Zenith Transport Model

Technical Note 2
The Household Segmentation Model

Zenith Version: 2.0.0

VEITCH LISTER CONSULTING PTY LTD

Date: 06-04-2013

Zenith Transport Model

Technical Note 2: The Household Segmentation Model

Date	Revision	Prepared By	Checked By	Approved By	Description
05-04-2013	A	TV			Draft for comment

Contents

1 Introduction 1
1.1 Scope of This Document 1
2 The Household Segmentation Model 2
2.1 Context 2
2.2 Motivation 3
2.3 Methodology 5
2.3.1 Household Attributes for which Segmentation Is Required 5
2.3.2 Functional Form 6
2.4 Data Sources for Model Estimation 9
2.5 Parameter Estimation 10
2.6 Locational stability 11
3 Sources of Further Information 15

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 Household Segmentation model, the first step in the Zenith Model Run process.

1.1 Scope of This Document

The primary objective of this document is to detail the motivation and methodologies underpinning the Household Segmentation model.

This scope of this document does not extend to include the definition or validation of model parameters for specific markets. Refer to Section 3 for links to this information.

The remainder of this technical note is structured as follows:

- Section 2 describes the Household Segmentation model, the motivation for its development, its functional form, the data sources used for estimation, and its locational stability,
- Section 3 lists further sources of information relating to the Household Segmentation model.

2 The Household Segmentation Model

2.1 Context

The Household Segmentation model is the first component to be executed in a Zenith Model Run, as illustrated in Figure 1 below.

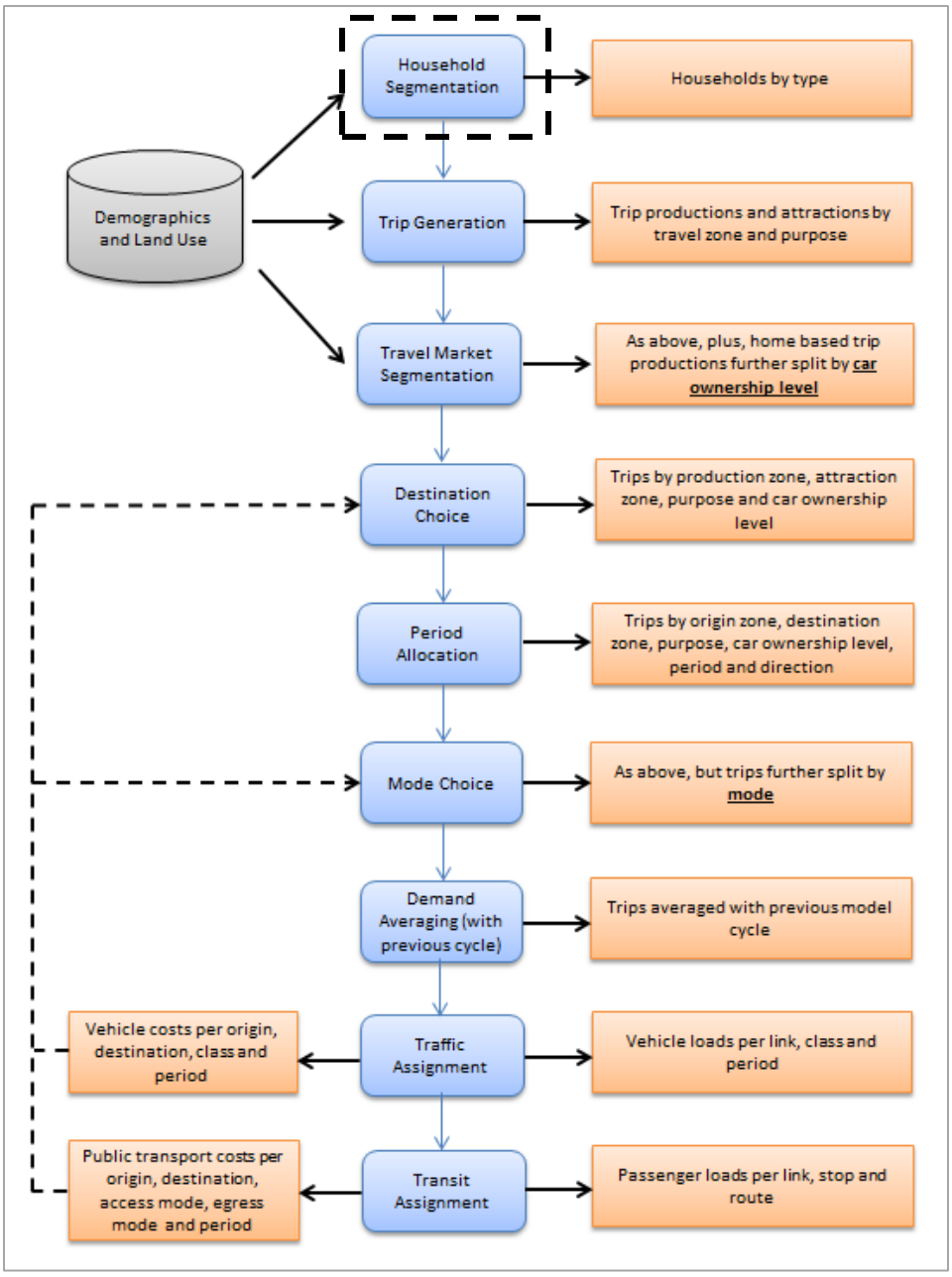

Figure 1 - Zenith Model Run Process

The Household Segmentation model's sole input is zonal demographic data (number of households, average household size, average cars per household, etc). Its output – households by type – is fed directly into the Trip Generation model.

2.2 Motivation

The aim of the Household Segmentation model is to segment households according to their level of a single attribute (e.g. level of car ownership), given an average (zonal) value for that attribute. This is perhaps best illustrated through an example.

Household Segmentation Example

Given that a travel zone has an average household car ownership of 1.3, what proportion of households are likely to own 0, 1, 2, or 3+ cars?

There is no single correct answer to this question; there are in fact an infinite number of car ownership distributions which all achieve an average of 1.3.

For example:

- 0 car - 0%
- 1 car - 70%
- 2 car - 30%
- 3+ car - 0%

The average of 1.3 is calculated as (0.7 x 1) + (0.3 x 2)

While this set of example proportions achieves the correct average, it is not realistic, as it predicts that all households have at least one car, and that none own more than 2 cars.

In this case, a more realistic distribution might be (taken from the Victorian model):

- 0 car - 15.5%
- 1 car - 48.5%
- 2 car - 28.3%
- 3+ car - 7.7%

The motivation for segmentation is two-fold:

- The use of segmented households results in a more accurate trip generation model,

- It would be impractical for the modeller to define segmented households as a model input. Therefore, we need a model that reliably converts averages (e.g. average car ownership), into distributions.

Exploring these issues further, the trip generation model used in Zenith takes as input the following variables *for each household*:

- Number of White Collar workers
- Number of Blue Collar workers
- Number of dependants aged 0-17
- Number of dependants aged 18-64
- Number of dependants aged 65+
- Number of cars owned

As such, an average car ownership (eg. 1.3) is not enough; we need to know whether a specific household has 0, 1, 2 or 3+ cars, and 0, 1, 2 or 3+ white collar workers, etc.

We could ask the user to exogenously define these distributions as an input to the model; they are, after all, available directly from the ABS Census. However, defining these distributions for future year scenarios would be onerous and impractical.

Instead, it is more realistic for economists and demographers to produce estimates of *average household sizes, average car ownership, average workers per household, etc*, and to supply these as inputs to the model. The role of the Household Segmentation model, therefore, is to act as a bridge between the demographer and the trip generation model, converting averages into realistic distributions.

2.3 Methodology

2.3.1 Household Attributes for which Segmentation Is Required

A unique Household Segmentation model is required for each of the following six attributes which are input to the trip generation model:

- Blue collar workers
- White collar workers
- Dependants aged 0-17
- Dependants aged 18-64
- Dependants aged 65+
- Cars owned

Note that white and blue collar have been defined using ANZSCO1 as follows in Table 1 below.

ANZSCO1 classification	White / blue
1. Managers	White
2. Professionals	White
3. Technicians and Trades Workers	Blue
4. Community and Personal Service Workers	White
5. Clerical and Administrative Workers	White
6. Sales Workers	White
7. Machinery Operators and Drivers	Blue
8. Labourers	Blue

Table 1 - Definition of White and Blue Collar (using ANZSCO1)

Also note that dependants are defined as follows:

- Anyone under the age of 15, plus
- Non-workers aged 15+

Finally, note that car ownership does not include motorcycles, as motorcycle ownership is not collected by the ABS Census.

2.3.2 Functional Form

Research conducted by VLC in the 1990s found that a family of modified logistic curves could be developed to accurately predict the distribution of households across the levels of an attribute, given a zonal average of that attribute.

This methodology was also adopted by the Government models in Queensland and Victoria – the Brisbane Strategic Transport Model (BSTM) and Melbourne Integrated Transport Model (MITM).

The most appropriate form of logistic (logit) relationship was found to be as follows:

$$H_n(x) = \frac{200 - A_n}{1 + e^{1/B_n(x - C_n)}}$$

Where: $H_n(x)$ is the percentage of households which have a value of n or less for the given attribute

x is the average value of the attribute for the travel zone

A_n, B_n and C_n are the model parameters for level n

Note: this equation is a rearranged (but equivalent) version of the model form developed and published by VLC in its original Zenith Technical Notes.

Some example curves are presented in Figure 2, with each curve defined by its unique set of model parameters, A, B and C. Broadly speaking, A is a vertical placement factor, B is a logit slope factor and C is a horizontal offset.

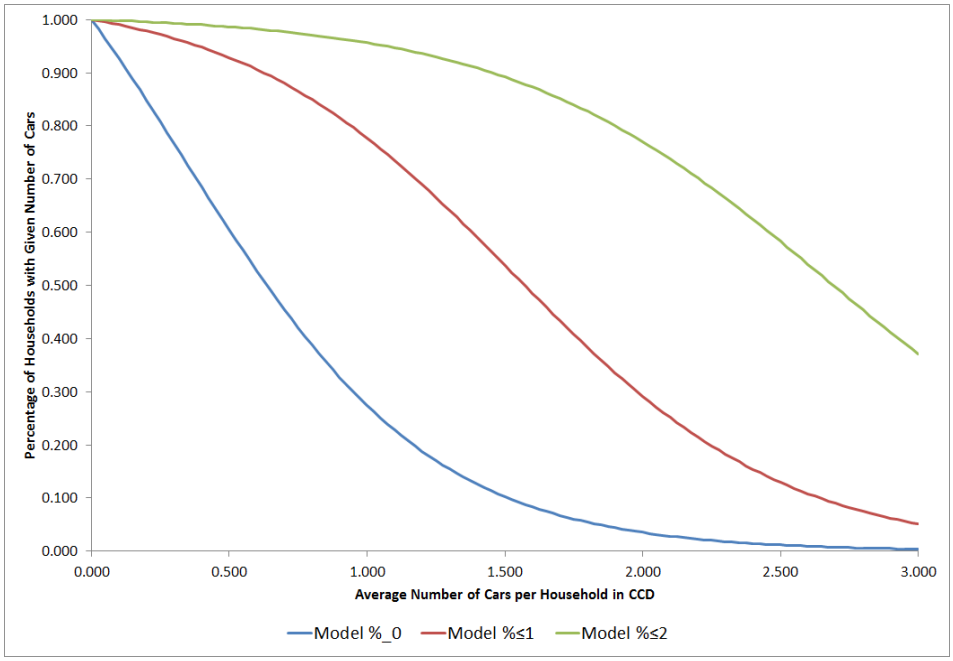

Figure 2 – An Example Family of Logistic Curves

The interpretation of these curves is made clearer in Figure 3, which highlights that the curves are, in fact, the boundaries between different levels of a variable.

The example shows the estimated segmentation of households into discrete car ownership levels given an average car ownership level of a travel zone (the x-axis). The grey region (below the first curve) represents the proportion of households that don't own a car, the blue region (between the first and second curves) represents households which own one car, with the red and green regions representing households which own 2 and 3+ cars respectively.

It can be seen that when the average number of cars per household (the x-axis) is zero, all households don't own a car. However, as the average car ownership increases, the proportion of households with no car decreases, with more households having higher car ownership.

The estimated distribution for an average car ownership level of 1.3 is marked on the Figure. This particular case was presented in the example in Section 2.1.

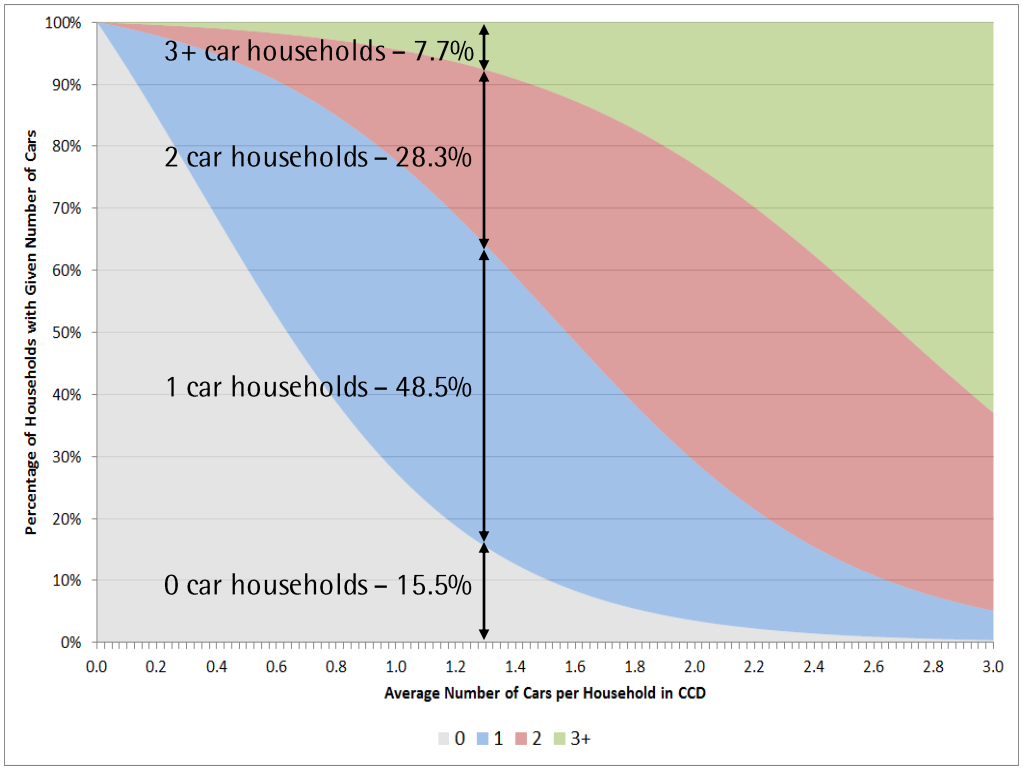

Figure 3 – Car Ownership Segmentation Example

2.4 Data Sources for Model Estimation

The parameters for each of the 6 household segmentation models are estimated using data obtained from the ABS Census. The most recent model estimation exercise was conducted using the 2006 Census.

The dataset, which is described in Table 2 below, was requested for all the Census Collector Districts (CCDs) in Australia, to allow us to test the spatial transferability of the Household Segmentation model.

Census Variable (for each CCD)
Average number of white collar workers per household
The number of households with 0, 1, 2, or 3+ white collar workers
Average number of blue collar workers per household
The number of households with 0, 1, 2, or 3+ blue collar workers
Average number of dependants 0-17 per household
The number of households with 0, 1, 2, or 3+ dependants aged 0-17
Average number of dependants 18-64 per household
The number of households with 0, 1, 2, or 3+ dependants aged 18-64
Average number of dependants 65+ per household
The number of households with 0, 1, 2+ dependants aged 65+
Average number of cars owned per household
The number of households with 0, 1, 2, 3+ cars

Table 2 - The Custom 2006 ABS Census Variables used for Model Recalibration

Conceptually, the dataset consists of average CCD attribute values (e.g. average car ownership), as well as the distribution among discrete levels of each attribute (e.g. car ownership distribution).

2.5 Parameter Estimation

The aim of parameter estimation is to determine parameters A, B, and C, which best fit the data obtained from the ABS Census.

An example, relating to car ownership, is shown in Figure 4 below. Each CCD is represented by vertically aligned blue, red and green dots (one of each), indicating the proportion of households in that CCD which own 0, 1, 2, or 3+ cars.

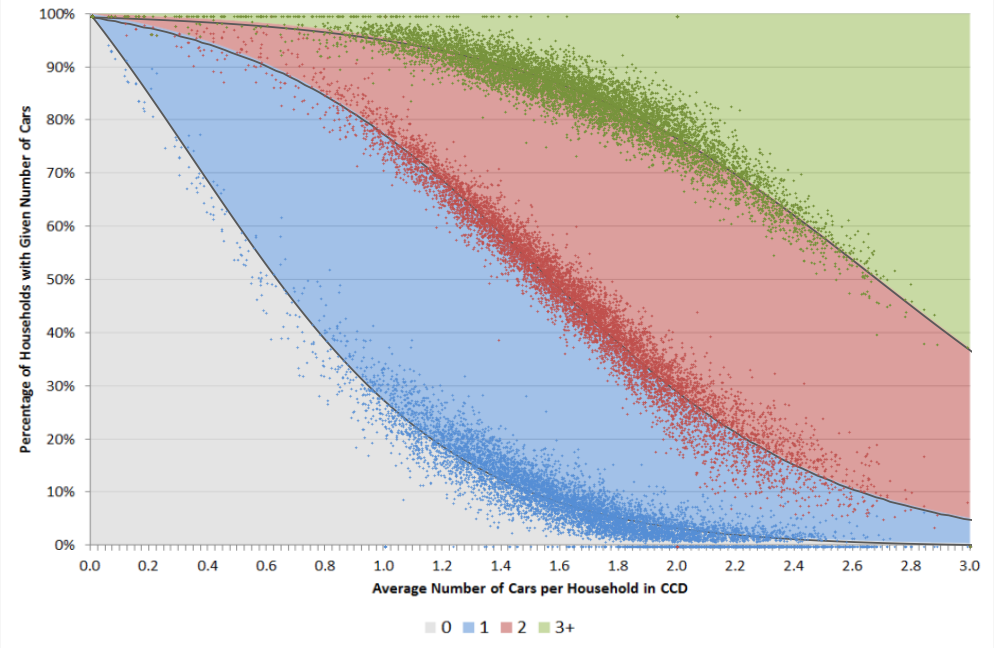

Figure 4 - Curve Fitting for Car Ownership Segmentation

The fitted curves are drawn in the Figure. It is clear that the curves accurately fit the data (all of the curves have an $R^2 > 0.91$), confirming that it is possible to estimate the distribution of households based on a zonal average, with some degree of confidence.

2.6 Locational stability

Household Segmentation models have been developed and applied by VLC in many Australian cities since 1986, and have always exhibited remarkable consistency.

This suggests that the relationships are spatially stable and transferrable (temporal stability might be explored at a later date by analysing historical Census results).

To illustrate this, we have overlayed the fitted curves for three distinct regions: Melbourne, Sydney, and South East Queensland. The comparisons, for each segmented attribute, are found in **Figure 5** through **Figure 10** below.

Referring to **Figure 5**, it can be observed that:

- There is remarkable consistency between the segmentation curves of the three cities (they often sit on top of each other),
- There is a slight difference in the tail of the 2 / 3+ (topmost) set of segmentation curves; this is almost certainly due to a shortage of data points (CCDs) with averages at the high end of the spectrum (above, say, 1.3).

This pattern of consistency is repeated through all of the household attributes..

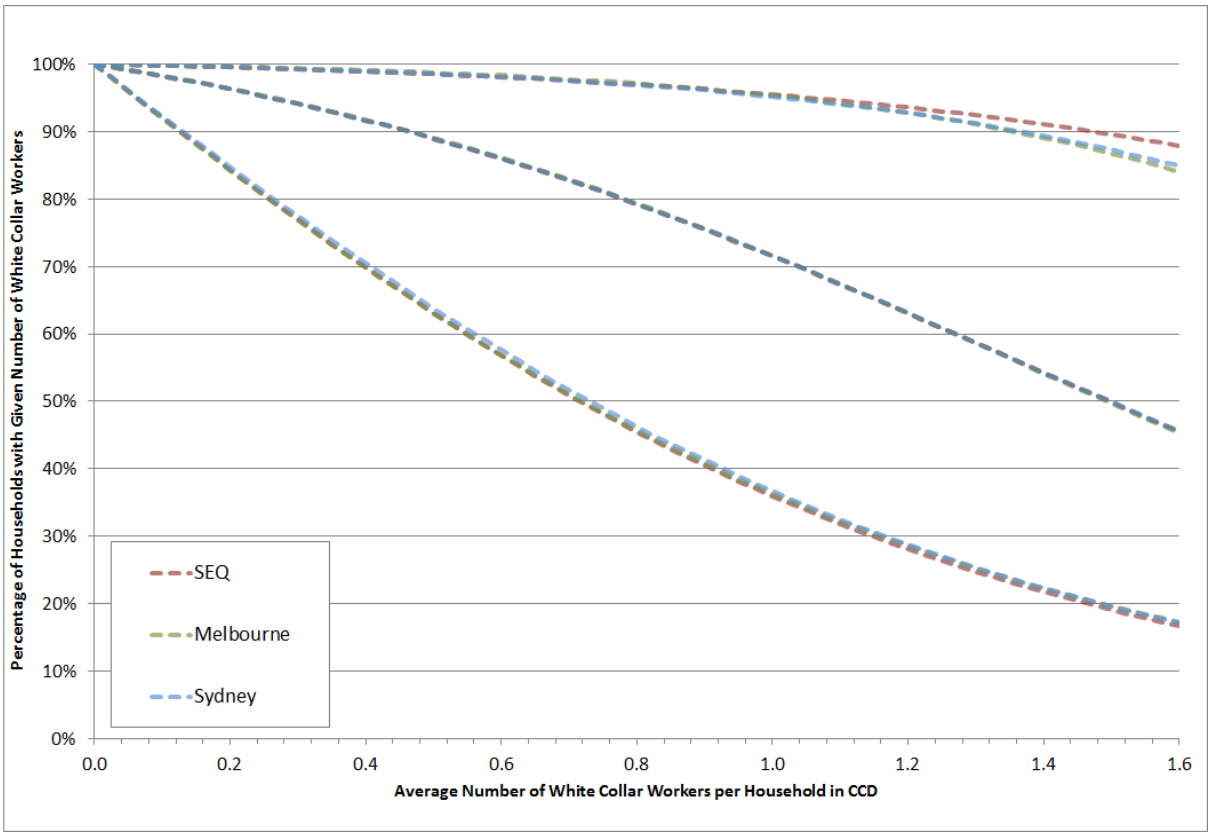

Figure 5 –Segmentation of White Collar Workers by Region

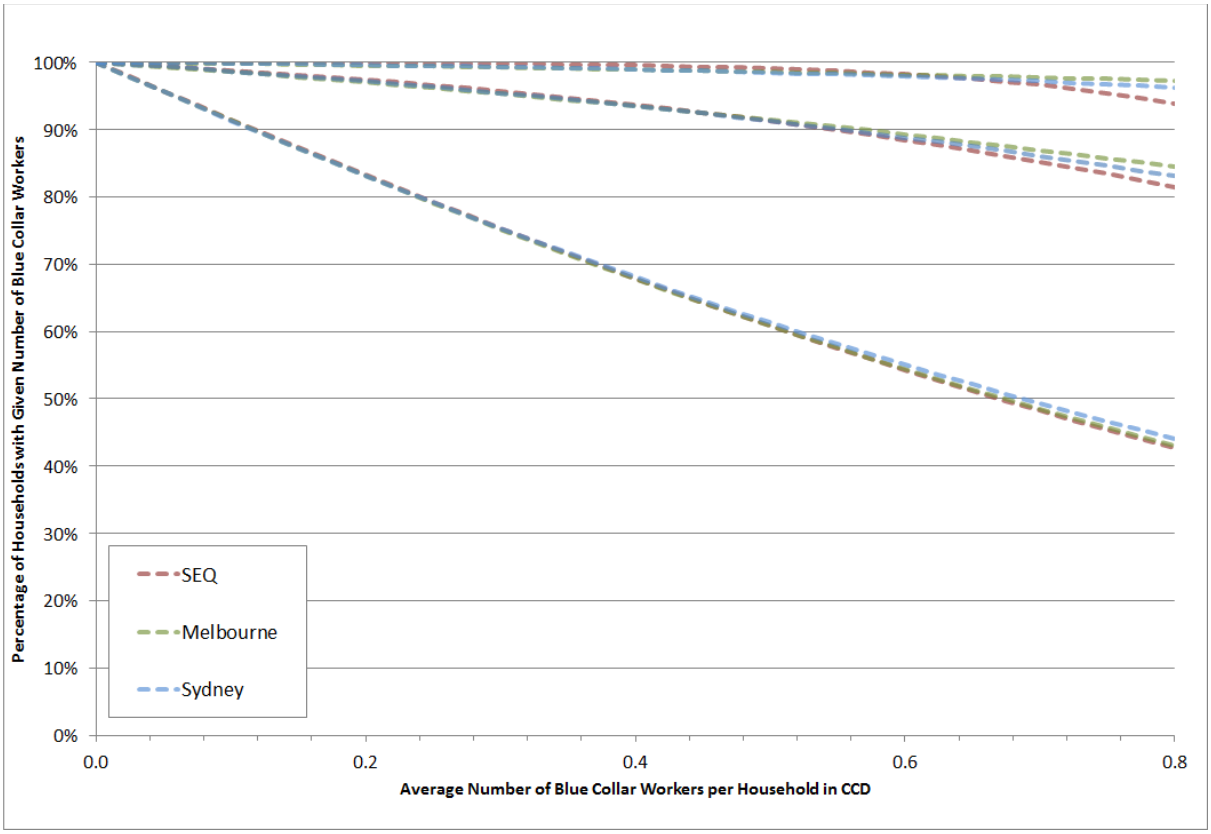

Figure 6 – Segmentation of Blue Collar Workers by Region

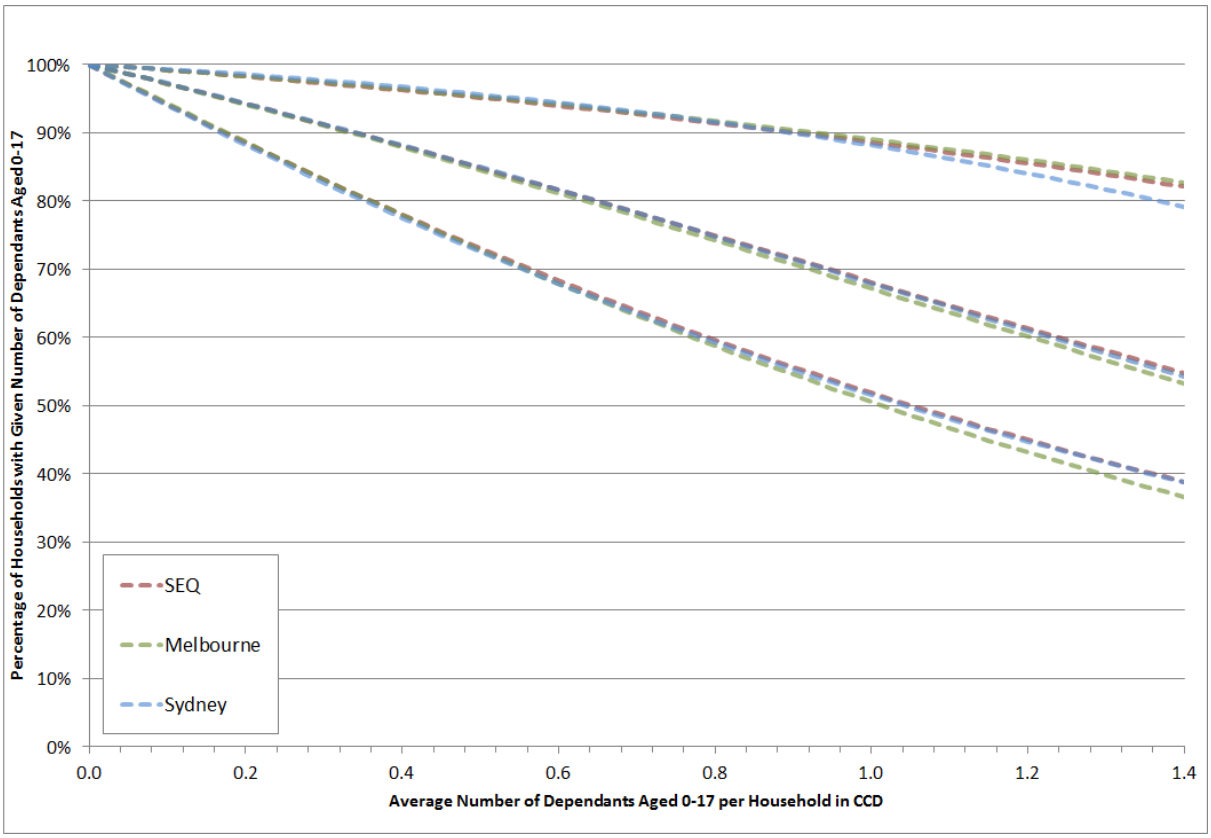

Figure 7 - Comparison of Segmentation of Dependants Aged 0-17

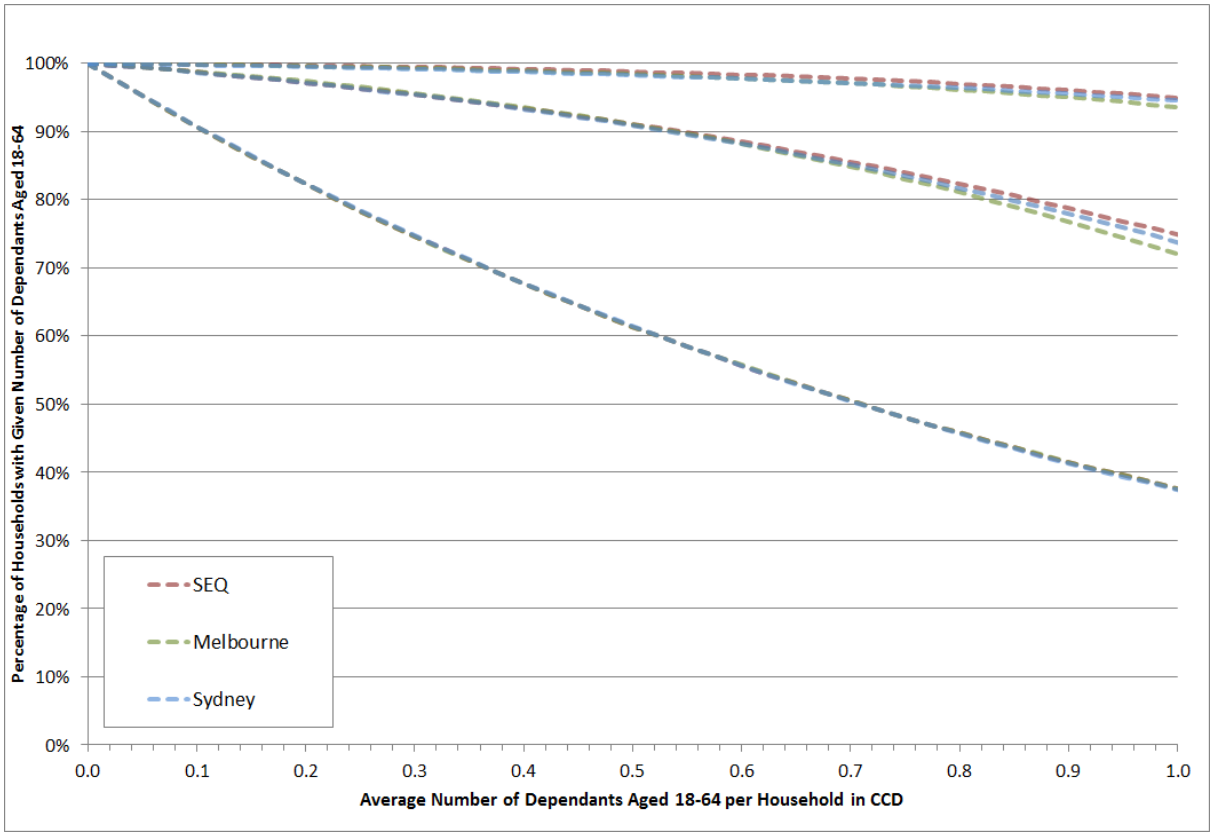

Figure 8 - Segmentation of Dependants Aged 18-64 by Region

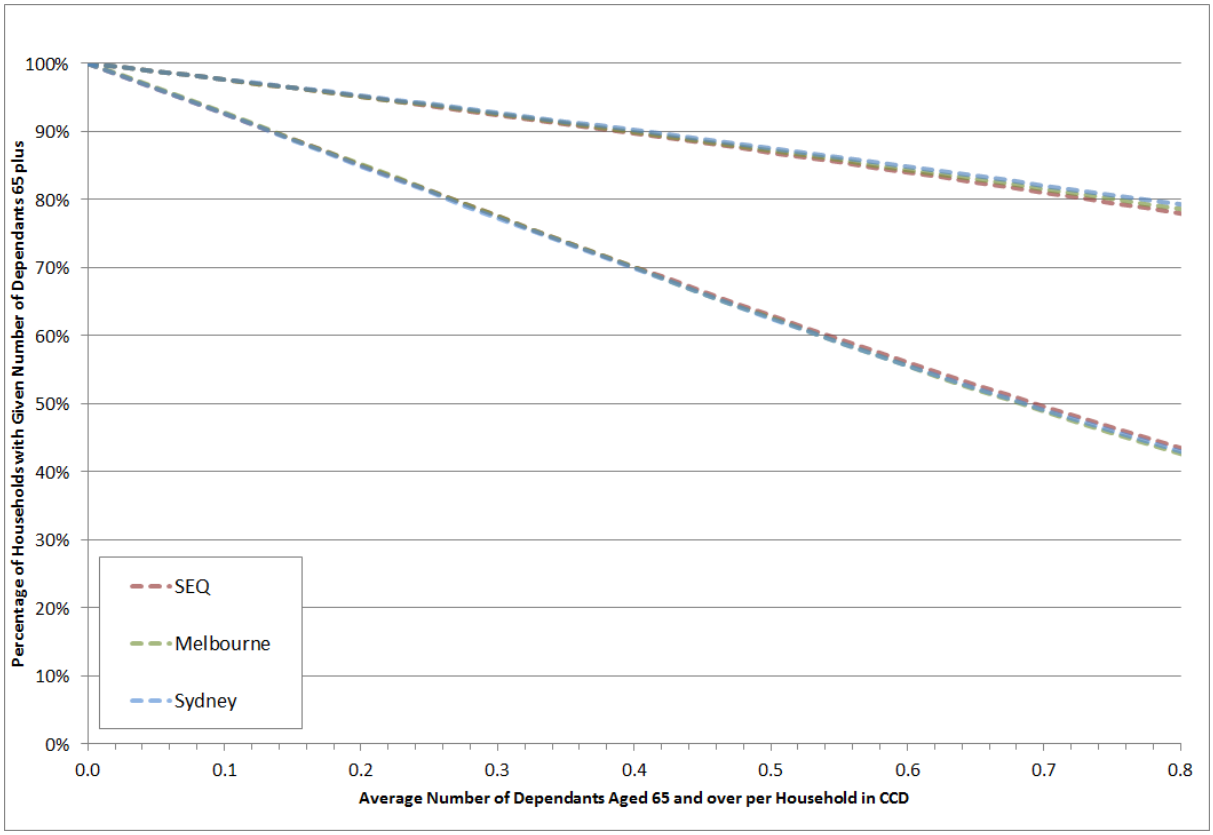

Figure 9 – Segmentation of Dependants Aged 65 plus by Region

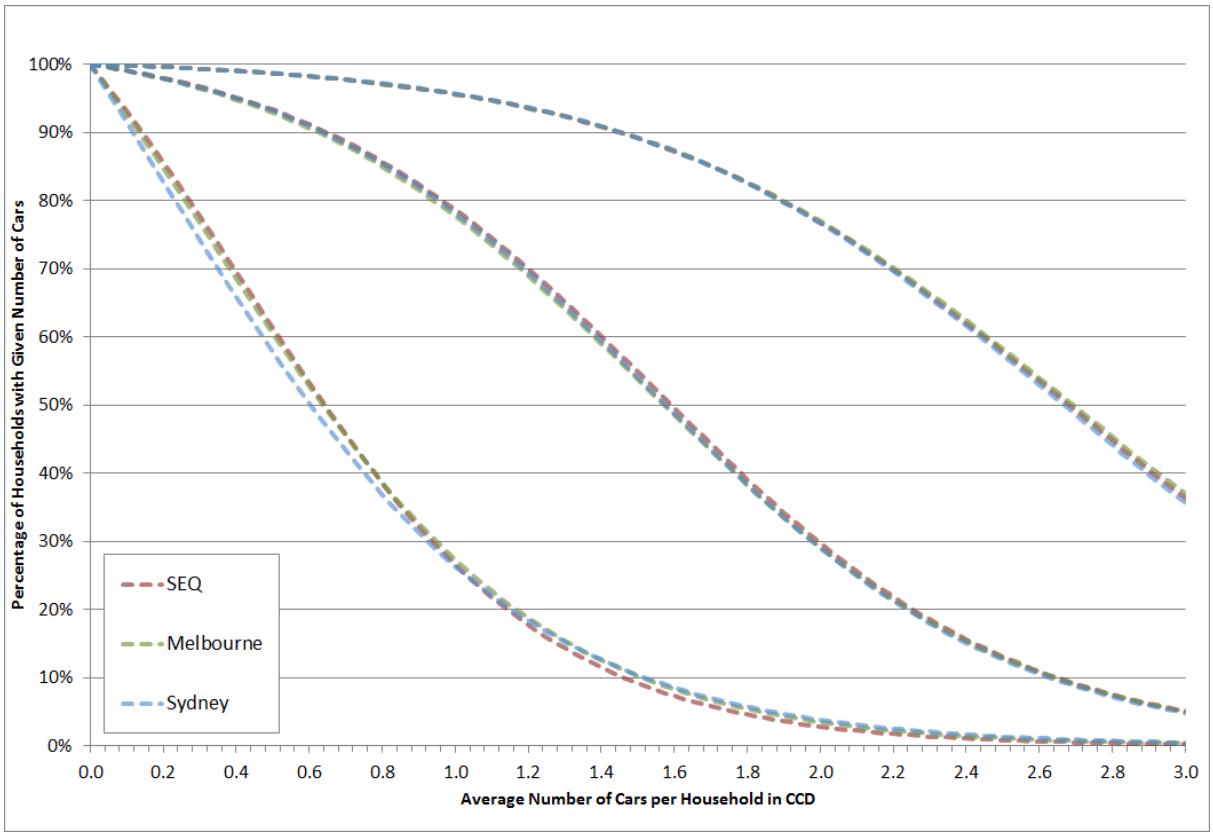

Figure 10 – Segmentation by Car Ownership Level by Region

3 Sources of Further Information

This technical note has described the objectives and methodology underpinning the Zenith Household Segmentation model.

The note has not, however, listed the actual parameters estimated and applied for each modelled region. These can be found in the *model specific* technical notes relating to household segmentation. Simply go to: http://zenith.veitchlister.com.au, and go the tab which reflects your region of interest.

For further information, please contact Tim Veitch as per the contact details below.

Tim Veitch

email: zenith@veitchlister.com.au

phone: +61 7 3870 4888

web: http://zenith.veitchlister.com.au