Veitch Lister Consulting Pty Ltd

Zenith Model Framework Papers - Version 3.0.1

Paper G – Mode Choice Model

May 2014

Page Intentionally Left Blank

Zenith Model Framework Papers – Version 3.0.1

Paper G – Mode Choice Model

Draft Report

Date	Revision	Prepared By	Checked By	Approved By	Description
04/04/2014	A	TV	JC	TV	Draft Report
02/05/2014	B	MP	JC	TV	Draft Report

Executive Summary

The Zenith Models are a family of four step transport models, developed by Veitch Lister Consulting (VLC) and implemented in the OmniTRANS software package for a range of Australian cities and regions. This document is one in a series of working papers that collectively describe the model structure and framework of the Zenith Model; in particular, this document describes the Mode Choice Model.

The aim of the Zenith Mode Choice Model is to predict the likelihood that a trip maker will choose to walk, cycle, go by car or take public transport given the attributes of each mode, and the attributes of the trip maker (i.e. car ownership).

The Zenith Mode Choice Model takes into account the generalised cost of each travel mode, including travel time, tolls, petrol costs, fares, transfer penalties, etc, as well as taking into account the car ownership of each household.

Contents

List of Figures

Page Intentionally Left Blank

1 Introduction

This Technical Note is one of a series of papers that collectively describe the Zenith Transport Model. Zenith is a four step transport model, implemented in the OmniTRANS software package for a range of Australian cities and regions.

This Technical Note details the Mode Choice Model implemented within Zenith. The aim of the Zenith Mode Choice Model is to predict the likelihood that a trip maker will choose to walk, cycle, go by car or take public transport given the attributes of each mode, and the attributes of the trip maker (i.e. car ownership). The Mode Choice Model forms an integral part of the overall Zenith Model as it directly determines the amount of travel that will utilise each mode of transport.

This document focuses on the methodology of the Mode Choice Model, and does not include parameter estimates or model validation for specific regions. Information about parameter estimates and model validation can be found in the region specific technical notes relating to the Mode Choice Model.

For further information, please contact our research and development team at zenith@veitchlister.com.au.

2 Mode Choice Methodology

2.1 Background

One of the key decisions made by every trip maker is what mode of travel to use for each trip. The answer to this question may depend on a wide range of factors, some of which can be included in a strategic travel model such as Zenith, and some which cannot.

Factors which *have* been included in the Zenith Model are:

- **The reason (purpose) for the trip**. For example, a primary school child travelling to school is more likely to be driven to school (by a parent) than an older child attending secondary school,
- **The structure of the transport network, and the ease with which each mode facilitates travel to different locations.** Trip makers will generally choose a mode which reaches their chosen destination with least *perceived* cost. The Zenith Model takes account of travel time, monetary costs such as tolls, petrol, and fares, and the inconvenience of interchanging between public transport services.
- **Access to a car.** Households with low car ownership are more likely to use public transport or active travel modes.

By taking these factors into consideration, the Zenith Model is able to test the sensitivity of mode choices to changes in these factors, and to measure the resulting impact on network patronage and performance.

2.2 Modal Alternatives

The Zenith Mode Choice Model typically includes the following modal alternatives:

- Walking
- Car
- Public transport – walk access
- Public transport – car access

Some Zenith Models (i.e. Victoria and ACT) also include cycling as an alternative. The Victorian Model also divides car access to public transport into separate alternatives for park and ride and kiss and ride.

In all cases, the car mode includes both car drivers and car passengers, and also includes those travelling by motorbike.

2.3 Model Structure

The Zenith Mode Choice Model takes the form of a Nested Logit Model.

An example of the nesting structure employed by the Zenith model for Home Based Work is presented below in Figure 2.1. Car and public transport are grouped into a nest called "*Mechanised Modes*", while walking and cycling are grouped into "*Non-Mechanised Modes*". While this structure is typical for Home Based Work, the tree structures used in the Zenith Model do vary across trip purposes.

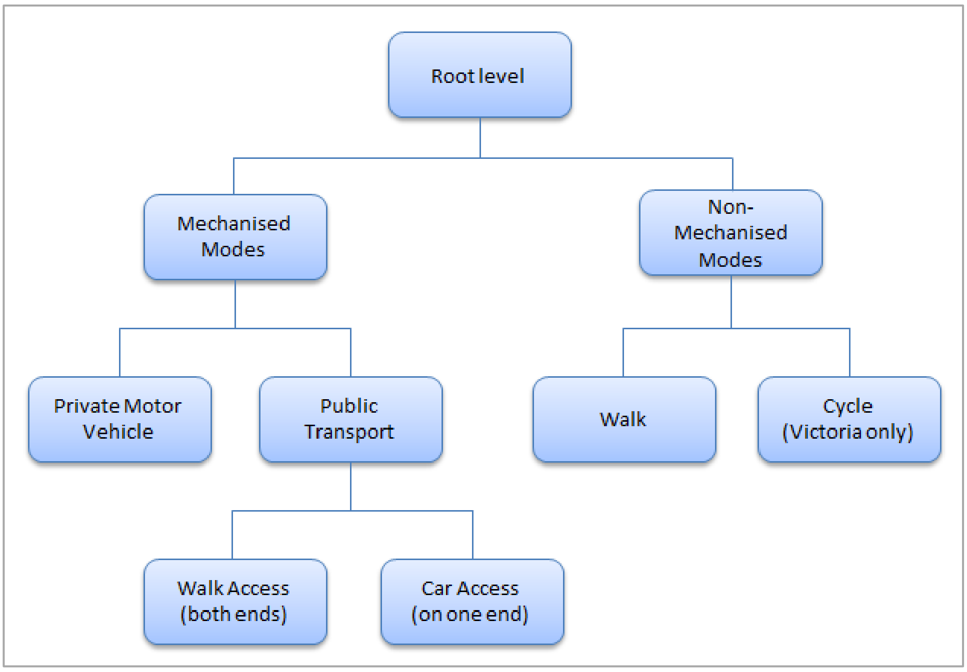

Figure 2.1: Nesting of Modal Alternatives

2.4 Utility Functions

The Zenith Mode Choice Model typically includes the following elements in the utility function for each alternative:

Walking

- Generalised cost (typically a function of time)

Cycling

- Generalised cost (typically a function of time)
- Alternative specific constant

Car

- Generalised cost (typically a function of travel time, fuel costs and tolls)
- Destination specific constants (typically used for the CBD and surrounds as a proxy for parking costs and supply constraints)
- Car ownership constants (separate constants for each car ownership level)
- Alternative specific constant

Public Transport

- Generalised cost (typically a function of walking time, car travel time, waiting time, in-vehicle time (train, bus, light rail, ferry, etc.), access and egress penalties, transfer penalties, fares, perceived cost of parking at train stations (as a function of spare parking capacity), perceived cost of crowding.
- Car ownership constants (separate constants for each car ownership level), which affect the utility for car access to transit

- Alternative specific constants which vary for each access / egress mode pair (i.e. walk access, park and ride, kiss and ride)

The definition of generalised cost for each alternative, and the estimated parameters which enter into the utility functions vary by region, and are contained in region specific technical notes.

2.5 Segmentation

In the Zenith Mode Choice Model, trips are segmented according to a range of trip purposes. Typically, separate mode choice models are applied for each of the following trip purposes:

- Home based work (white collar)
- Home based work (blue collar)
- Home based education (primary)
- Home based education (secondary)
- Home based education (tertiary)
- Home based shopping
- Home based recreation
- Home based other
- Work based travel (including work based work, work based shopping and work based other)
- Shopping and other (including shopping based shopping, shopping based other and other non-home based).

The reason for this approach is that modal preferences tend to vary by trip purpose. This is partly because the modal preferences of particular demographic groups tend to be correlated with trip purpose (i.e. children make education trips, workers make work trips, etc.).

2.6 Inclusion of Mode Choice in the Overall Zenith Model

The Zenith Model includes the choice of mode after the choice of destination and departure time.

The ordering of these three key choices (destination, departure time, and mode choice) is a matter of some conjecture in the industry, with different models adopting a different ordering.

The ordering is a matter of some importance, as it directly determines the relative sensitivity of each choice to changes in model inputs.

For example, take an individual who currently takes the train to work during the morning peak. If fares were increased by 50% during the peak, the challenge of the model is to determine whether the trip maker would:

- Continue to take the train during the peak,
- Continue to take the train but travel earlier or later to avoid the higher fare,
- Switch to ferry,
- Switch to bus,
- Switch to car,
- Switch to walking,
- Switch to cycling,

- Quit their job and find another job in a different location.

Of these alternatives, one might expect that a change in destination would be least likely, and therefore, the choice of destination should be included *first* in the ordering.

It could be argued that the same applies to Home Based Education trips; that the choice of education location is primarily related to preferences towards individual schools, with the choice of mode being a secondary and more flexible choice. But this no doubt varies within the population, depending on the importance placed by parents on the education of their children.

Given the key role that Home Based Work and Home Based Education trips play in public transport demand (which is a major focus of the Zenith Model), it is considered intuitive to place the destination choice before mode choice.

An important (and attractive) consequence of this approach is that the Zenith Mode Choice Model is applied separately for trips in each time period, taking account of the generalised costs of travel in the specific period within which each trip is made. Therefore, effects such as higher peak congestion, higher peak transit frequencies, and peak period public transport crowding can be taken into account when predicting mode choice.

In the Victorian and ACT models, the Mode Choice Model is more sophisticated and takes into account the generalised cost of full *return* journeys. Therefore, for a trip maker who travels to work in the morning peak, and returns home in the evening peak, the mode choice model will take into account the relative cost of each mode for the full return journey, reflecting the times at which the outward and return journeys are made. This is effectively a limited form of tour based modelling.

3 Theoretical Background

This section introduces the theoretical background to the Zenith Mode Choice Model, and is aimed at those readers who wish to understand the mechanics of the Nested Logit Model.

3.1 The Logit Model

In a transport modelling context, discrete choice models (such as logit) are employed to predict the travel choices made by individuals under a wide range of real and hypothetical circumstances.

In the case of mode choice (as it is applied in Zenith), individuals may choose from a set of available travel modes; typically car, walking, cycling or public transport, given a known origin, destination, trip purpose, departure time, etc. Each individual is assumed to derive a certain amount of *utility* from each alternative mode, where utility can be thought of as a measure of "*value*" or "*usefulness*". Travel cost is generally thought of as the opposite of utility.

Furthermore, it is assumed that each individual chooses the mode which provides them with the *greatest utility*. Therefore, individuals are assumed to be "*utility maximisers*".

The exact utility derived by an individual from a specific travel mode depends on a range of factors. Generally, these can be grouped as follows:

- Factors relating to the travel alternative (mode) (e.g. in the case of a public transport mode, in-vehicle travel time, waiting time, fare, walking time, number of transfers, reliability, etc.)
- Factors relating to the individual making the choice (e.g. car availability, income, occupation, fitness, as well as tastes and preferences held by the individual towards each mode, etc.)
- External factors (e.g. day of the week, the weather, strikes, etc.)

It is impossible for us to measure and account for all of these factors, particularly those relating to individual tastes and preferences. It would be impractical to ask each individual in a modelled area about their preferences towards certain modes, and even if this was done, the best that could be done is to improve the estimate of each individual's utility. The actual, exact, utility derived by each individual would still remain a mystery.

Given that the actual utility derived by an individual is unknowable, the best that can be done is to treat utility as a *probability distribution*. The probability distribution for utility is composed of two parts:

1. A modelled estimate of the utility, *V*, based on the known, observable factors relating to the travel mode, the individual, and externalities, and
2. A random error term, $\boldsymbol{\varepsilon}$, which captures the effect of all of the unknown and unobservable factors.

In mathematical terms,

$$\boldsymbol{U} = V + \boldsymbol{\varepsilon}$$

It is the random error term, ε, which gives utility its distribution (assuming no random parameters in the function for V).

Obviously, it is the aim of modellers to make the estimate of utility (component 1) as accurate as possible, by including as many of the key factors which affect an individual's utility as possible into the model. Doing this reduces the number of unknown factors, consequently reducing the spread of the random error (component 2), giving a narrower probability distribution. In essence, this allows the modeller to make an accurate and confident prediction about people's choices.

In the example shown in Figure 3.1 below, the probability distributions for car and public transport utility are represented by solid blue and red lines, respectively. The probability distribution for car sits to the right of public transport, which means that the car mode is likely to have a higher utility. Given that people maximise their utility, a majority of people will choose car, in this example.

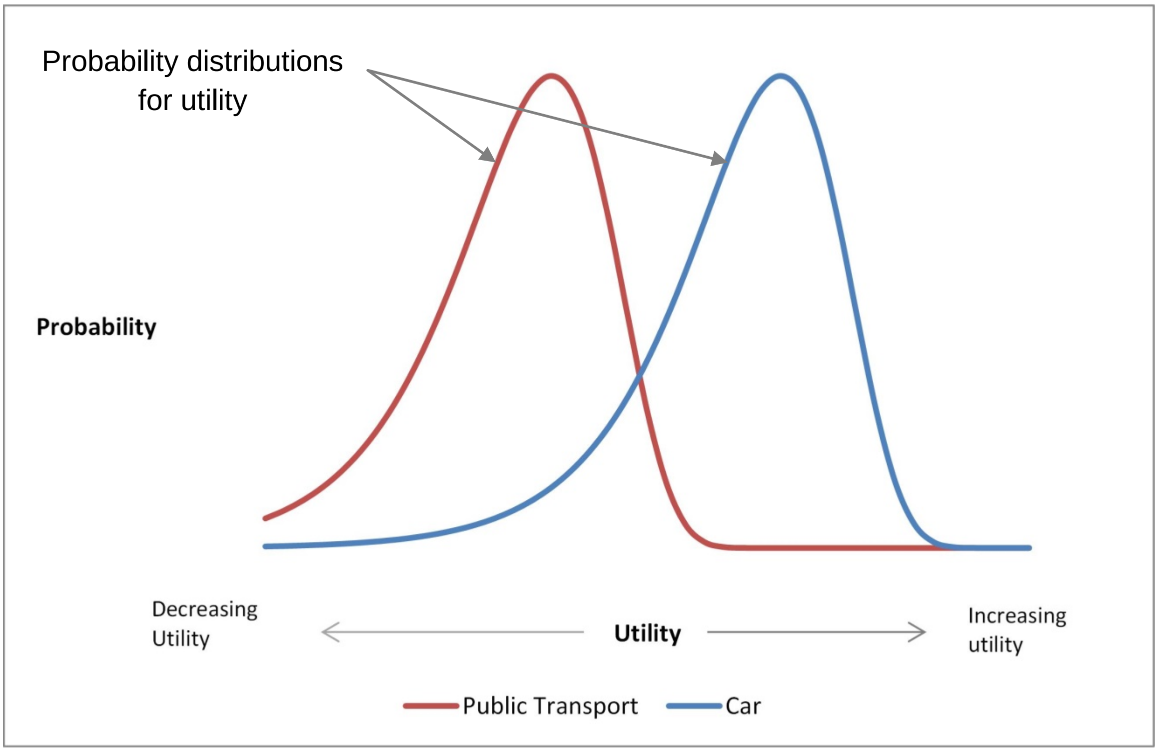

Figure 3.1: Example Probability Distributions of Car and Public Transport Utility

It is important to understand that for an actual individual, there is no probability distribution. An individual derives a specific utility from both car and public transport, and chooses the mode which offers the highest utility. The problem is that it is impossible to determine exactly what these utilities are. The probability distribution, therefore, is merely a probabilistic *estimate* of the end user's utility.

To see how this applies to an individual, consider Figure 3.2 below. The figure shows the actual utility derived by a hypothetical individual from car and public transport alternatives. The individual's utility for car is higher than their utility for public transport – as such, this individual will choose to travel by car.

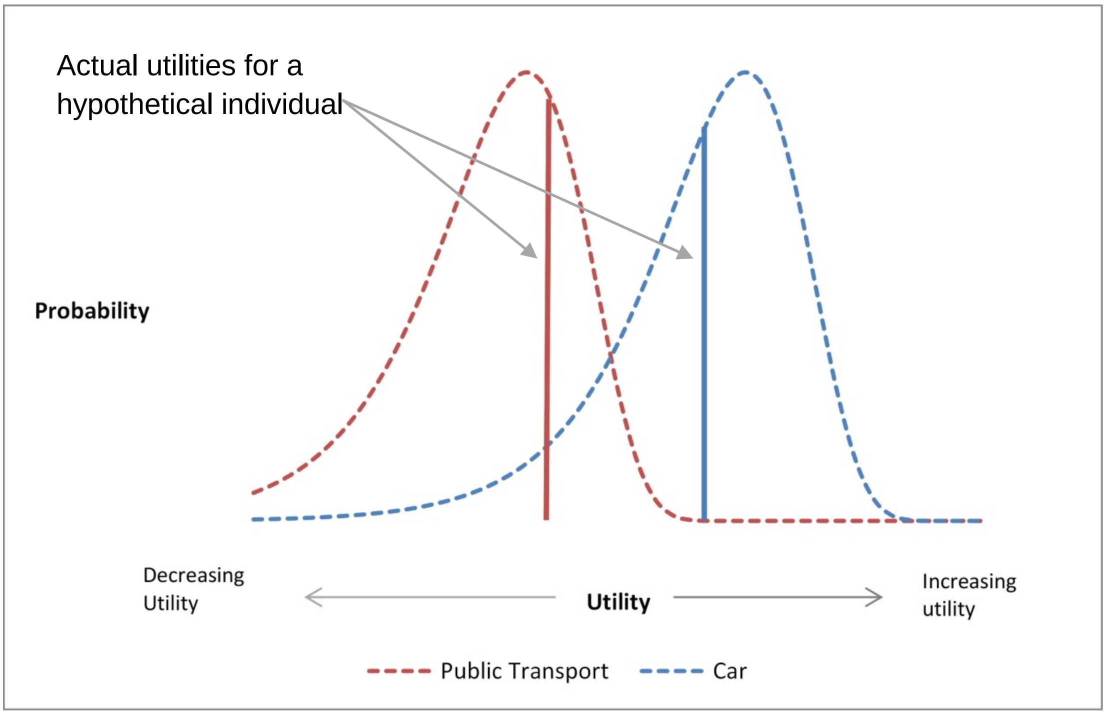

Figure 3.2: Actual Utilities for a Hypothetical Person

In contrast, Figure 3.3 presents another hypothetical individual; this time a person who chooses public transport. This individual sits far to the left on the car probability distribution – perhaps they don't have access to a car because their partner took the car to work, and so their only car option is to take a taxi. Whatever the reason, their public transport utility (the solid red line) is to the right of their car utility, so they choose public transport.

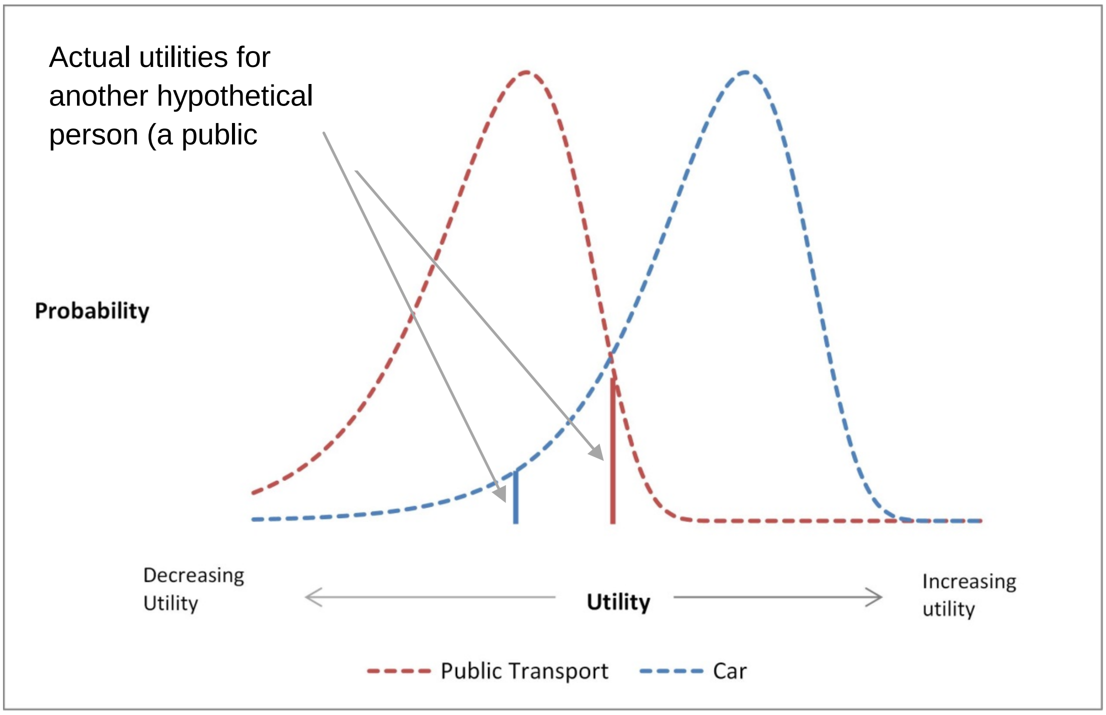

Figure 3.3: Actual Utilities for Another Hypothetical Person (who chooses Public Transport)

This person is a rarity, however. By definition, people are most likely to fall near the centre of the two probability distributions. This is illustrated in Figure 3.4 below, which illustrates that the majority of people have a utility near the centre of our probability distributions. In this example, most people have a car utility greater than their public transport utility (95% of people, in fact).

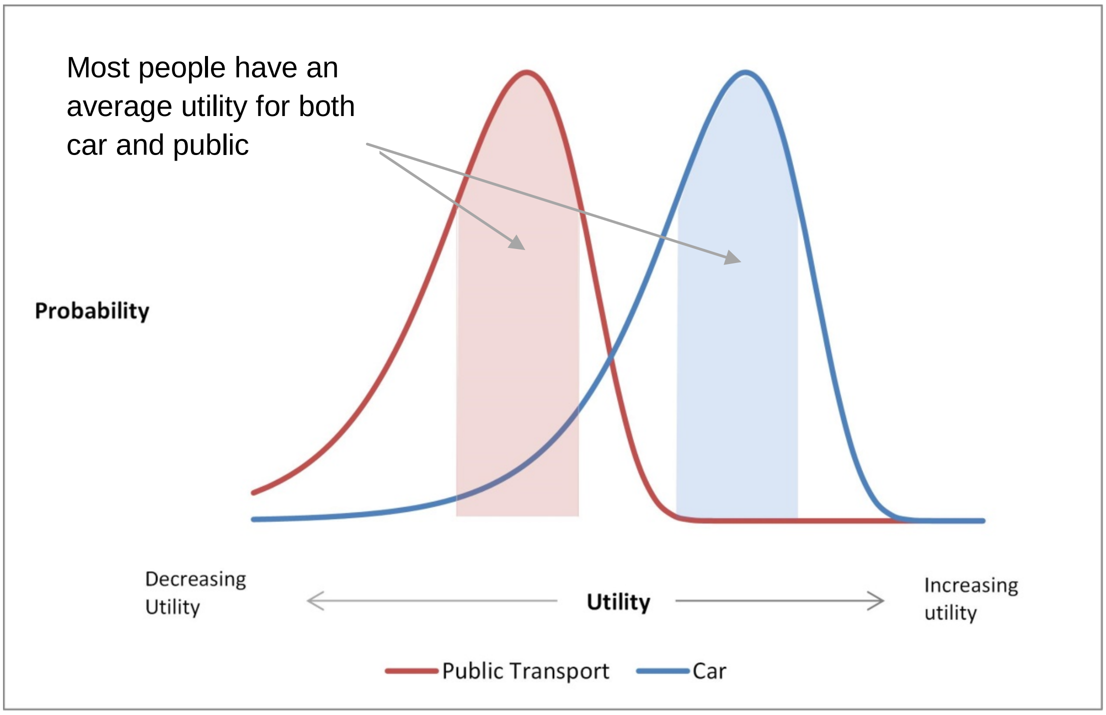

Figure 3.4: People Commonly Reside at the Centre of the Distributions

Within this framework, scenarios (such as changes in travel time, policy variables, etc.) can be viewed as shifts in the utility for one or more modes. For example, in Figure 3.5 below, a leftward shift is depicted in the utility for car travel. A decrease in utility is equivalent to an increase in cost, so this scenario might represent an increase in fuel costs, or parking costs, etc. Because the cost applies to all car travellers, the entire probability distribution moves left. In the example, the first hypothetical individual (the car user), has now switched to public transport; note that the car utility for our individual has shifted to the left of their public transport utility.

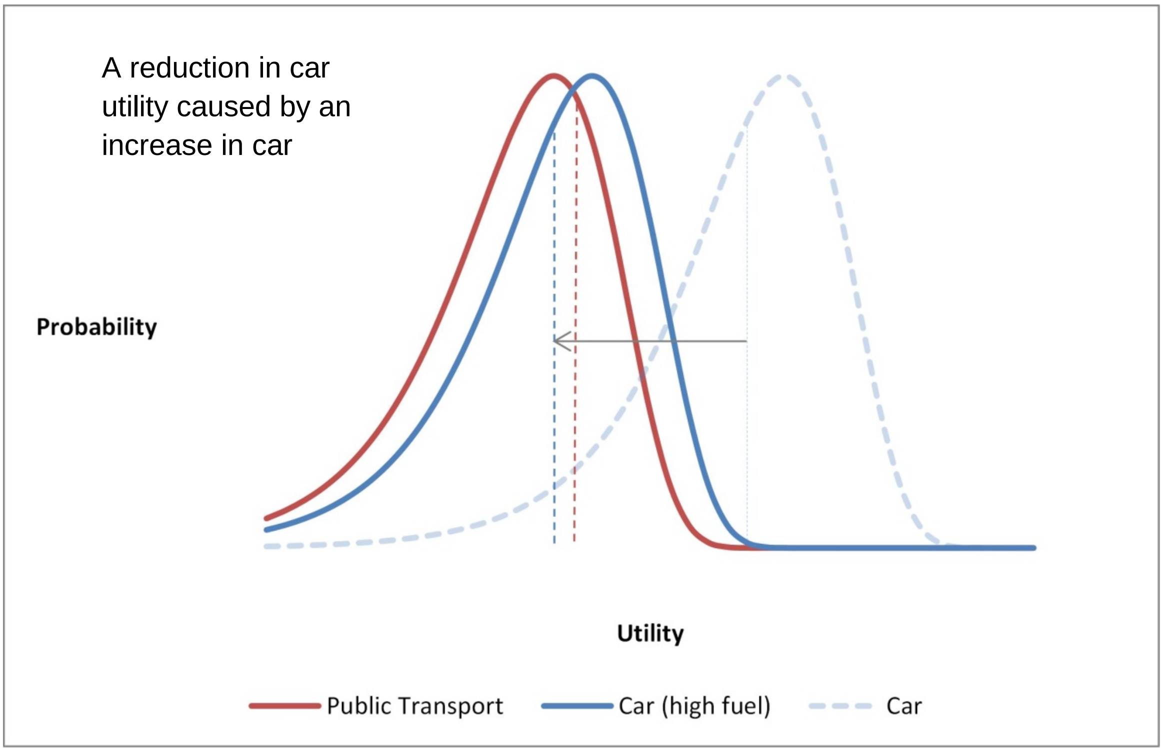

Figure 3.5: A Shift in the Utility for Car

It should be evident that by moving the distribution of car utilities left or right, a change in the proportion of individuals who choose car and public transport will occur. The proportion is often illustrated on a graph where the x-axis is the **difference** in the modelled estimates of utility (V) between two alternatives, as in Figure 3.6 below.

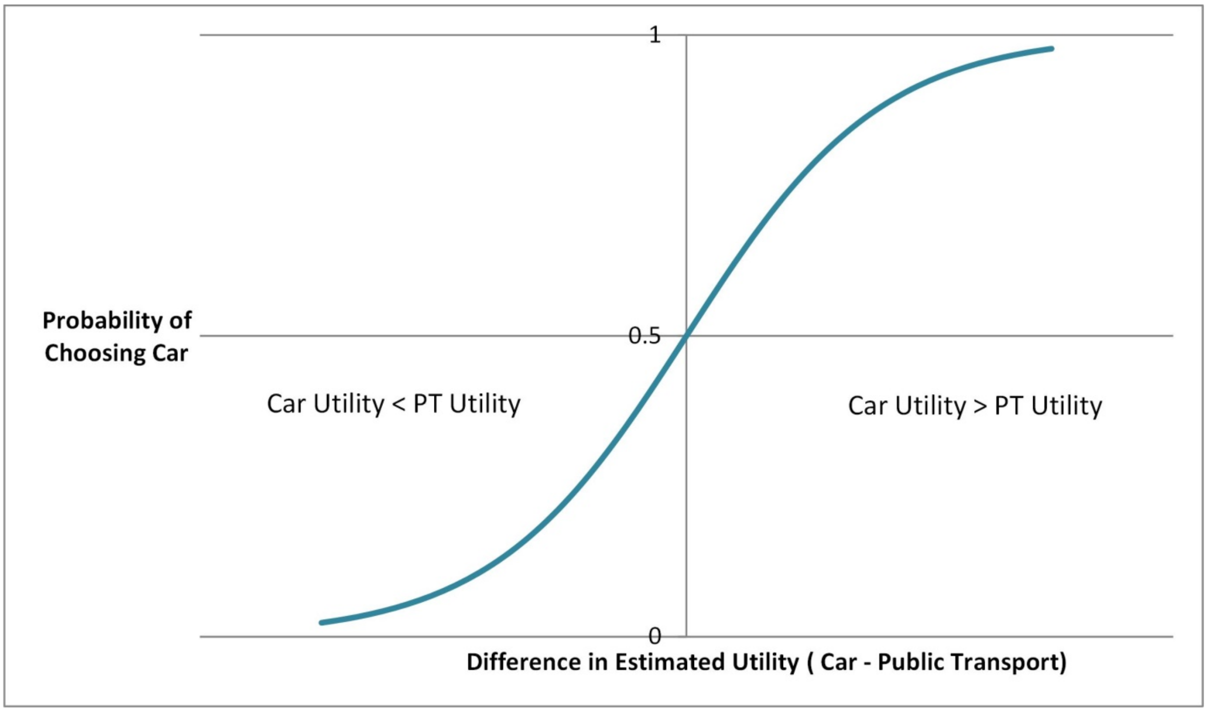

Figure 3.6: Probability of Choosing Car Given Differences in Modelled Estimates of Utility

<u>Estimating V</u>

At the beginning of section 0, the probability distribution for utility is comprised of two parts: $V,$ which is the modelled estimate of utility (and which determines the centre of the probability distribution), and the random element $\boldsymbol{\varepsilon}$ which gives the probability distribution its distribution.

For the model to be accurate and useful, the estimated utility, V, should depend on variables such as travel time, fuel price, fares, etc. Then, by changing these variables, a prediction can be made as to its impact on mode share.

For example, for car travel, it might be assumed that the key factors affecting utility are travel time, cost of fuel, cost of tolls, cost of parking and car availability. In this case, the estimate of utility, *V*, would be defined as follows:

$$V = \beta_1 TT + \beta_2 Fuel + \beta_3 Toll + \beta_4 Parking + \beta_5 Car\_Availability$$

where the βs are parameters which capture the effect of each variable on our estimate of utility; and TT, $FUEL$, $Toll$, $Parking$ and $Car\_Availability$ are the x's, the known attributes of the car trip.

If a factor does not affect the utility of a car trip (for example, if we were to include an irrelevant factor such as the first digit of the licence plate), then the parameter for that factor (the β) should be zero.

By varying the travel time, toll, parking and car availability, we change our estimate of $V,$ and thus change our predicted mode share. The βs determine how sensitive V is to changes in each variable.

<u>Estimating $\boldsymbol{\varepsilon}$</u>

The random error term, $\boldsymbol{\varepsilon}$, plays an extremely crucial role in the prediction of mode shares.

Consider the examples in Figure 3.7 and Figure 3.8 below.

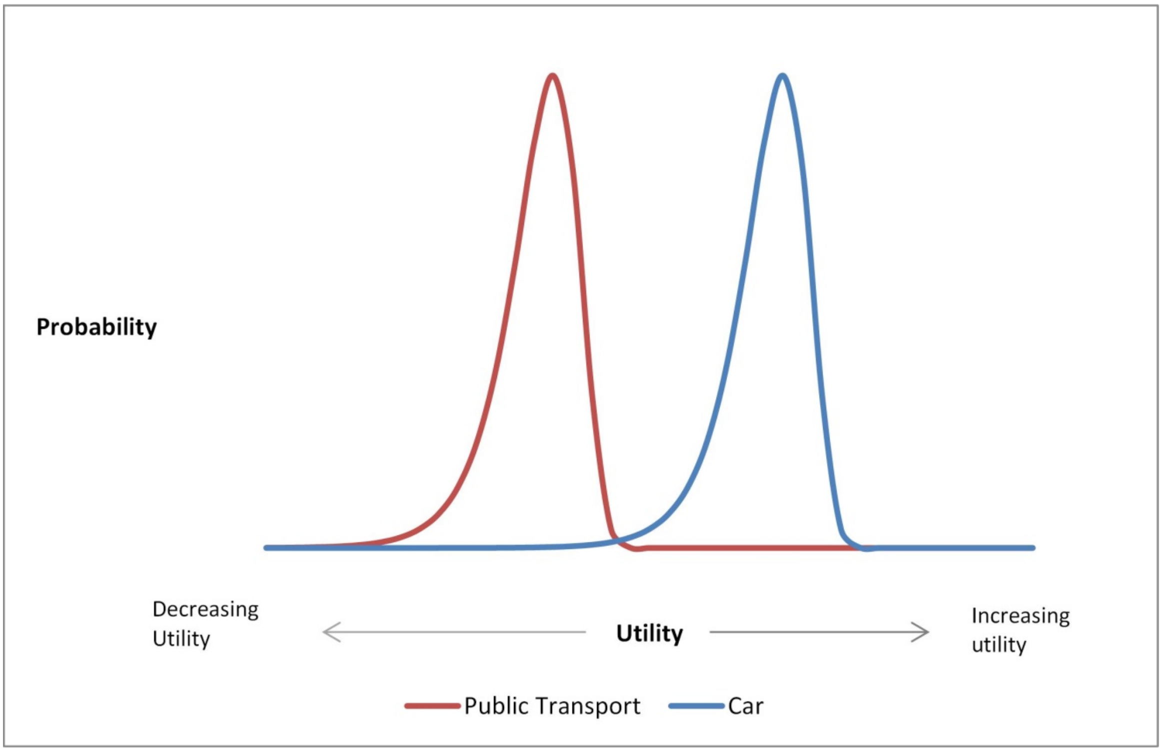

Figure 3.7: A Small Random Error Term

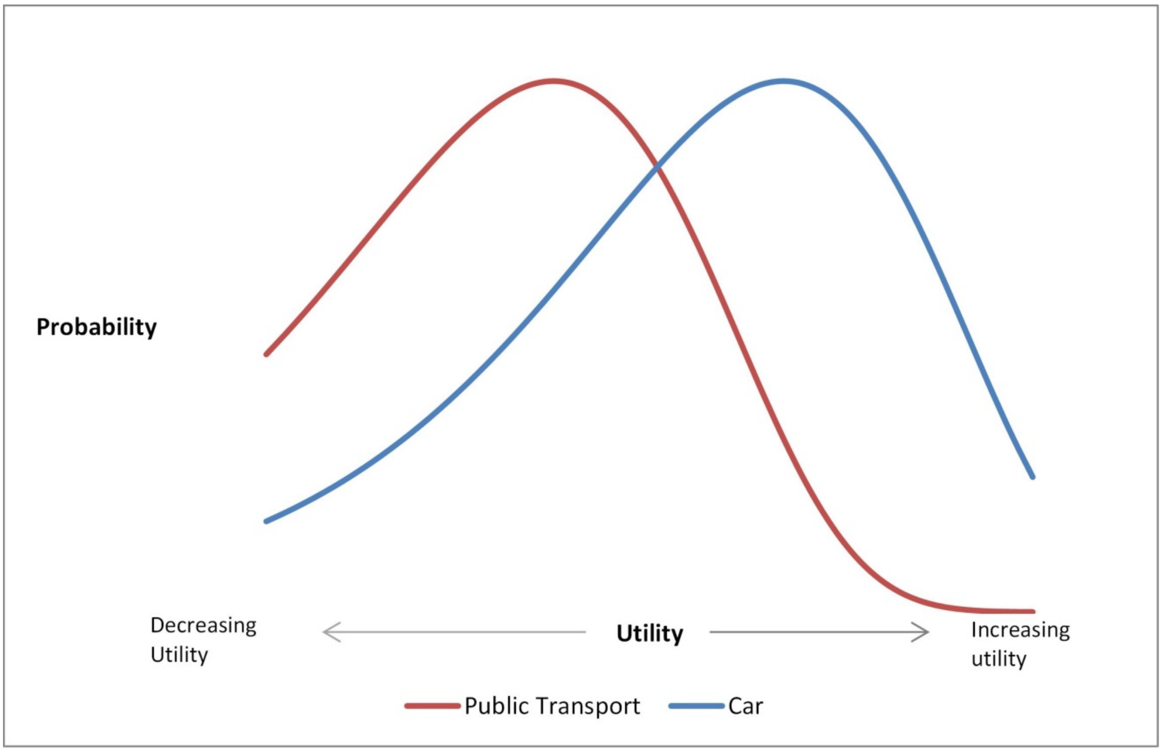

Figure 3.8: A Large Random Error Term

In the first case, the random error term is small in scale. As a result, the car and public transport probability distributions are quite narrow and separate, resulting in a high probability for car travel.

In the second case, the random error term is large in scale, and as a result, the probability distributions are wide and overlap to a greater degree. In this case, the probability of car use will be lower.

It needs to be noted that the mid-point of these distributions are identical in both examples. In other words, the modelled estimates of utility (V) were identical; all that varied was the scale of the random error term, and yet the predicted mode share changed considerably.

The scale of the random error term also dictates the sensitivity of the predicted mode shares. In the first example (small random error) the mode share will be insensitive to a small change in the car or public transport utilities. However, if the car utility were to move far enough left, and even just past the public transport distribution, then suddenly the mode share could flip from 100% car to 100% public transport.

By contrast, the changes in the second example will be much more gradual.

Conceptually, the random error represents all of the things which cannot be known about the individual and the choices they face. If V is the embodiment of <u>what is known</u>, then $\boldsymbol{\varepsilon}$ is the embodiment of <u>what is unknown</u>.

The more that is known, the more confident is the prediction; with a probability close to 1.0, a modeller will be very certain that an individual will choose a particular alternative.

The less that is known, the less confident the predictions become, and the more likely predictions are to hover near the mean mode share for the entire population. For example, a very poor model might predict an average public transport mode share of 14% for all Home Based Work trips, irrespective of the origin and destination.

So clearly it is important to know how much we don't know.

3.2 The Nested Logit Model

In a standard logit model (refer to section 0

The Logit Model), the distribution of the random error term, $\boldsymbol{\varepsilon}$**,** is assumed to have the same variance for all alternatives. Referring back to Figure 3.7 and Figure 3.8 in the previous section, note how the spread of the distributions was always identical for both car and public transport.

One other assumption was also implicitly made in the previous section, which was that these two distributions are unrelated, or independent; that is, one's car utility doesn't affect one's public transport utility, and vice versa.

Both of these assumptions can break down in the real world. Some alternatives (or modes, in our Mode Choice Model's case), have an inherently greater or lesser amount which is unknown about them (remembering that the scale of the random error term is reflective of the scale of unknown factors). The amount that is unknown about one's utility for cycling may differ from what is unknown about one's utility for car travel, or public transport. As such, the first assumption breaks down.

Furthermore, the utilities of competing alternatives are not always independent, especially when alternatives are very similar, or have a great deal in common.

Without going into great detail, violations of this kind can cause significant errors in prediction. Fortunately, a nested logit model allows the modeller to loosen these assumptions, which can improve the model's predictions.

The basic idea is that similar alternatives can be grouped together into what are commonly called "*nests*". A trip maker first chooses amongst nests, and then, having chosen a nest, then chooses amongst the alternatives which are grouped into that nest.

Nests can also contain nests, which lead to something of a tree-like structure. An example tree structure was presented earlier in Figure 2.1.