

Zenith Model Framework Papers -
Version 3.0.1

Paper B – Household Segmentation Model

May 2014

Page Intentionally Left Blank

Zenith Model Framework Papers – Version 3.0.1

Paper B – Household Segmentation Model

Draft Report

COPYRIGHT: The concepts and information contained in this document are the property of Veitch Lister Consulting Pty Ltd. Use or copying of this document in whole or in part without the written permission of Veitch Lister Consulting constitutes an infringement of copyright.

LIMITATION: This report has been prepared on behalf of and for the exclusive use of Veitch Lister Consulting Pty Ltd's Client, and is subject to and issued in connection with the provisions of the agreement between Veitch Lister Consulting and its Client. Veitch Lister Consulting accepts no liability or responsibility whatsoever for or in respect of any use of or reliance upon this report by any third party.

Date	Revision	Prepared By	Checked By	Approved By	Description
16/03/2014	A	MP	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 Household Segmentation Model.

The aim of the Household Segmentation Model is to take an average value for a travel zone attribute (i.e. Household Size) and estimating the distribution of households within the categories of that attribute (i.e. 1 person household, 2 person household, ... 6+ person household).

The reason for doing this is that such segmented data improves the accuracy of the Trip Generation Model, but it is very difficult to produce this data in future year scenarios. By allowing the model user to simply provide average values (which are much easier to predict) the usability of the model is greatly enhanced.



Contents

Executive Summary	ii
Contents	iii
List of Figures	iv
List of Tables	v
1 Introduction	7
2 The Household Segmentation Model	8
2.1 Background	8
2.2 Methodology	10
2.2.1 Household Attributes	10
2.2.2 Functional Form	10
2.2.3 Parameter Estimation	12
2.3 Locational stability	13



List of Figures

Figure 2.1: Zenith Model Run Process	8
Figure 2-2 – An Example Family of Logistic Curves	11
Figure 2.3: Car Ownership Segmentation Example.....	12
Figure 2.4: Curve Fitting for Car Ownership Segmentation	13
Figure 2.5: Segmentation of White Collar Workers by Region.....	14
Figure 2.6: Segmentation of Blue Collar Workers by Region.....	14
Figure 2.7: Comparison of Segmentation of Dependants Aged 0-17.....	15
Figure 2.8: Segmentation of Dependants Aged 18-64 by Region.....	15
Figure 2.9: Segmentation of Dependants Aged 65 plus by Region.....	16
Figure 2.10: Segmentation by Car Ownership Level by Region	16



List of Tables

Table 1: Definition of White and Blue Collar (using ANZSCO1).....	10
---	----



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 Household Segmentation model, which is the first sub-model within the overall Zenith Model framework.

This document focuses on the methodology of the Household Segmentation 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 Household Segmentation (located at <http://zenith.veitchlister.com.au>).

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



2 The Household Segmentation Model

2.1 Background

The Household Segmentation Model is the first component to be executed in a Zenith Model Run, as illustrated in Figure 2.1 below.

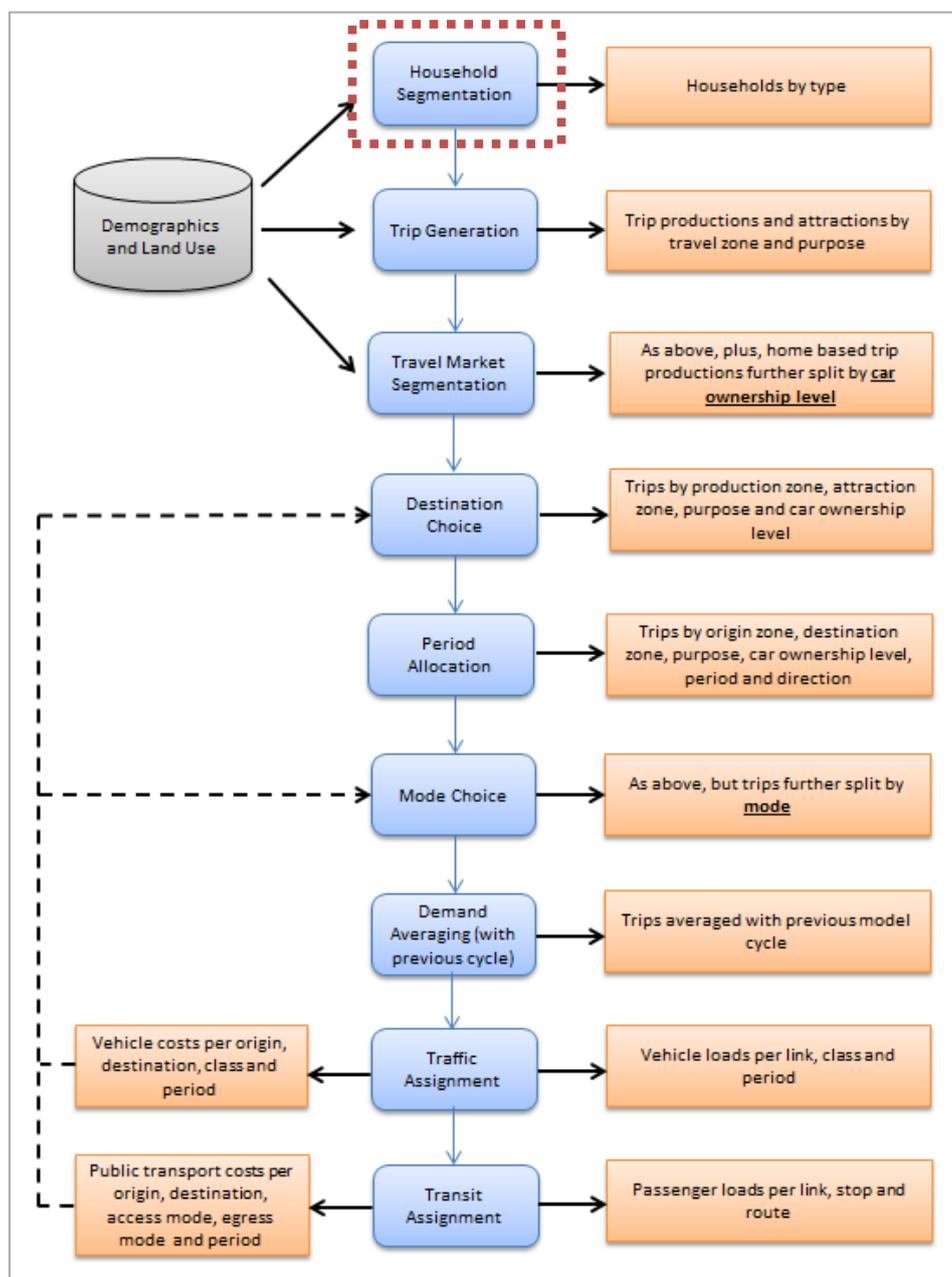


Figure 2.1: Zenith Model Run Process

The aim of the Household Segmentation Model is to segment households according to their level of particular attributes (e.g. level of car ownership), given an average (zonal) value for each attribute. This is illustrated in the example below.



Household Segmentation Example

If the average car ownership for households resident in a given travel zone is 1.3 cars / household, 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 $(0.7 \times 1) + (0.3 \times 2)$

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

In this case, a more likely 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 Household Segmentation Model takes as input zonal averages for a range of household attributes, including car ownership, household size, number of workers, etc. The output of the model – households by type – is fed directly into the Trip Generation Model.

The motivation for using a Household Segmentation Model is two-fold:

- The segmentation of households into different groups (i.e. 0 car households, 1 car households, etc.) results in a more accurate Trip Generation Model; and
- By using a Household Segmentation Model, it is not necessary for the model user to produce counts of households per segment (per travel zone) as input to the model. Instead, model users only need to supply the model with *average values* – i.e. average car ownership, average household size, etc. for each travel zone. This greatly reduces the effort required to create new demographic scenarios.



2.2 Methodology

2.2.1 Household Attributes

A unique Household Segmentation Model is required for each of the following seven household attributes which are input to the Trip Generation Model:

- Household size
- Number of blue collar workers
- Number of white collar workers
- Number of dependants aged 0-17
- Number of dependants aged 18-64
- Number of dependants aged 65+
- Number of 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.

2.2.2 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 value 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.

The main input to the model is the average value of the particular attribute (x). The model parameters (A, B, and C) vary by the level of the attribute, and by trip purpose.

Some example curves are presented in Figure 2-2, 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.

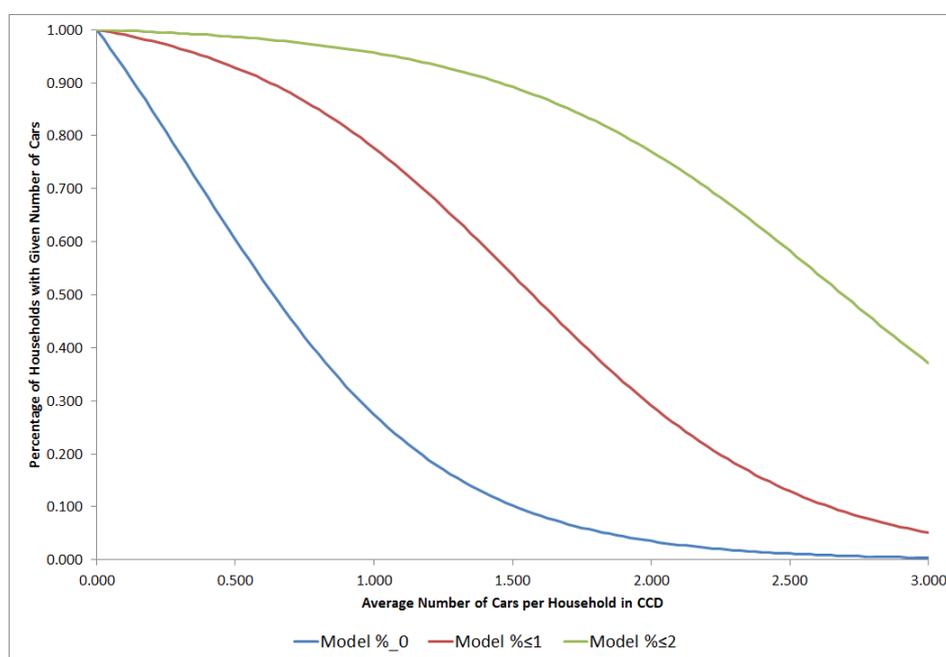


Figure 2-2 – An Example Family of Logistic Curves

The interpretation of these curves is made clearer in Figure 2.3, which highlights that the curves are, in fact, the boundaries between different levels of the 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 annotated on top of Figure 2.3 (this is the particular case that was provided as a worked example in Section 2.1).

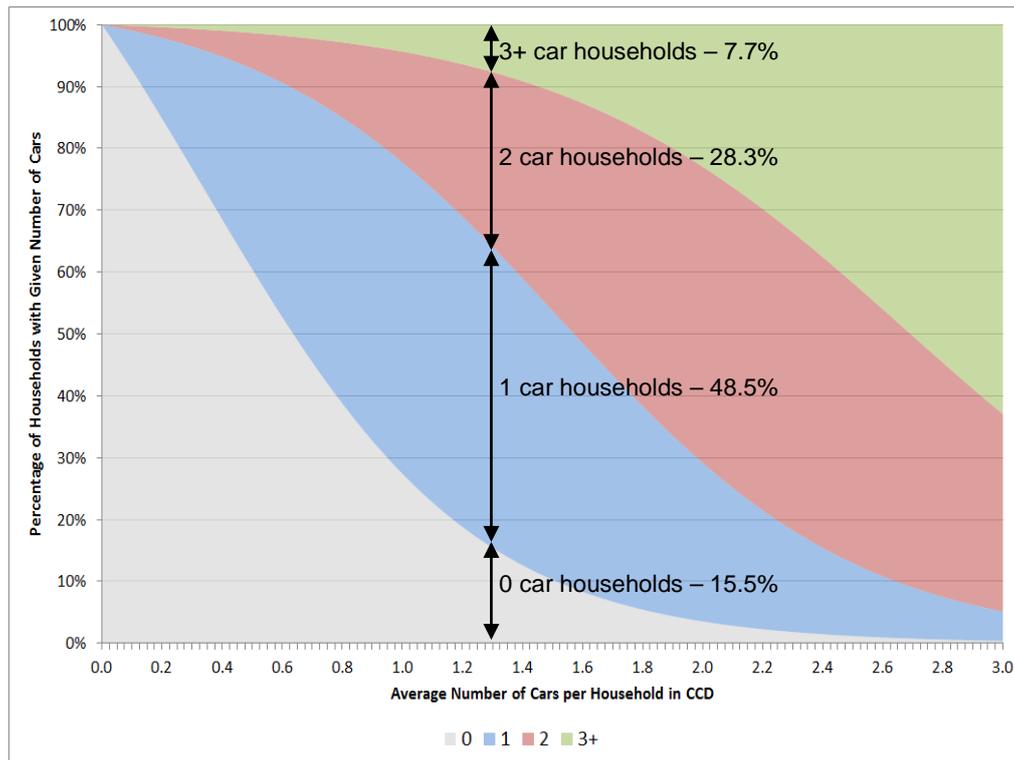


Figure 2.3: Car Ownership Segmentation Example

2.2.3 Parameter Estimation

The model parameters A , B and C for each of the 7 Household Segmentation Models are estimated using data obtained from the Census. Parameter estimation is performed using non-linear least squares regression.

An example, relating to car ownership, is shown in Figure 2.4 below. Each Census Statistical Area is represented by vertically aligned blue, red and green dots (one of each), indicating the proportion of households in that area which own 0, 1 or less, or 2 or less cars.

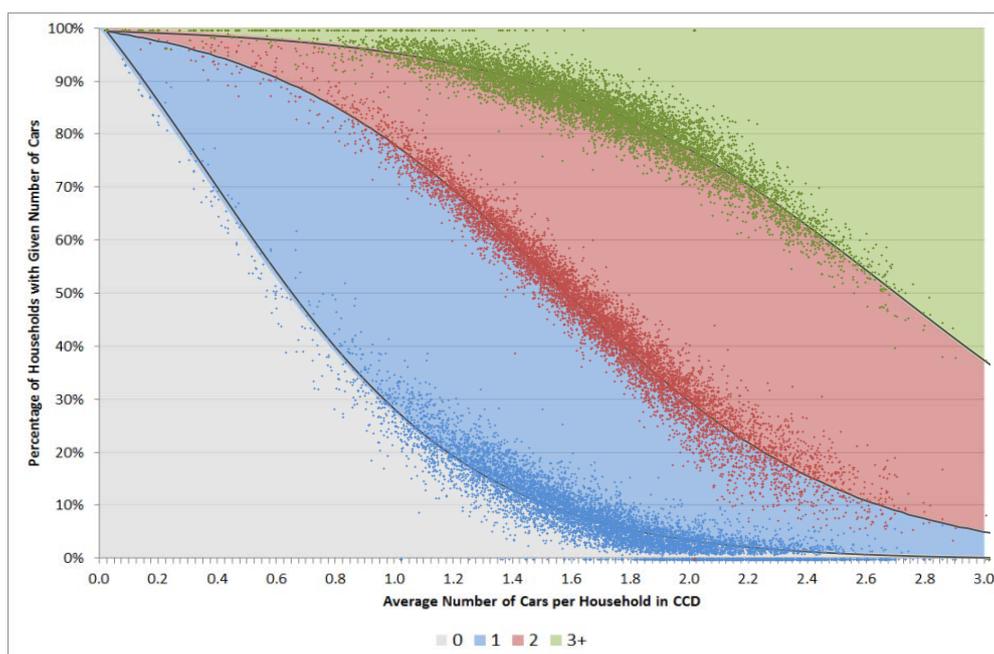


Figure 2.4: Curve Fitting for Car Ownership Segmentation

The fitted curves are drawn in Figure 2.4; from the closeness of the fit, it is clear that the curves accurately fit the data confirming that it is possible to reliably estimate the distribution of households based on a zonal average.

2.3 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 and temporally stable and transferrable between Australia capital cities.***

In this section, locational stability is examined by comparing the curves estimated for three Australian regions (Victoria, Sydney, and South East Queensland). The results are presented in Figure 2.5 through to Figure 2.10 below.

Referring to Figure 2.5, it can be observed that:

- There is remarkable consistency between the segmentation curves of the three cities;
- 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 with large average values.

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

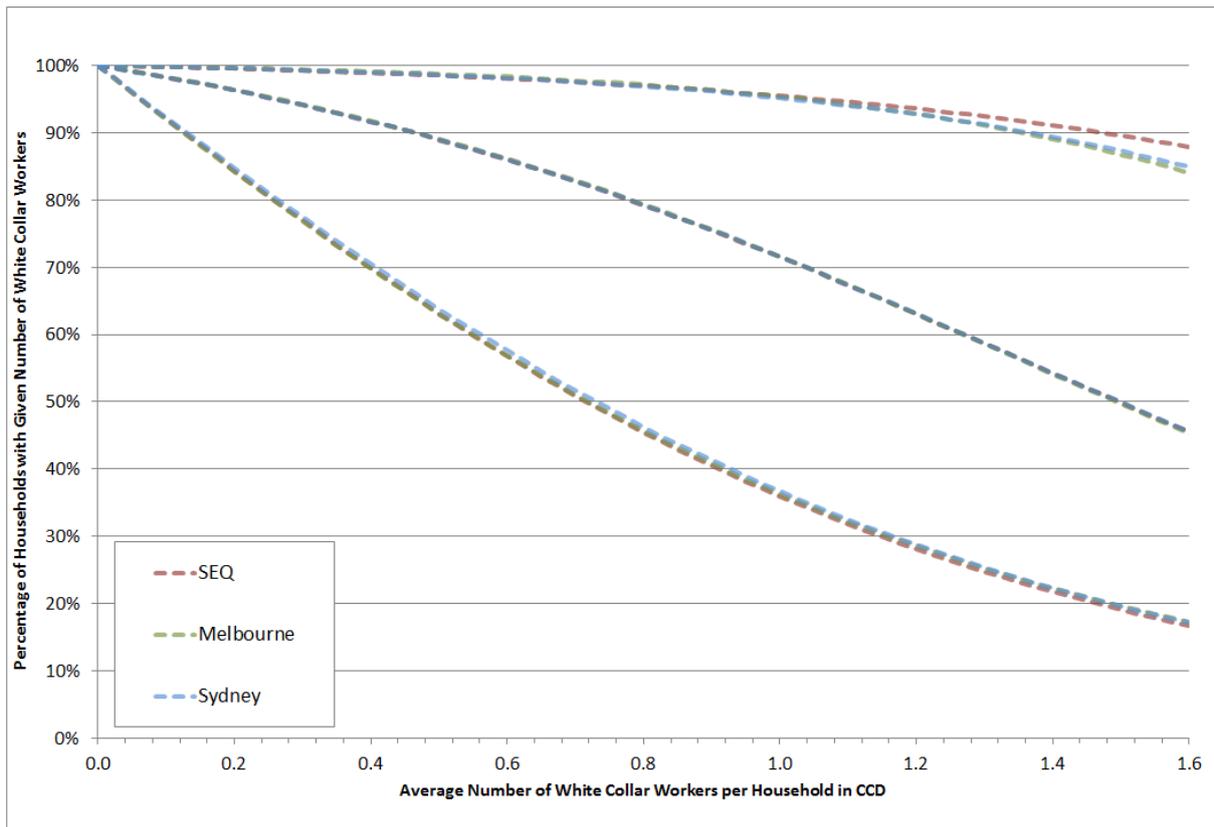


Figure 2.5: Segmentation of White Collar Workers by Region

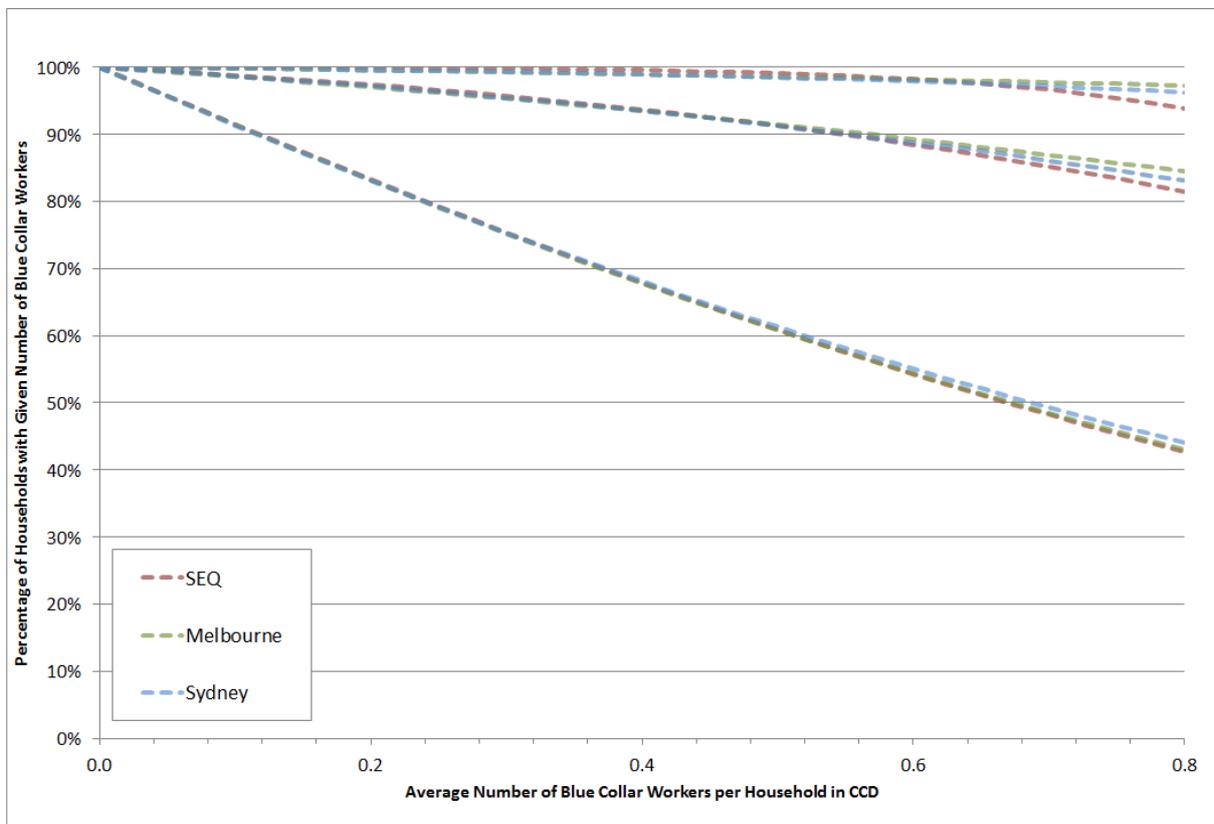


Figure 2.6: Segmentation of Blue Collar Workers by Region

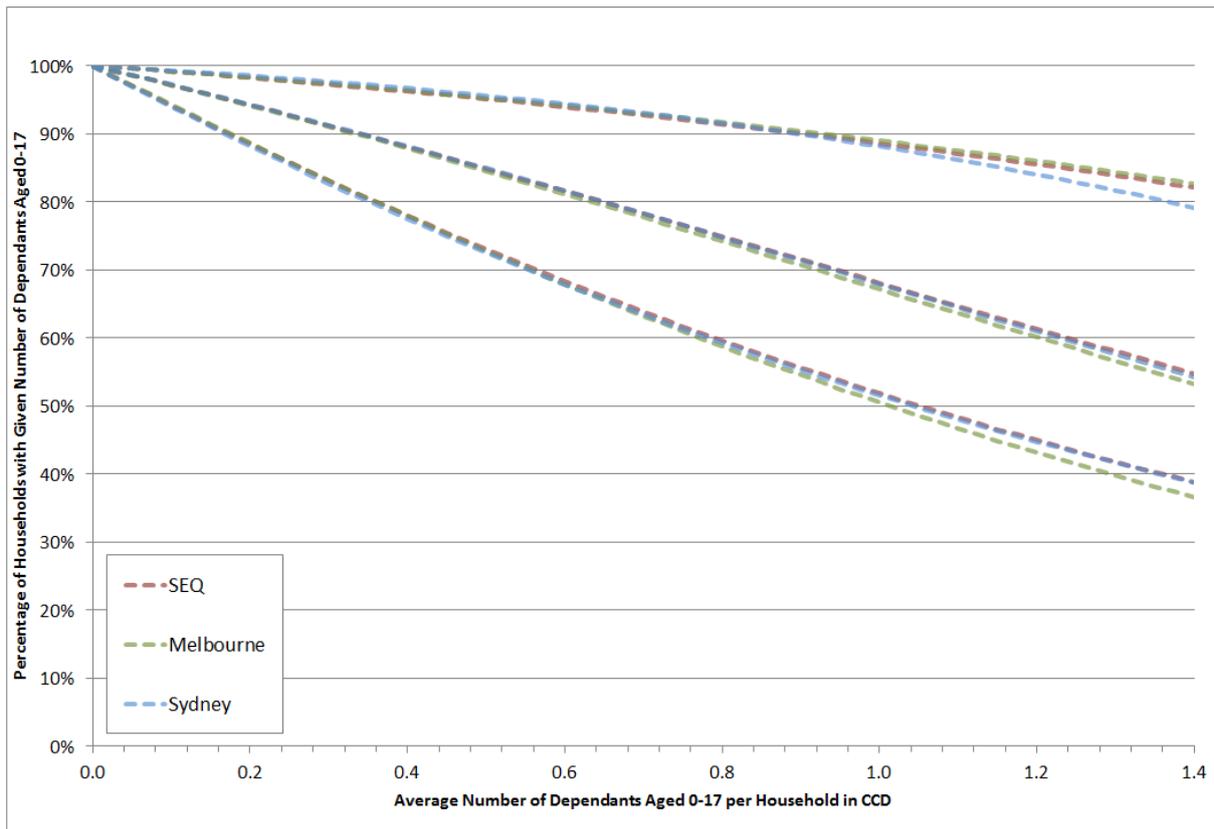


Figure 2.7: Comparison of Segmentation of Dependants Aged 0-17

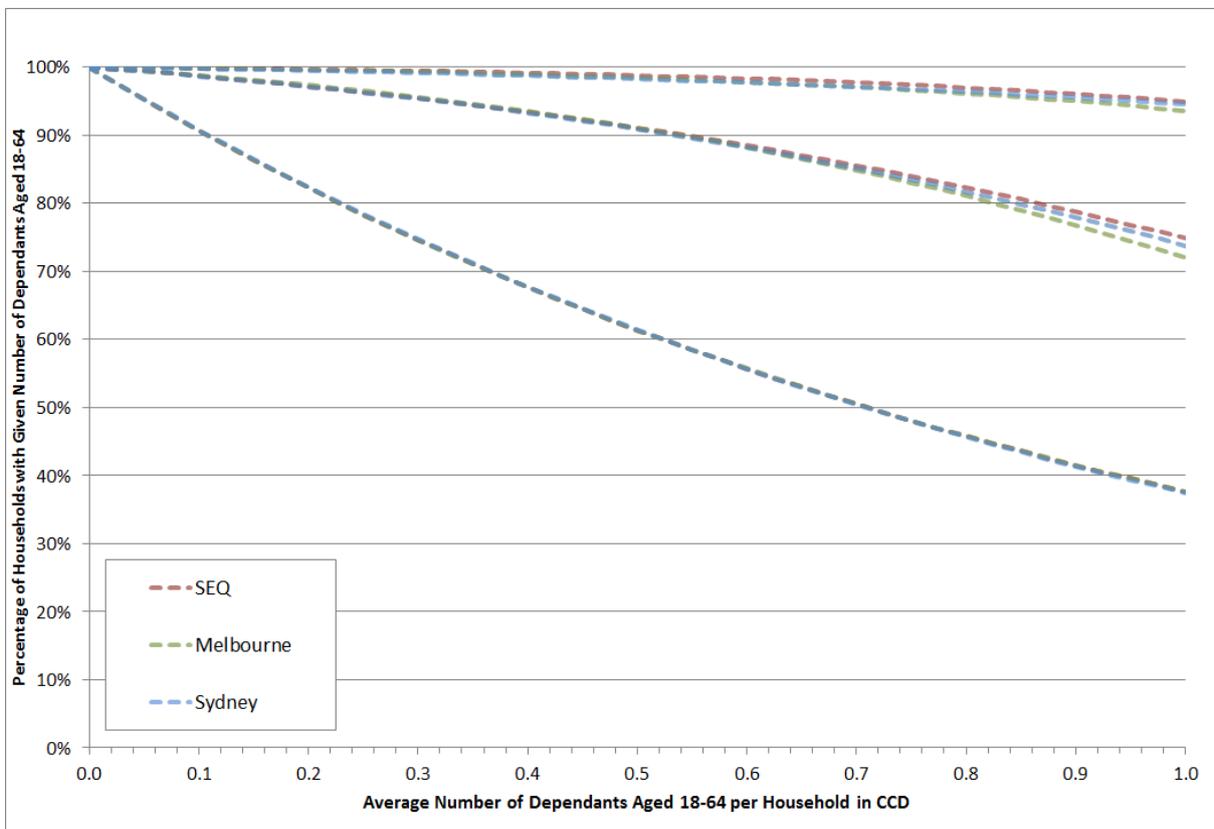


Figure 2.8: Segmentation of Dependants Aged 18-64 by Region

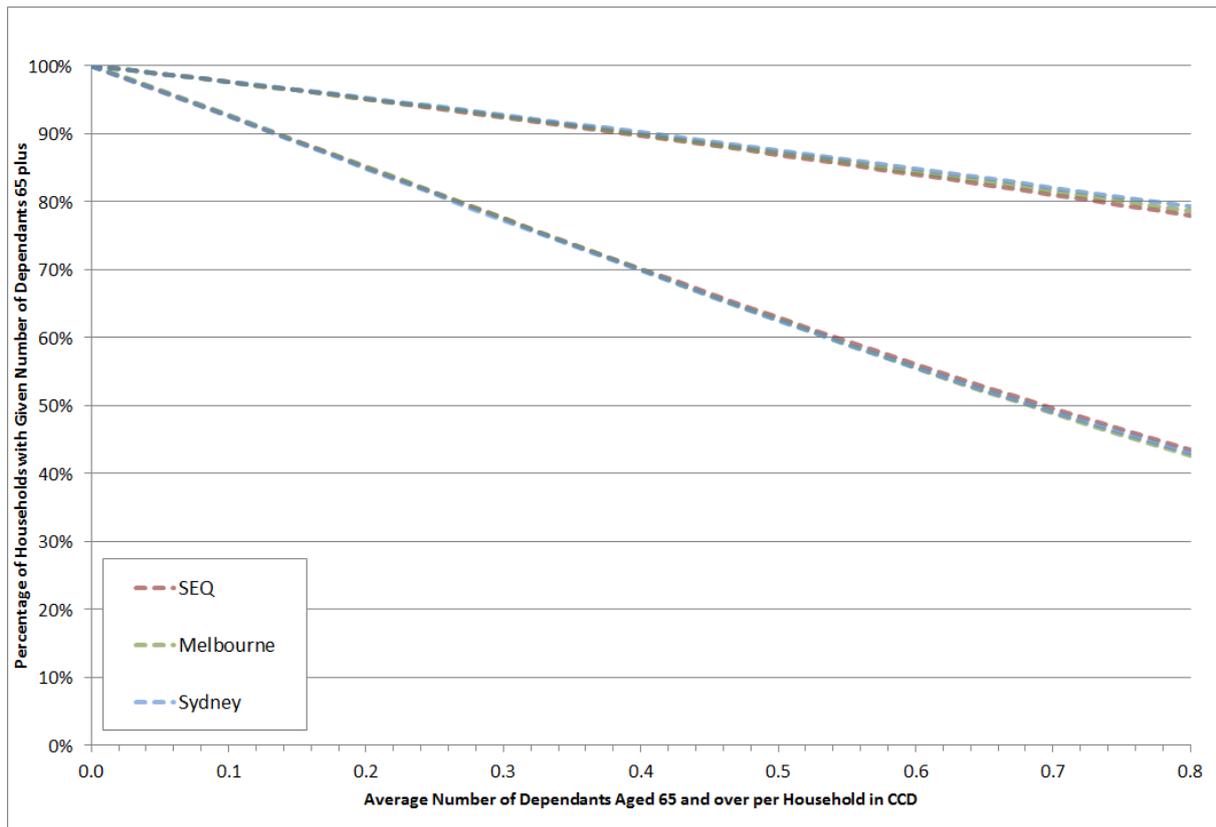


Figure 2.9: Segmentation of Dependants Aged 65 plus by Region

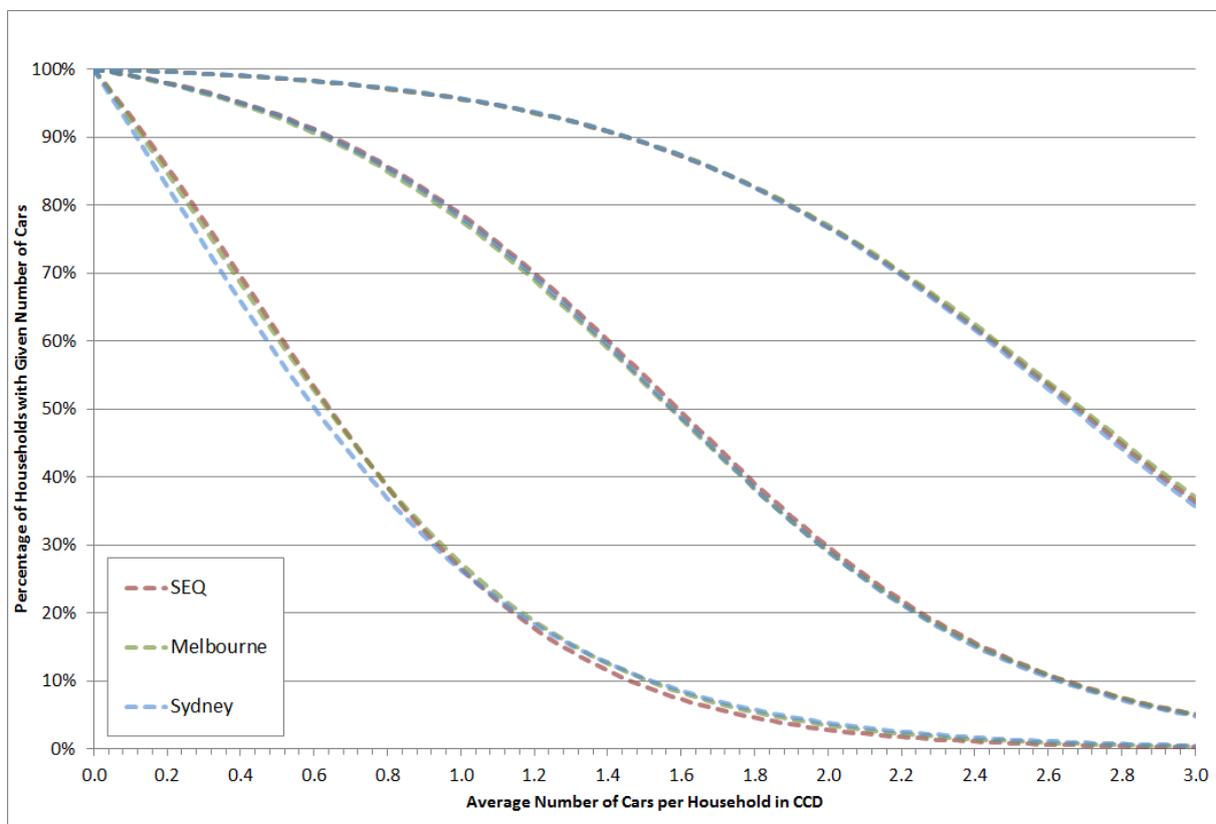


Figure 2.10: Segmentation by Car Ownership Level by Region