlation Analysis of Weekly Trip Chaining Behaviour", Transportation Research A 20(5), pp. 385-399.

Golob, T.F. (2000) "A Simultaneous Model of Household Activity Participation and Trip Chain Generation", Transportation Research B 34(5), pp. 355-376.

Goodwin, P.B. (1992) "A Review of New Demand Elasticities with Special Reference to Short and Long Run Effects of Price Changes ", Journal of Transport Economics and Policy 26(2), pp. 155-169.

Goodwin, P. (1999) "Transformation of Transport Policy in Great Britain", Transportation Research A 33, pp. 655-669.

Greene W.H. (2003a) *Econometric Analysis*, Macmillan, New York.

Greene, W.H. (2003b) *Limdep Version 8.0 for Windows*, Econometric Software, New York.

Griffiths W.E., R.C. Hill and G.G. Judge (1993) *Learning and Practising Econometrics*, John Wiley & Sons, New York.

Hensher D.A. and M. Bradley (1993) "Using Stated Response Choice Data to Enrich Revealed Discrete Choice Models", Marketing Letters, 4(2), pp. 139-151.

Hensher, D. (1979) "Individual Mode Choice Modelling with Discrete Commodities: Theory and Application to the Tasman Bridge Reopening", Economic Record 55.

Hensher, D. (1997) "A Practical Approach to Identifying the Market for High-Speed Rail in the Sydney-Canberra Corridor", Transportation Research A 31(6), pp. 431-446.

Hensher, D. (2001a) "The Valuation of Commuter Travel Time Savings for Car Drivers: Evaluating Alternative Model Specifications", Transportation 28, pp. 101-118.

Hensher, D. (2001b) "Measurement of the Valuation of Travel Time Savings", Journal of Transport Economics and Policy 35(1), pp. 71-98.

Hensher, D. (2001c) "The Sensitivity of the Valuation of Travel Time Savings to the Specification of Unobserved Effects", Transportation Research E 37(2-3), pp. 129-142.

Hensher, D. and W. Greene (2003) "The Mixed Logit Model: The State of Practice", Transportation 30(2), pp. 133-176.

Hensher, D. and W. Greene (2002) "Specification and Estimation of the Nested Logit Model: Alternative Normalisations", Transportation Research B 36(1), pp. 1 -17.

Hensher, D., P. Barnard, N. Smith and F. Milthorpe (1989) "An Empirical Model of Household Automobile Holdings", Applied Economics 21(1), pp. 35-57.

Hensher, D.A. (1994) "Stated Preference Analysis of Travel Choices: The State of the Practice", Transportation 21, pp. 107-133.

Hensher, D.A. and A.J. Reyes (2000) "Trip Chaining as a Barrier to Use Public Transport", Transportation 27, pp. 341-361.

Herriges, J.A. and C.L. Kling (1996) "Testing the Consistency of Nested Logit Models with Utility Maximization", Economics Letters 50, pp. 33-39.

Hoyle B. and R. Knowles (Eds.) (1998) *Modern Transport Geography*, Wiley, Chichester.

Hunt, G.L. (2000) "Alternative Nested Logit Structures and the Special Case of Partial Degeneracy", Journal of Regional Science 40(1), pp. 89-113.

Jara-Díaz, S.R. and M. Farah (1987) "Transport Demand and Users' Benefits with Fixed Income: The Goods/Leisure Trade Off Revisited", Transportation Research B 21(2), pp. 165-170.

Jara-Díaz, S.R. and J. de D. Ortúzar (1989) "Introducing the Expenditure Rate in the Estimation of Mode Choice Models", Journal of Transport Economics and Policy 23(3), pp. 293-308.

Jara-Díaz, S.R. and J. Videla (1989) "Detection of income effect in mode choice: Theory and application", Transportation Research B 23(6), pp. 393-400.

Jara-Díaz, S.R. (1998) "Time and Income in Travel Choice: Towards a Microeconomic Activity Framework", in T. Garling, T. Laitia and K. Westin (eds.), *Theoretical Foundations of Travel Choice Modelling*, Pergamon Press, Oxford.

Jara-Díaz, S.R. and C.A. Guevara (2003) "Behind the Subjective Value of Travel Time Savings. The Perception of Work, Leisure and Travel from a Joint Mode Choice Activity Model", Journal of Transport Economics and Policy 37(1), pp. 29-46.

Johnson, E. J. and R.J. Meyer (1984) "Compensatory Choice Models of Noncompensatory Processes: The Effect of Varying the Context", Journal of Consumer Research 11(1), pp. 528-541.

Kingham, S., J. Dickinson and S. Copsey (2001) "Travelling to Work: Will People Move out of their Cars", Transport Policy 8(2), pp. 151-160.

Koppelman, F.S. and C.H. Wen (1997) "Alternative Nested Logit Models: Structure, Properties and Estimation", Transportation Research B 32(5), pp. 289-298.

Krinsky, I. and A.L. Robb (1986) "On Approximating the Statistical Properties of Elasticities", The Review of Economics and Statistics 68(4), pp. 715-719.

Lancaster K.J. (1966) "A New Approach to Consumer Theory", Journal of Political Economy, 14, pp. 132-157.

Louviere, J. J. (1988) "Conjoint Analysis Modelling of Stated Preferences", Journal of Transport Economics and Policy 22(1), pp. 93-119.

Louviere, J.J., D.A. Hensher and J.D. Swait (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press, Cambridge.

Luce, D. (1959) *Individual Choice Behaviour*, John Wiley and Sons, New York.

Mackie, P.J., S. Jara-Diaz and A.S. Fowkes (2001) "The Value of Travel Time Savings in Evaluation", Transportation Research E 37(2-3), pp. 91-106.

Marschak J. (1960) "Binary Choice Constraints on Random Utility Indicators", in K. Arrow (ed.), *Stanford Symposium on Mathematical Methods in the Social Sciences*, Stanford University Press, Stanford.

McFadden (1978) "The Theory and Practice of Disaggregate Demand Forecasting for Various Modes of Urban Transportation", in *Emerging Transportation planning methods*, Department of Transportation, Washington D.C.

McFadden (1981) "Econometric Models of Probabilistic Choice" in C.F. Manski and D.L. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, Massachusetts.

McFadden D.L & K.E. Train (2000) "Mixed MNL Models for Discrete Response", Journal of Applied Econometrics 15, pp. 447-470.

McFadden D.L. (2001) "Disaggregate Behavioral Travel Demand's RUM Side: A 30-Year Retrospective", in D. Hensher (ed.), *The Leading Edge of Travel Behaviour Research*, Pergamon Press, Oxford.

McFadden, D.L. (1973) "Conditional Logit Analysis of Qualitative Choice Behaviour", In P. Zarambeka (ed.), *Frontiers in Econometrics*, Academic Press, New York.

McFadden, D.L. (1974) "The Measurement of Urban Travel Demand", Journal of Public Economics 3, pp. 303-328.

McFadden, D.L. (1978a) "Modeling the Choice of Residential Location", in A. Karlqvist et. al. (eds.), *Spatial Interaction Theory and Residential Location*, 75-96, North-Holland, Amsterdam.

McFadden, D.L. (1978b) "Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments", in D. Hensher and P. Stopher (eds.), *Behavioral Travel Modeling*, pp. 279-318, Croom Helm, London.

McFadden, D.L. (1986) "The Choice Theory Approach to Market Research", Marketing Science 5, pp. 275-297.

McFadden, D.L. (1999) "Rationality for Economists?", Journal of Risk and Uncertainty 19(1-3), pp. 73-105.

McFadden, D.L. (2001) "Economic Choices", American Economic Review 91(3), pp. 351-378.

Mohring, H. and D. Anderson (1994) "Congestion Pricing for the Twin Cities Metropolitan Area", Department of Economics working paper, University of Minnesota.

Morikawa T. (1994), "Correcting State Dependence and Serial Correlation in the RP/SP Combined Estimation Method", Transportation 21, pp. 153-165.

Morikawa, T. (1989) *Incorporating Stated Preference Data in Travel Demand Analysis*. Ph.D. thesis, Department of Civil Engineering, MIT.

Munizaga, M.A., B.G. Heydecker and J. de D. Ortúzar (2000) "Representation of Heteroskedasticity in Discrete Choice Models", Transportation Research B 34(3), pp. 219-240.

Nerhagen, L. (2001) *Travel Demand and Value of Time – Towards an Understanding of Individuals' Choice Behaviour*, Ph.D. thesis, Department of Economics, Gothenburg University.

Nielsen, O.A. and Jovicic, G. (2003) "The AKTA Road Pricing Experiment in Copenhagen", paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August.

Nielsen, O.A. and Sørensen, M.V. (2004) "Sensitivity of Variable Definitions in SP Analyses - An Empirical Study of Car-users' Evaluation of Length, Cost and time", paper presented at the 10th World Conference on Transport Research, Istanbul, July.

Noland, R.B. and H. Kunreuther (1995) "Short-Run and Long-Run Policies for Increasing Bicycle Transportation for Daily Commuter Trips", Transport Policy 2(1), pp. 67-79.

Nutley, S. (1998) "Rural areas: The accessibility problem", in B. Hoyle and R. Knowles (eds.), *Modern Transport Geography*, pp. 185-215, Wiley, Chichester.

Næss, P. And S.L. Sandberg (1996) "Workplace location, Modal Split and EnergyUse for Commuting Trips", Urban Studies 33(3), pp. 557-580.

Ortúzar, J. de D., A. Iacobelli, C. Valeze (2000) "Estimating demand for a cycle-way network", Transportation Research Part A 34(5), pp. 353-373.

Ortúzar, J. de D. (1999) "Fundamentals of Stated Preference", In J. de D. Ortúzar (ed.), *Stated Preference Modelling Techniques*, PTRC Education and Research Services Limited, London.

Ortúzar, J. de D. and L.G Willumsen (2001) *Modelling Transport*, John Wiley and Sons Ltd.

Oster, C.V. (1979) "Second Role of the Work Trip – Visiting Non-work Destinations", Transportation Research Record 1203, pp. 79-82.

Oum, T.H., W.G. Waters II and J.S. Yong (1992) "Concepts of Price Elasticities of Transport Demand and Recent Empirical Estimates", Journal of Transport Economics and Policy 26(2), pp. 139-154.

Ouwersloot, H. and P. Rietveld (1996) "Stated Choice Experiments with Repeated Observations", Journal of Transport Economics and Policy 30(2), pp. 203-212.

Pendyala, R.M., L.P. Kostyniuk and K.G. Goulias (1995) "A Repeated Cross Sectional Evaluation of Car-Ownership", Transportation 22(2), pp. 165-184.

Peter, J.P. and J. Olson (1994) *Understanding Consumer Behaviour*. Irwin Inc, Burr Ridge, Illinois.

Porter, R.C. (1999) *Economics at the Wheel. The Costs of Cars and Drivers*, Academic Press, San Diego.

Revelt D. and K.E. Train (1998) "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level", Review of Economics and Statistics 80(4), pp. 647-657.

Ruud, P. (1996) "Approximation and Simulation of the Multinomial Probit Model: An Analysis of Covariance Matrix Estimation", Department of Economics working paper, University of California, Berkeley.

Rye, T. (2002) "Travel Plans: Do They Work", *Transport Policy* 9(4), pp. 287-298.

Scottish Executive (2003) *Scottish Household Survey User Guide*, Edinburgh.

Shiftan, Y. (1998) "Practical Approach to Model Trip Chaining", Transportation Research Record 1645, pp. 17-23.

Shoup, D. (1997) "The High Cost of Free Parking", Journal of Planning Education and Resarch 17(1), pp. 3-20.

Shoup, D.C and R. Willson (1990) "Parking Subsidies and Parking Choices: Assessing the Evidence", Transportation 17, pp. 141-157.

Simon, H.A. (1955) "A Behavioural Model of Rational Choice", Quarterly Journal of Economics 69, pp. 99-118.

Small, K. (1983) "The Incidence of Congestion Tolls on Urban Highways", Journal of Urban Economics 13, pp. 90-111.

Small, K. (1992a) *Urban Transportation Economics*, Routledge, New York.

Small, K. (1992b) "Using the Revenues from Congestion Pricing", Transportation 19, pp. 359-381.

Small, K. and H. Rosen (1981) "Applied Welfare Economics with Discrete Choice Models", Econometrica 49(1), pp. 105-130.

Smyth, A. (2003) "Devolution and Sustainable Transport", in I. Docherty and J. Shaw (eds.), *A New Deal for Transport*, pp. 30-50, Blackwell, Oxford.

Stradling, S.G, M.L. Meadows and S. Beatty (1999) *Factors Affecting Car Use Choices*, Report to the Department of the Environment, Transport and the Regions, Transport Research Institute, Napier University, Edinburgh.

Strathman, J.G., K.J. Dueker and J.S. Davis (1994) "Effects of Household Structure and Selected Travel Characteristics on Trip Chaining", Transportation 21, pp. 23-45.

Swait J. and M.E. Ben-Akiva (1987) "Empirical Test of a Constrained Discrete Choice Model: Mode-Choice in Sao Paulo, Brazil", Transportation Research B 21(2), pp. 103-115.

Swait, J. and J.J. Louviere (1993) "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models" Journal of Marketing Research 15, pp. 305-314.

Swait, J. (2001) "Choice Set Generation within the Generalized Extreme Value Family of Discrete Choice Models", Transportation Research B 35, pp. 643 -666.

Sydsaeter K. and P. J. Hammond (1995) *Mathematics for Economic Analysis*, Prentice-Hall, New Jersey.

Sælensminde, K. (2001) "Inconsistent Choices in Stated Choice Data. Use of the Logit Scaling Approach to Handle Resulting Variance Increases", *Transportation* 28, pp. 269-296.

Thobani, M. (1984) "A Nested Logit Model of Travel Mode to Work and Auto Ownership", Journal of Urban Economics 15(3), pp. 287-301.

Thurstone, L. (1924) "A Law of Comparative Judgement", Psychological Review 34, pp. 273-286.

Train K.E. (1980) "A Structured Logit Model of Auto Ownership and Mode Choice", Review of Economic Studies 47(2), pp. 357-370.

Train K.E. (1998) "Recreation Demand Models with Taste Differences Over People", Land Economics, 74(2), pp. 230-239.

Train K.E. (1999) "Halton Sequences for Mixed Logit", Working Paper, Department of Economics, University of California, Berkeley.

Train K.E. (2003) *Discrete Choice Models with Simulation*, Cambridge University Press, Cambridge.

Train, K. and D.L. McFadden (1978) "The Goods/ Leisure Trade-off and Disaggregate Work Trip Mode Choice Models", Transportation Research 19, pp. 349-353.

Truong P. and D. Hensher (1985) "Measurement of Travel Time Values and Opportunity Cost from a Discrete-Choice model", The Economic Journal 95, pp. 438-451.

Tversky, A. (1972) "Elimination by Aspects: A Theory of Choice", Psychological Review 79(4), pp. 281-299.

Tversky, A. and D. Kahnemann (1974) "Judgement Under Uncertainty: Heuristics and Biases", Science, 185, pp. 1124-1131.

Tversky, A. and D. Kahnemann (1981) "The Framing of Decisions and the Psychology of Choice", Science, 211, pp. 453-458.

University of St Andrews (2002) *Travel Plan*, assisted by Colin Buchanan and Partners.

Varian, H. (1992) *Microeconomic Analysis*, W.W. Norton, New York.

Verplanken, B., H. Aarts and A. van Knippenberg (1994) "Habit, information acquisition and the process of making travel mode choices", European Journal of Psychology 27, pp. 539-560.

Walker, J. (2002) "The Mixed Logit (or Logit Kernel) Model: Dispelling Misconceptions of Identification", Transportation Research Record 1805, pp. 86-98.

Walker, J., M. Ben-Akiva and D. Bolduc (2003) "Identification of the Logit Kernel (or Mixed Logit) Model", Working Paper, Massachusetts Institute of Technology.

Wardman, M. (1988) "Comparison of RP and SP Models of Travel Behaviour", Journal of Transport Economics and Policy 22(1), pp. 71-91.

Wardman, M. (1998) "The Value of Travel Time: A Review of British Evidence", Journal of Transport Economics and Policy 32(3), pp. 285-316.

Williams, H. (1977) "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", Environment and Planning A 9, pp. 285-344.

Williams, H. and J. de D. Ortúzar (1982) "Behavioural Theories of Dispersion and Misspecification of Travel Demand Models", Transportation Research B 16(3), pp. 167-219.

Williams, P.A. (1989) "A recursive model of intraurban trip-making", Environment and Planning A 10, pp. 535-546.

Willson, R. (1992) "Estimating the Travel and Parking Demand Effects of Employer Paid Parking", Regional Science and Urban Economics 22, pp. 133-145.