Zenith Model of Victoria

Technical Note 5
Household and Travel Market Segmentation

Zenith Version: 2.0.0

VEITCH LISTER CONSULTING PTY LTD

Date: 26-03-2013

Zenith Model of Victoria

Technical Note 5: Household and Travel Market Segmentation

Date	Revision	Prepared By	Checked By	Approved By	Description
23-02-2011	A	TV	MV	MV	Draft for comment
24-01-2012	B	TV	MV	MV	Final Report

Contents

1 Introduction

The Zenith travel model of Victoria is one of a family of models developed by Veitch Lister Consulting (VLC) for transport planning in Australian cities and regions.

This document is one in a series of technical notes that collectively describe the Zenith Model of Victoria.

1.1 Related Documents

This technical note is the fifth of eleven. The other working papers are:

- Working Paper 1: Model Validation Framework and Data Sources
- Working Paper 2: Review of VISTA07
- Working Paper 3: Home Based Trip Production Model
- Working Paper 4: Non-Home Based Trip Production Model
- Working Paper 5: Household Segmentation & Travel Market Segmentation Models
- Working Paper 6: Period Allocation and Vehicle Occupancy Models
- Working Paper 7: Mode Choice Model
- Working Paper 8: Destination Choice and Trip Attraction Model
- Working Paper 9: Overall Model Validation
- Working Paper 10: Backcasting and Sensitivity Testing
- Working Paper 11: Reference Case Model Assumptions

1.2 Scope of This Document

The primary focus of this document is segmentation models. The Zenith model employs two segmentation models:

- The Household Segmentation model
- The Travel Market Segmentation model

Section 2 of this paper will describe the household segmentation model, including its aims, data sources, methodology and model results.

Section 3 will then describe the same aspects of the travel market segmentation model.

2 Household Segmentation

2.1 Motivation

The aim of the Household Segmentation model is to segment households according to their level of an attribute (eg. 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 will be found to own 0, 1, 2, or 3+ cars?

The answer to this question cannot be inferred easily; there are in fact an infinite number of combinations of proportions which all achieve an average of 1.3.

One of them is:

- 70% 1 car
- 30% 2 car

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; it is likely, of course, that there are households with 0 cars, and 3+ cars.

In this case, the model we have developed would output:

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

The motivation for segmenting is simple: the Home Based Trip Production model (HBTPM) requires these distributions as input.

In the HBTPM (described in Working Paper 3), each household in the model is described according to its:

- 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 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 for 2006 from the 2006 ABS Census. However, defining these distributions for a range of land use scenarios for each and every future year would be onerous.

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 HBTPM, converting averages into realistic distributions.

2.2 Data Sources

The 2006 ABS Census has been used to calibrate and validate the Household Segmentation model.

A custom dataset was purchased from the ABS (by VLC) for the purposes of model recalibration. The dataset, which is described in Table 1 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.

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

Table 1 - The Custom 2006 ABS Census Variables used for Model Recalibration

2.3 Methodology

2.3.1 Household Attributes for which Segmentation Is Required

A unique Household Segmentation model is required for each of the six attributes which are input to the Home Based Trip Production model (HBTPM):

- 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 2 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 2 - 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 was not collected in the 2006 ABS Census.

2.3.2 Model 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 (along with most of the rest of the Zenith model structure) by 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 generalised graphical representation of the above relationship is presented in Figure 1, where the three calibration parameters dictate the shape of each curve. Broadly speaking, A is a vertical placement factor, B is a logit slope factor and C is a horizontal offset.

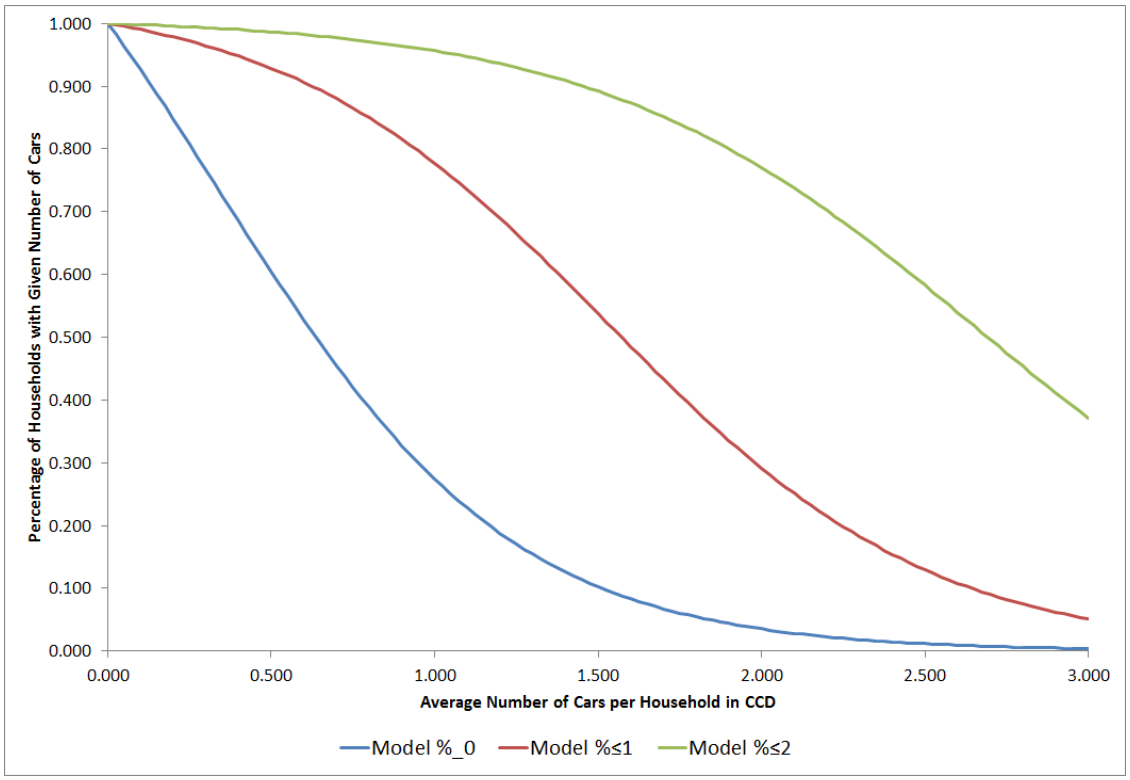


Figure 1 - Graphical Representation of a Family of Logistic Curves

A more revealing view of these curves is presented in Figure 2, which shows that the curves are, in fact, the boundaries between the segments. The grey region (below the first curve) represents households with no car, the blue region (between the first and second curves) represents households with one car, with the red and green regions representing households with 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 have no car. However, as the average car ownership increases, the proportion of households with no car decreases, with more households having higher car ownership.

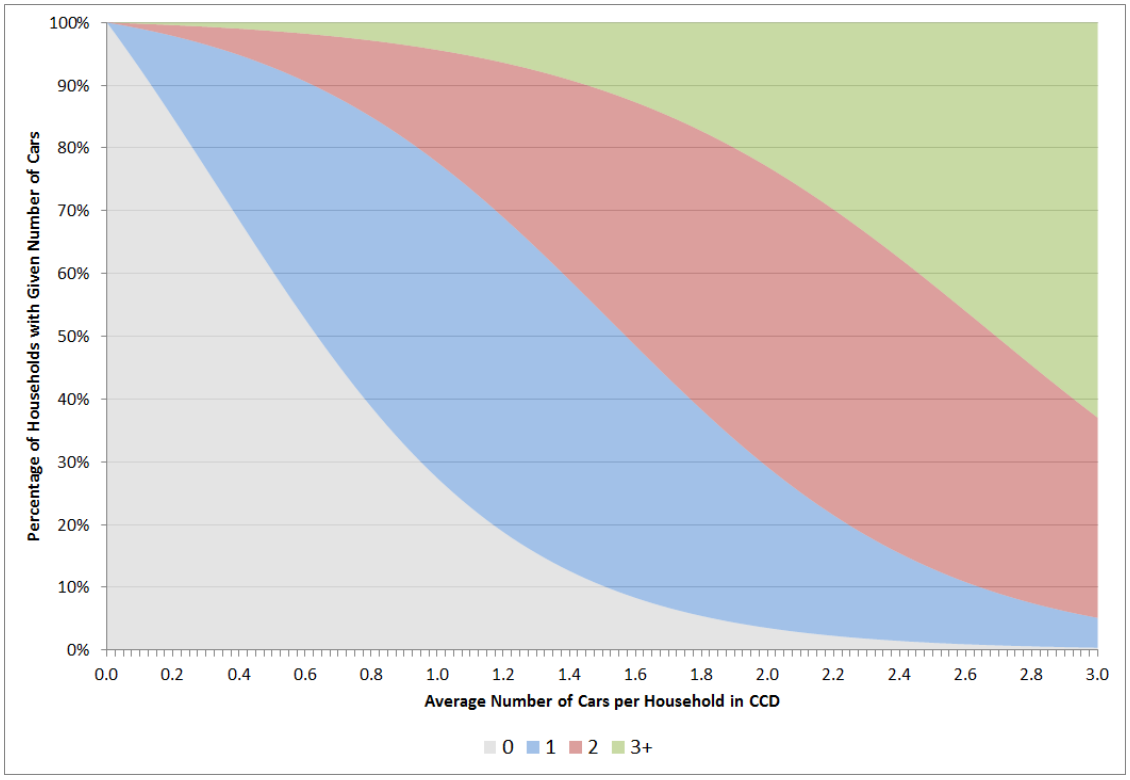


Figure 2 - Graphical Representation of Household Segments

2.4 Results

The Household Segmentation model has been recalibrated using 2006 ABS Census data, with the model parameters presented in Table 3 below. An R^2 is provided for each curve.

Attribute	Level	Model coefficients			
		A	*B*	*C*	R^2
Workers (White Collar)	≤ 0	-39.0199	0.7188	-0.2368	0.9439
	≤ 1	89.1454	0.6174	1.3711	0.9310
	≤ 2	99.3376	0.4709	2.3626	0.5556
Workers (Blue Collar)	≤ 0	-54.3561	0.6918	-0.3003	0.9574
	≤ 1	87.6109	0.8142	1.7004	0.7895
	≤ 2	97.6795	0.9956	3.7469	0.3196
Dependants (Aged 0-17)	≤ 0	-49.5424	1.0320	-0.4153	0.9132
	≤ 1	62.7388	0.9703	0.9579	0.9205
	≤ 2	89.0018	1.2344	2.7249	0.7114
Dependants (Aged 18-64)	≤ 0	-406780	1.0251	-8.5201	0.8957
	≤ 1	94.7138	0.4618	1.3578	0.7824
	≤ 2	99.3048	0.4179	2.0764	0.4898
Dependants (Aged 65+)	≤ 0	38.6725	0.5292	0.2588	0.9391
	≤ 1	62.8294	1.1513	1.1394	0.7334
Cars	≤ 0	53.1913	0.4484	0.3404	0.9606
	≤ 1	95.3751	0.4970	1.5277	0.9671
	≤ 2	98.9871	0.5837	2.6808	0.9121

Table 3 - Re-estimated Household Segmentation Parameters

The resulting segmentation curves, and the 2006 ABS Census data used to estimate the curves, are presented in Figure 3 through Figure 8 below. It is worthwhile describing what these plots are showing.

Referring to Figure 3 (white collar workers) as an example:

- Each CCD (in the Zenith model area) has an average number of white collar workers, which places it on the x-axis. Each CCD also has a breakdown of households according to their number of white collar workers (0, 1, 2 or 3+), which results in 3 dots (blue, red and green), which mark the segmentation of households into each of these categories
- The segmentation curves, which are shown as grey lines, have been "fitted" through the Census data points

- The segmentation curves act as boundaries between the segments; the grey region represents households with 0 white collar workers, while the blue, red and grey regions represent households with 1, 2 and 3+.
- As an example of how the model would be applied, if we take an average of 0.8 white collar workers per household, then the model would predict that roughly 45% of households would have 0 white collar workers, 35% would have 1 worker, 17% would have 2 workers, and the remaining 3% would have 3+.

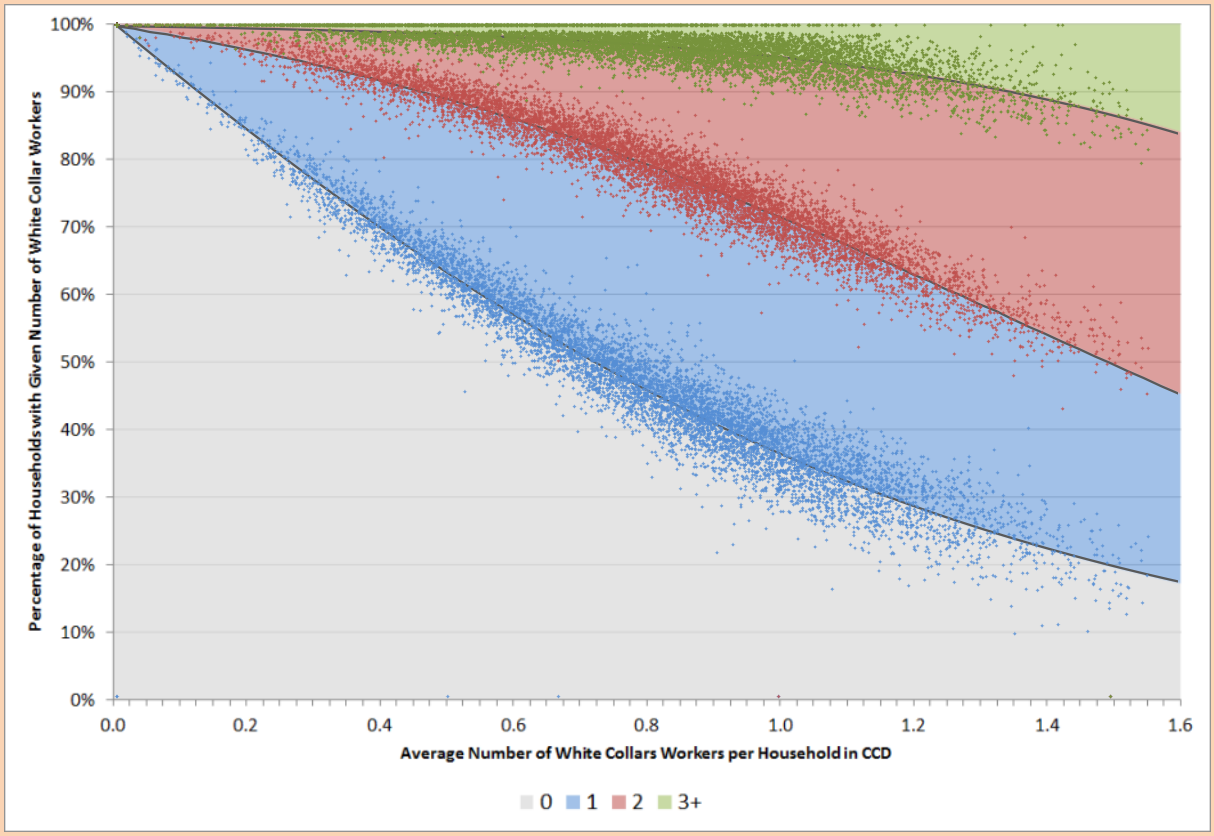


Figure 3 – Segmentation of Workers (white collar)

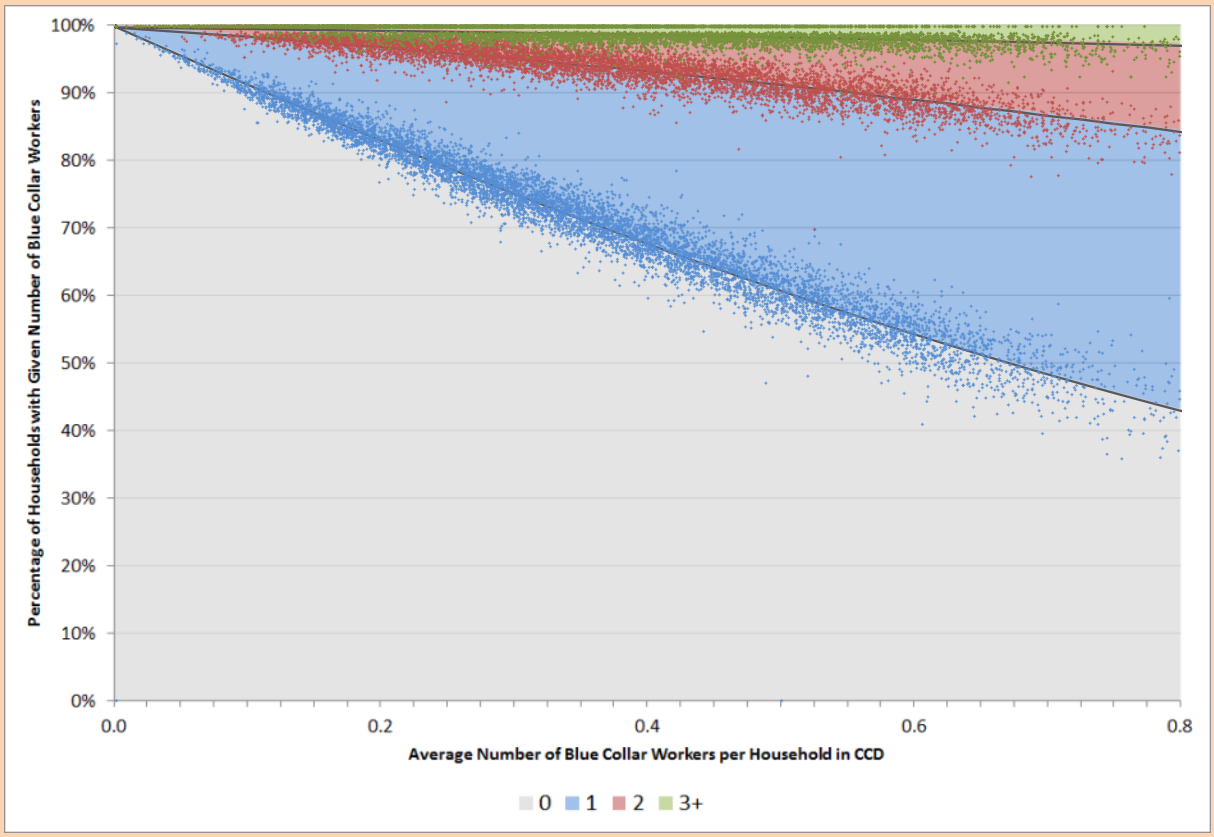


Figure 4 - Segmentation of Households by Number of Workers (blue collar)

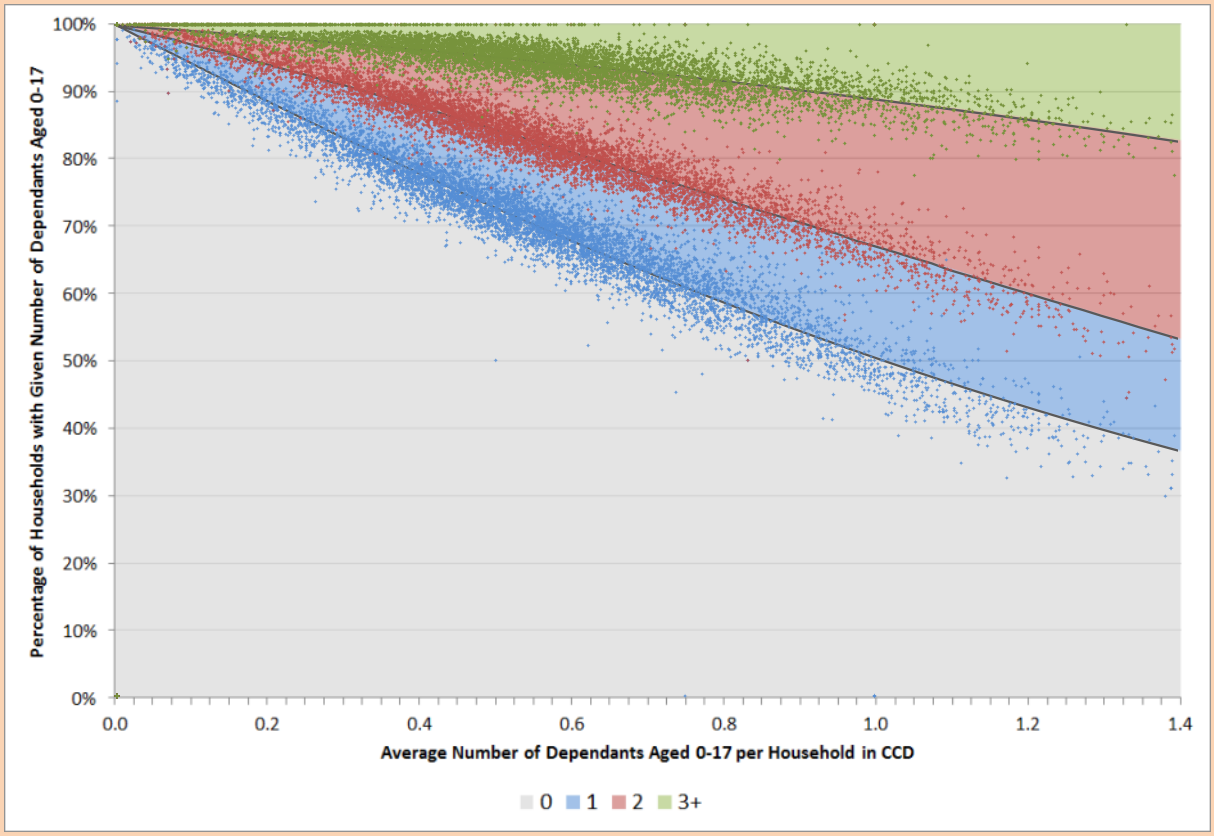


Figure 5 - Segmentation of Households by Number of Dependants (aged 0-17)

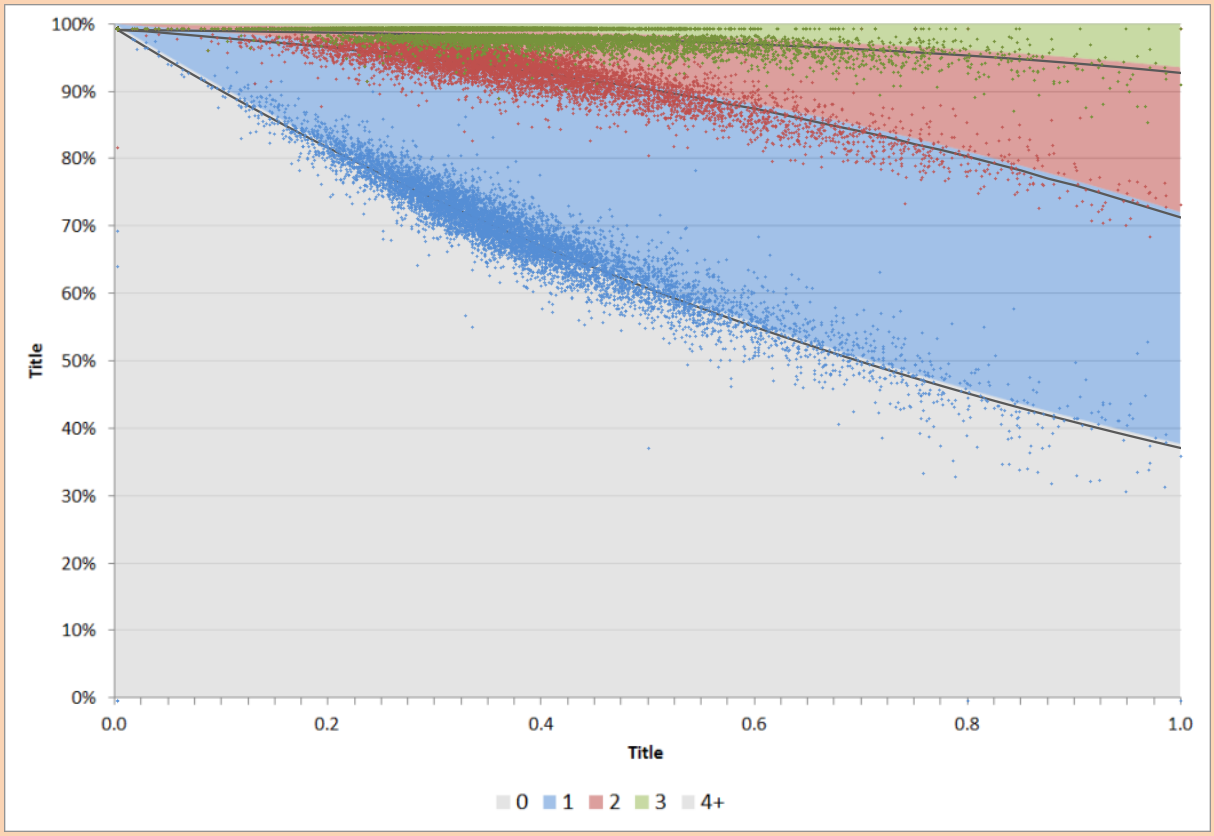


Figure 6 - Segmentation of Households by Number of Dependants (aged 18-64)

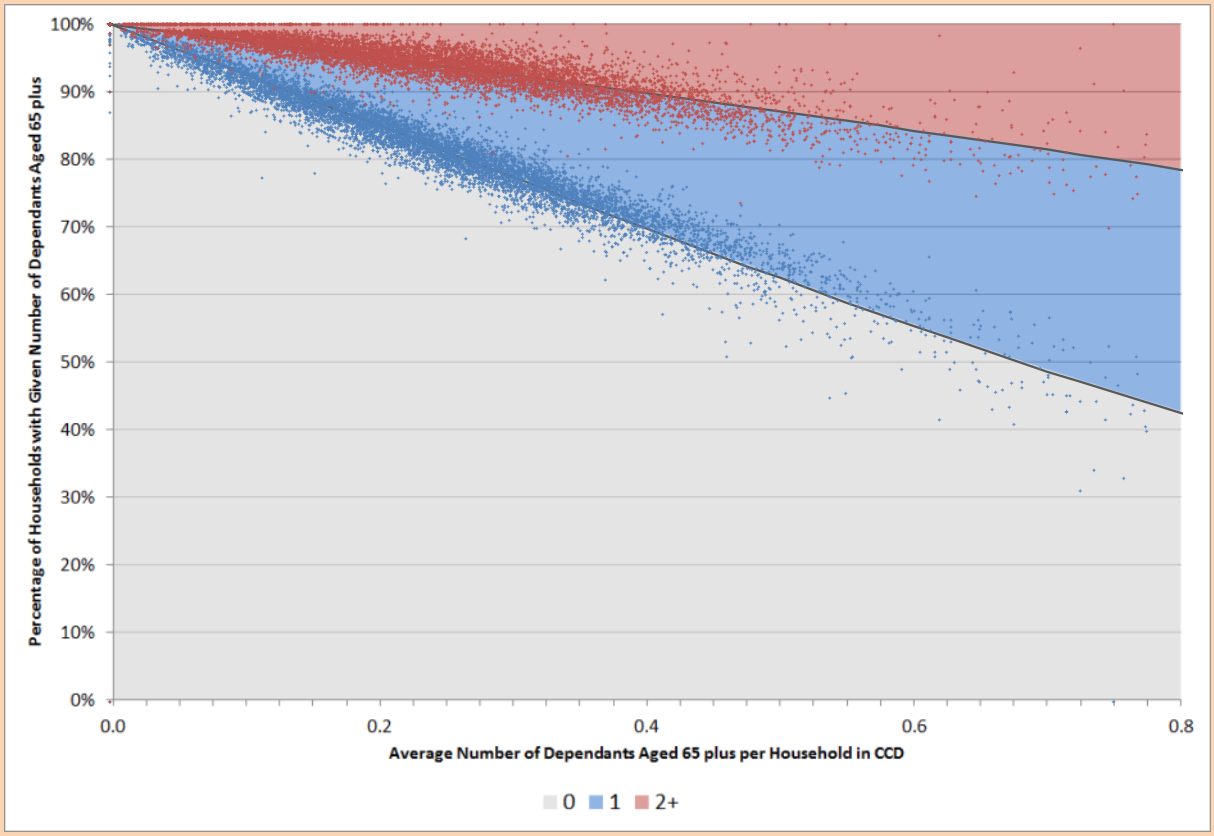


Figure 7 - Segmentation of Households by Number of Dependants (aged 65+)

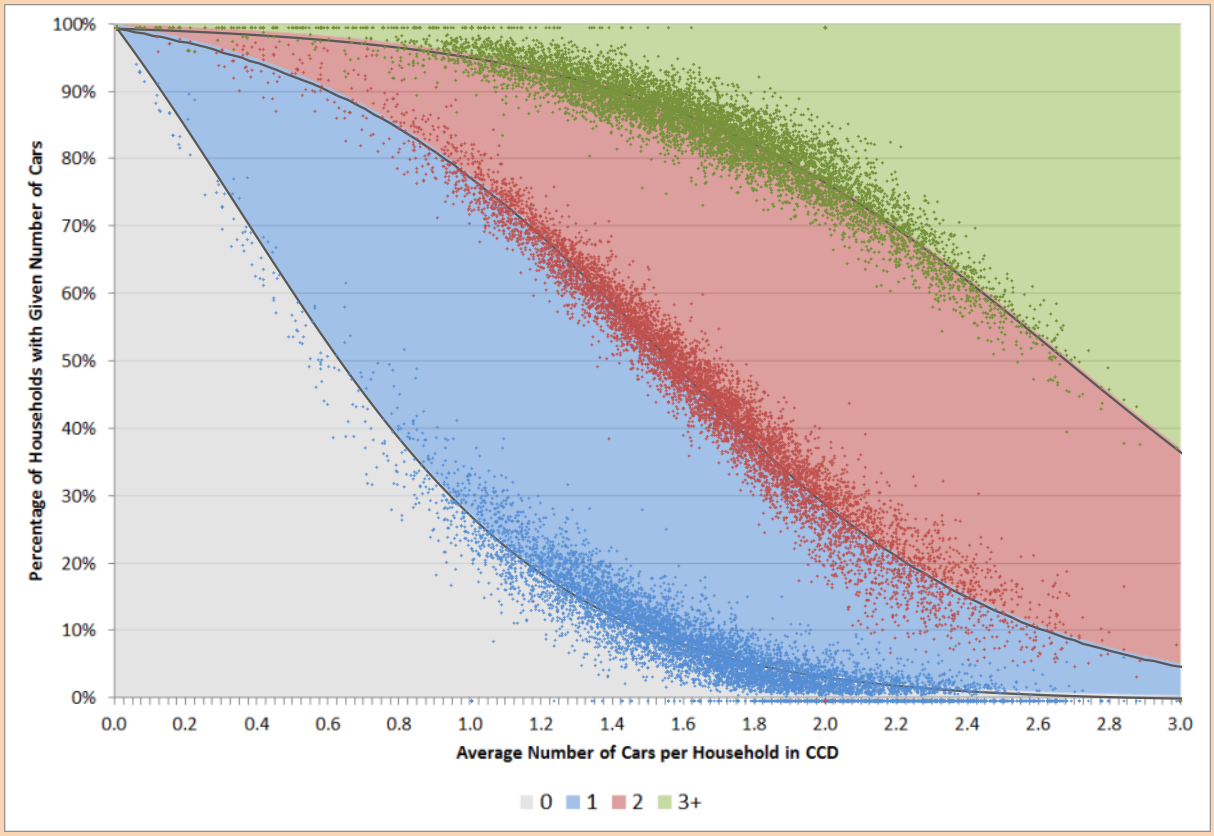


Figure 8 - Segmentation of Households by Number of Cars Owned

2.5 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 we have derived are spatially stable and transferrable (temporal stability might be explored at a later date by analysing historical Census results).

To illustrate this, we have estimated parameters for Sydney and South East Queensland, and compared the resulting segmentation curves. The comparisons are found in Figure 9 through Figure 14 below.

Referring to Figure 9, 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 can be seen by referring back to Figure 3.

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

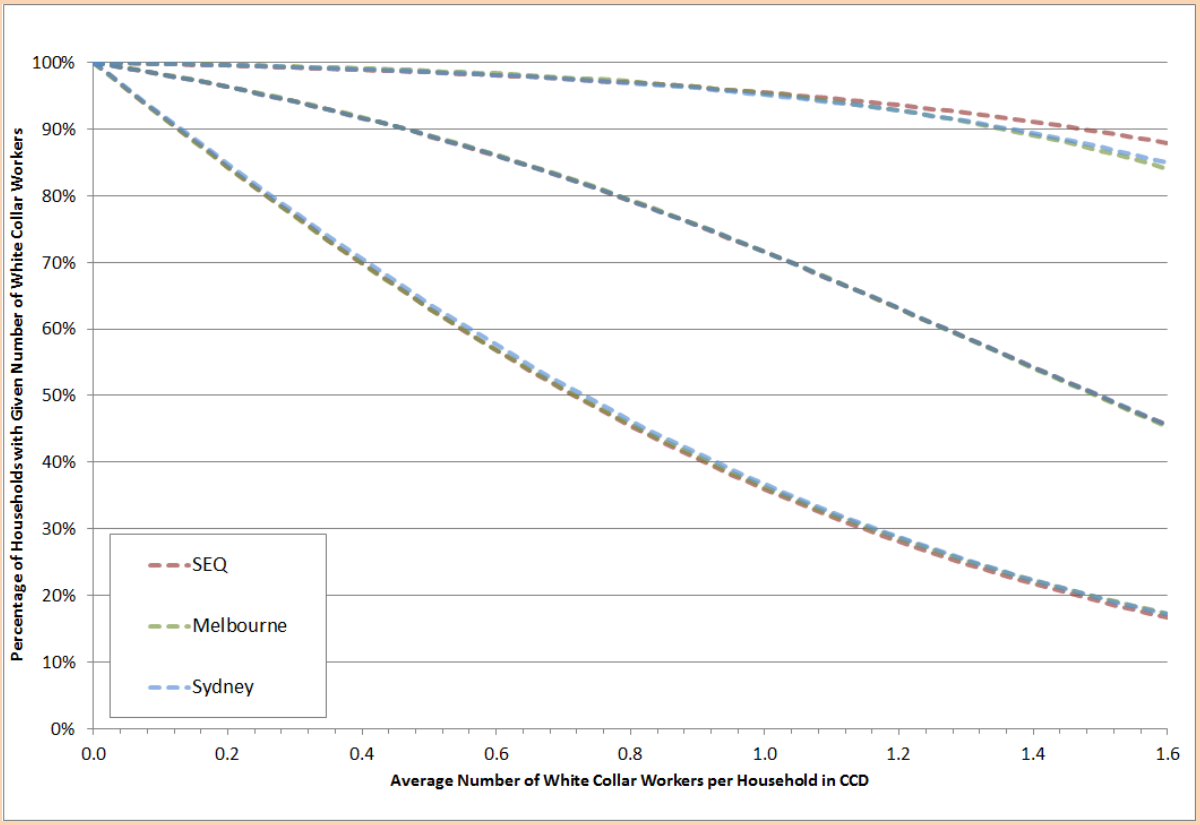


Figure 9 – Comparison of Segmentation of Workers (White Collar)

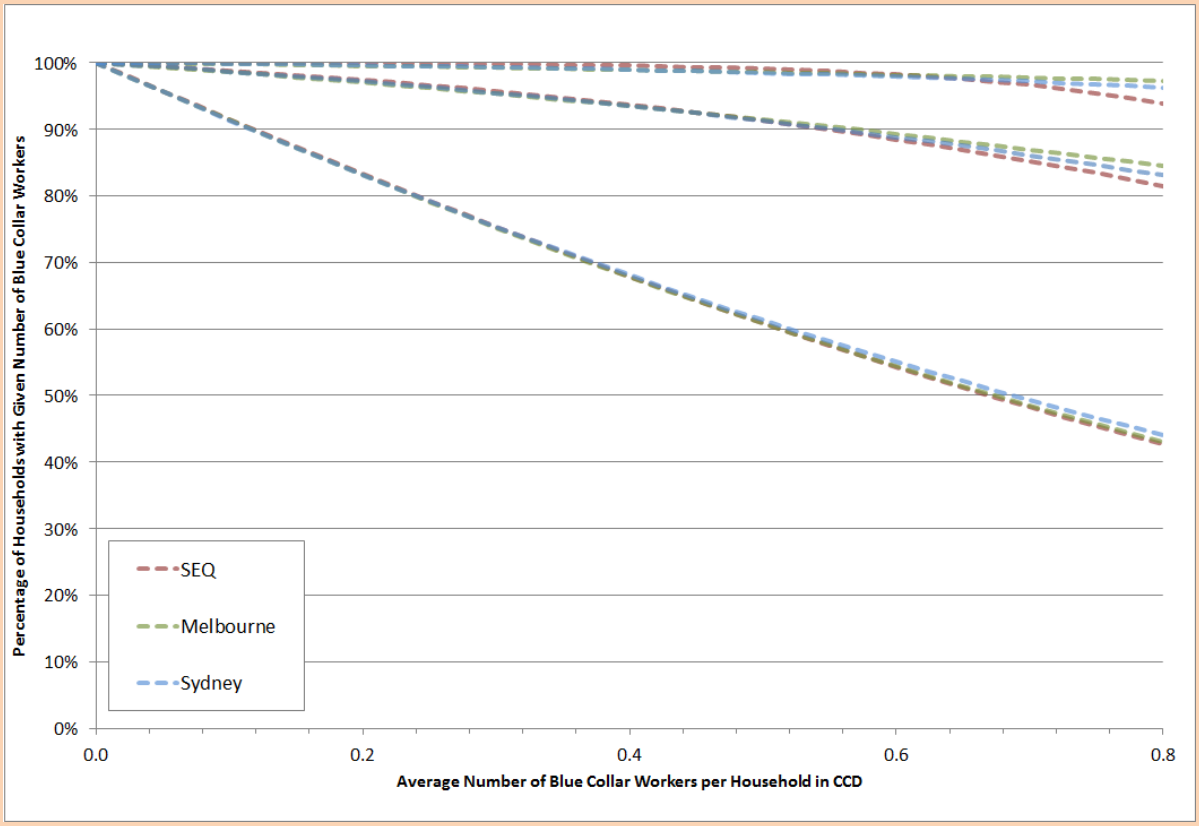


Figure 10 - Comparison of Segmentation of Workers (Blue Collar)

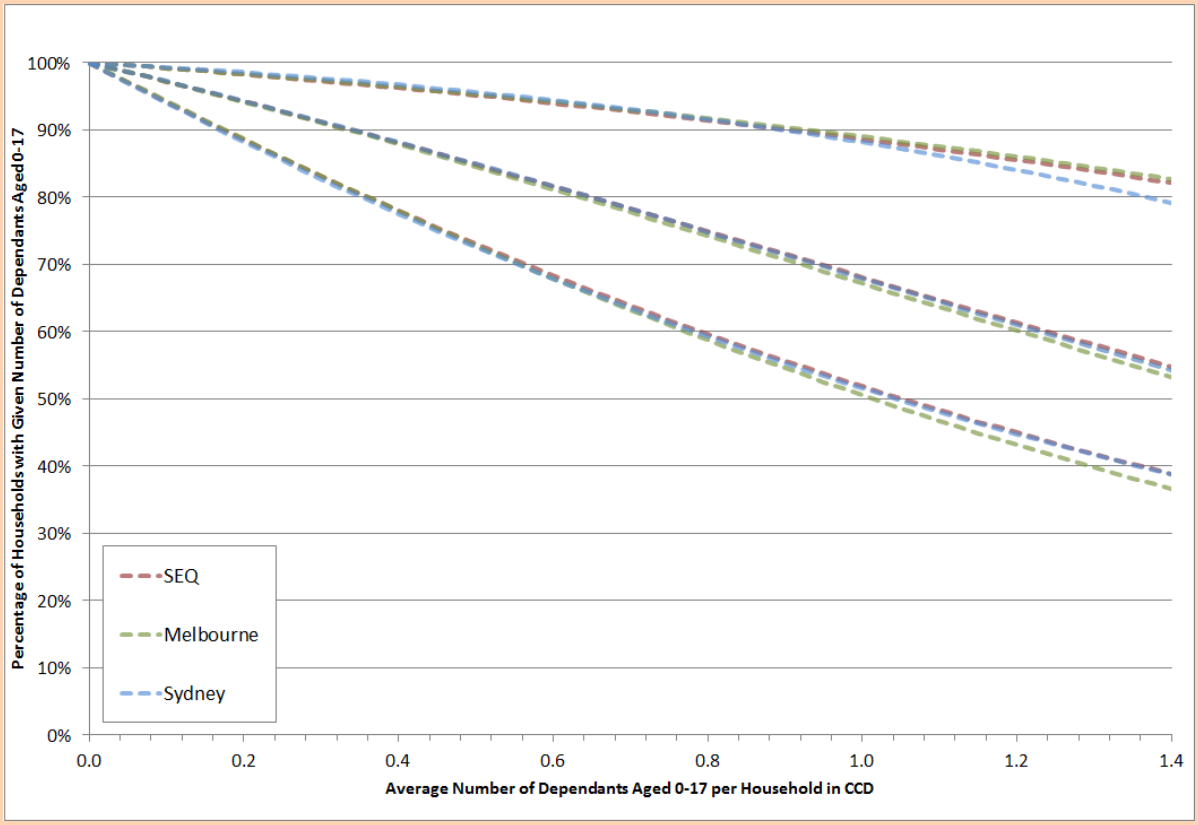


Figure 11 - Comparison of Segmentation of Dependants Aged 0-17

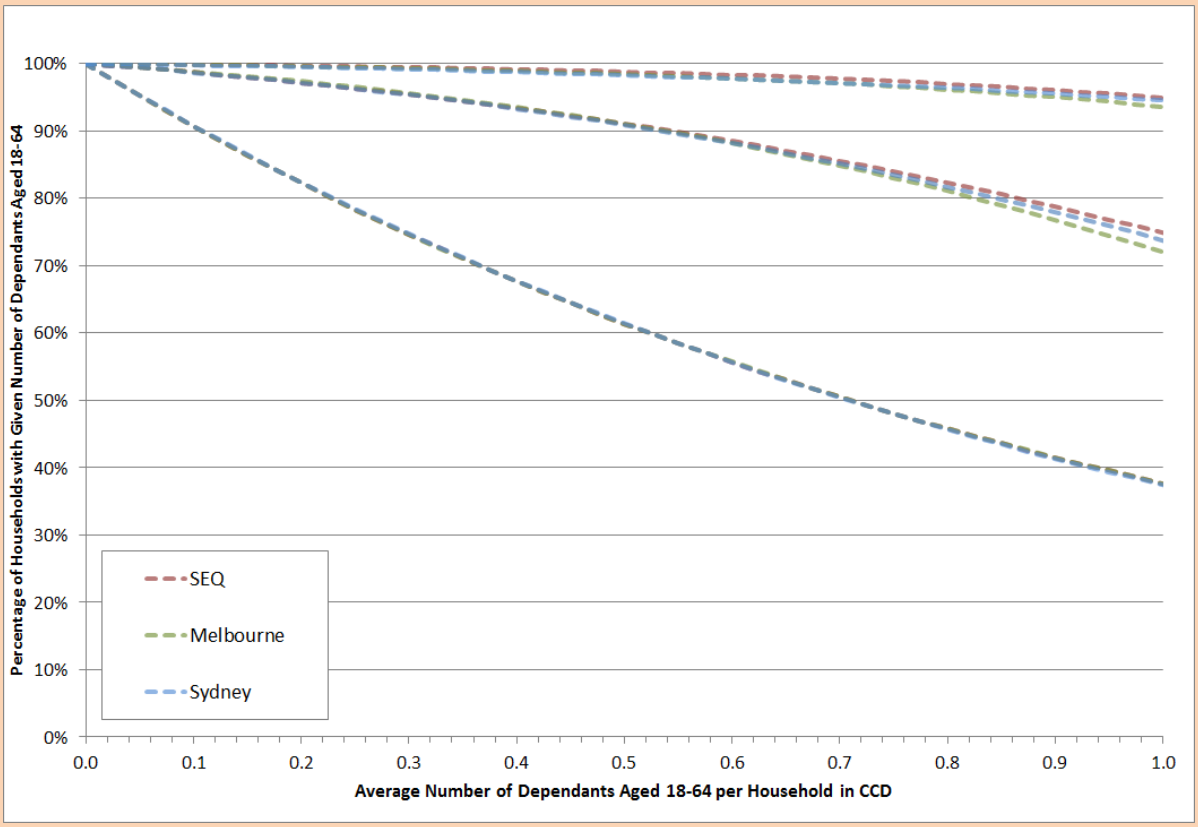


Figure 12 - Comparison of Segmentation of Dependants Aged 18-64

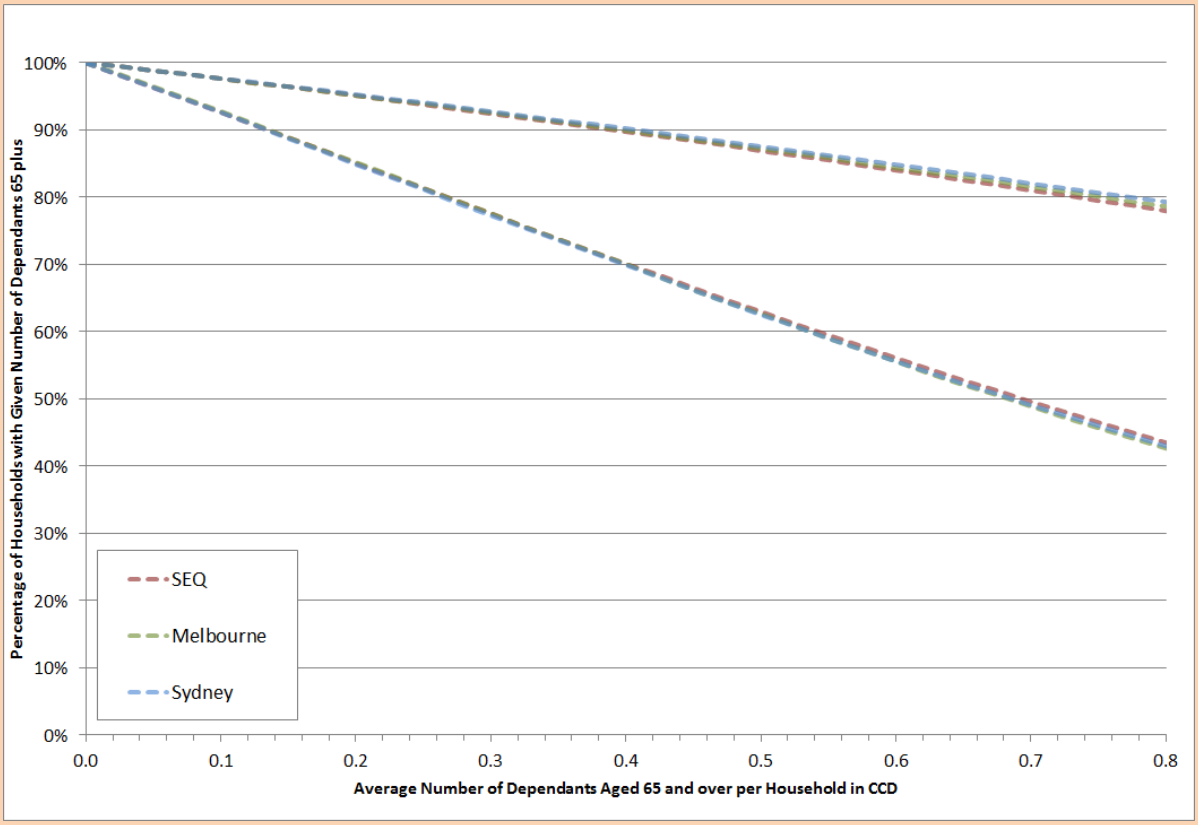


Figure 13 - Comparison of Segmentation of Dependants Aged 65plus

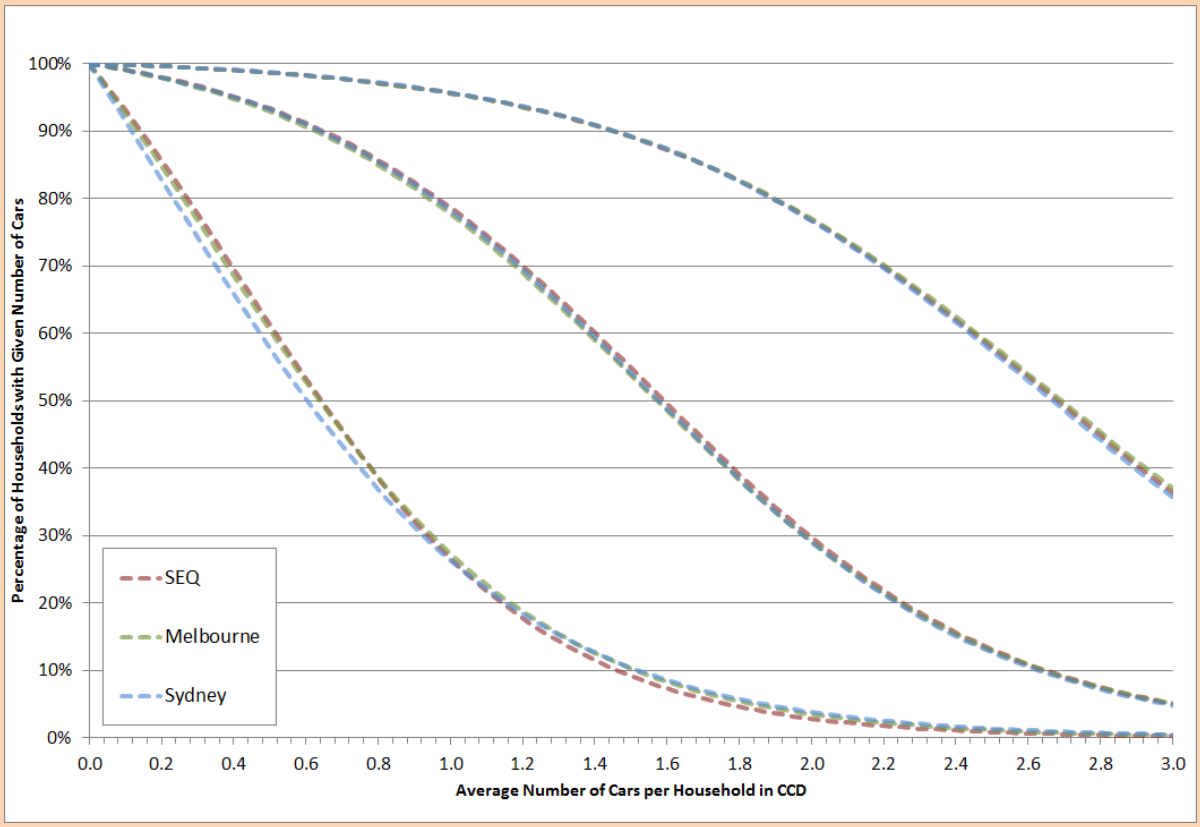


Figure 14 - Comparison of Segmentation of Car Ownership

3 Travel Market Segmentation

3.1 Motivation

The Household Segmentation model and the Home Based Trip Production model (as described in Section 2 of this paper, and Working Paper 3, respectively) enable estimates to be made of zonal trip productions for each of the home based trip purposes.

Zonal trip production and attraction estimates are subsequently linked between origin and destination zones by a Trip Distribution model, before being allocated to time periods, and split by mode.

Prior to submitting zonal trip production estimates to the trip distribution process, however, it has been found advantageous to further segment the home based trip productions based on the car ownership level of the household making each trip (ie. separate trips made by 0 car, 1 car, 2 car and 3+ car households). An example of this process is provided in the inset below.

It has been found during model development in Melbourne, Adelaide and South East Queensland that this further segmentation of the travel market, prior to trip distribution and mode choice, significantly increases the accuracy of the resulting models - by acknowledging that households with limited private motor vehicle access are also likely to display atypical destination and mode choice decision-making behaviour.

This further segmentation of home based trip productions is performed by the Travel Market Segmentation model.

Travel Market Segmentation Example:

Given that a travel zone has an average household car ownership of 1.3, what proportion of Home Based Work (white collar) trips produced in this zone will be made by households owning 0, 1, 2, or 3+ cars?

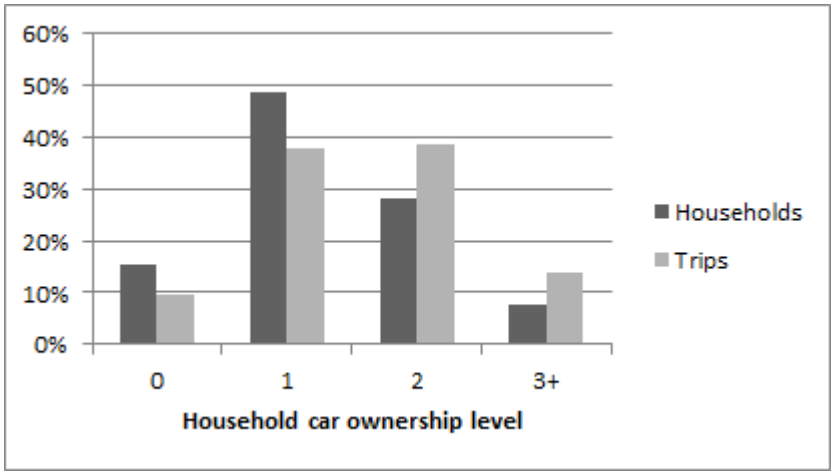
In this case, the model we have developed would output:

- 9.6% of trips would be made by 0 car households
- 37.9% of trips would be made by 1 car households
- 38.7% of trips would be made by 2 car households
- 13.8% of trips would be made by 3+ car households

Note that these percentages do not equal the breakdown of households according to their car ownership level which is listed below (source: Section 2.1):

- 15.5% 0 car
- 48.6% 1 car
- 28.3% 2 car
- 7.7% 3+ car

The distribution of households and trips by the level of household car ownership is presented below:



While nearly half of households own 1 car, 1 car households are responsible for only 38% of home based work (white collar) trips. More trips are made by 2 car households (39%), even though only 28% of households own 2 cars. Clearly, high car ownership is correlated with high trip making.

3.2 Data Sources

The data source used to calibrate the Travel Market Segmentation model is the Victorian Integrated Survey of Travel and Activity (VISTA07).

3.3 Methodology

3.3.1 Trip Purposes for Which Segmentation is Required

Travel market segmentation is separately performed for each of the seven home based trip purposes:

- Home based work (white collar)
- Home based work (blue collar)
- Home based education (secondary)
- Home based education (tertiary)
- Home based shopping
- Home based recreation
- Home based other

Ideally, we would segment non-home based trips also. The difficulty, however, is that unlike home based purposes, we don't have a simple rational basis for doing so. In the case of home based trips, we know home location of the trip maker (as it is the production end of the trip), and can make use of the average household car ownership of that location as a means of segmenting trips.

Incorporating car availability into the non-home based purposes is something which we might explore at a later date (or better yet, we could switch to modelling tours).

3.3.2 Model Form

The functional form of the Travel Market Segmentation model is presented below. Note that this functional form is identical to that of the Household Segmentation model (Section 2.3.2).

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

Where: $T_n(x)$ is the percentage of trips which are made by households owning n cars or less

x is the average car ownership of the travel zone

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

3.3.3 Estimation Procedures

A calibration dataset was formed by repeatedly drawing sub-samples from the VISTA07 survey, and for each sub-sample, calculating the average car ownership, as well as the proportion of trips made by 0, 1, 2 and 3+ car households for each trip purpose. Sub-samples of 200 households were drawn, with replacement.

To ensure that our sub-samples exhibited a fair cross section of average household car ownership levels, we stratified the VISTA sample of households by car ownership level, then drew unequally from the strata to achieve a desired level of average household car ownership for each sub-sample.

The Household Segmentation model for car ownership was used to define the number of households to be drawn from each strata to achieve a given average.

For example, to produce a data point with an average car ownership level of 1.3, we would randomly select:

- 31 households with no car (15.5%)
- 97 households with 1 car (48.5%)
- 57 households with 2 cars (28.3%)
- 15 households with 3+ cars (7.7%)

These proportions are drawn from the example in Section 2.1.

3.4 Results

The Travel Market Segmentation model has been recalibrated, with the resulting model parameters presented in Table 4 below.

The resulting segmentation curves, and the survey data upon which the model was fitted, are then presented in Figure 15 through Figure 21.

For each survey data point (which represents a sub-sample of 200 households from VISTA07), we have an average household car ownership level, which places it on the x-axis. We then have the breakdown of trips by 0, 1, 2 and 3+ car households for the sub-sample, which results in three dots (blue, red and green), which mark the boundaries between the segments.

The segmentation curves themselves are plotted as grey lines which pass through the survey data, and divide the market into our four segments; the grey, blue, red and green regions representing trips made by 0, 1, 2, and 3+ car owning households respectively.

Referring to the Figures, it can be observed that the strength of the relationship tends to vary by trip purpose. In the non-education purposes, the fit is very good, with the survey data nicely hugging our segmentation curves. However, in the case of the educational purposes (particularly tertiary), the fit is weaker. The quality of the fit is most likely related to the strength of the (correlative) relationship between car ownership and trip making. It seems that car ownership is not highly correlated with making tertiary education trips. It is also likely, however, that household car ownership will play a lesser role in the destination and modal decisions of education trips, relative to other trip purposes, making this less of an issue.

Overall, however, the Travel Market Segmentation model performs very satisfactorily.

Trip Type	Level	Model coefficients			
		A	*B*	*C*	R^2
Home-based Work (White Collar)	≤ 0	49.4057	0.3856	0.2627	0.9304
	≤ 1	90.3855	0.4979	1.1661	0.9435
	≤ 2	96.9863	0.6965	2.4392	0.8916
Home-based Work (Blue Collar)	≤ 0	36.7549	0.3896	0.1785	0.8081
	≤ 1	88.6390	0.4849	1.0546	0.8546
	≤ 2	95.8269	0.7285	2.3142	0.7330
Home-based Education (Secondary)	≤ 0	-67.2863	0.3902	-0.2008	0.6511
	≤ 1	79.6505	0.5733	0.9127	0.6962
	≤ 2	98.4449	0.7576	3.1545	0.3988
Home-based Education (Tertiary)	≤ 0	78.7245	0.3944	0.6104	0.5796
	≤ 1	95.4307	0.4273	1.3187	0.6173
	≤ 2	96.7090	0.6093	2.0800	0.4643
Home-based Shopping	≤ 0	42.7477	0.4451	0.2482	0.9461
	≤ 1	94.4171	0.5038	1.4536	0.9421
	≤ 2	98.6373	0.6043	2.5957	0.8807
Home-based Recreation	≤ 0	-31.6952	0.4268	-0.1175	0.8970
	≤ 1	90.9263	0.5289	1.2692	0.8899
	≤ 2	97.9403	0.6784	2.6341	0.7663
Home-based Other	≤ 0	-1601800000	0.3990	-6.6199	0.7858
	≤ 1	84.7453	0.5609	1.0546	0.8651
	≤ 2	97.5810	0.7348	2.7348	0.7331

Table 4 - Coefficients of Travel Market Segmentation Models

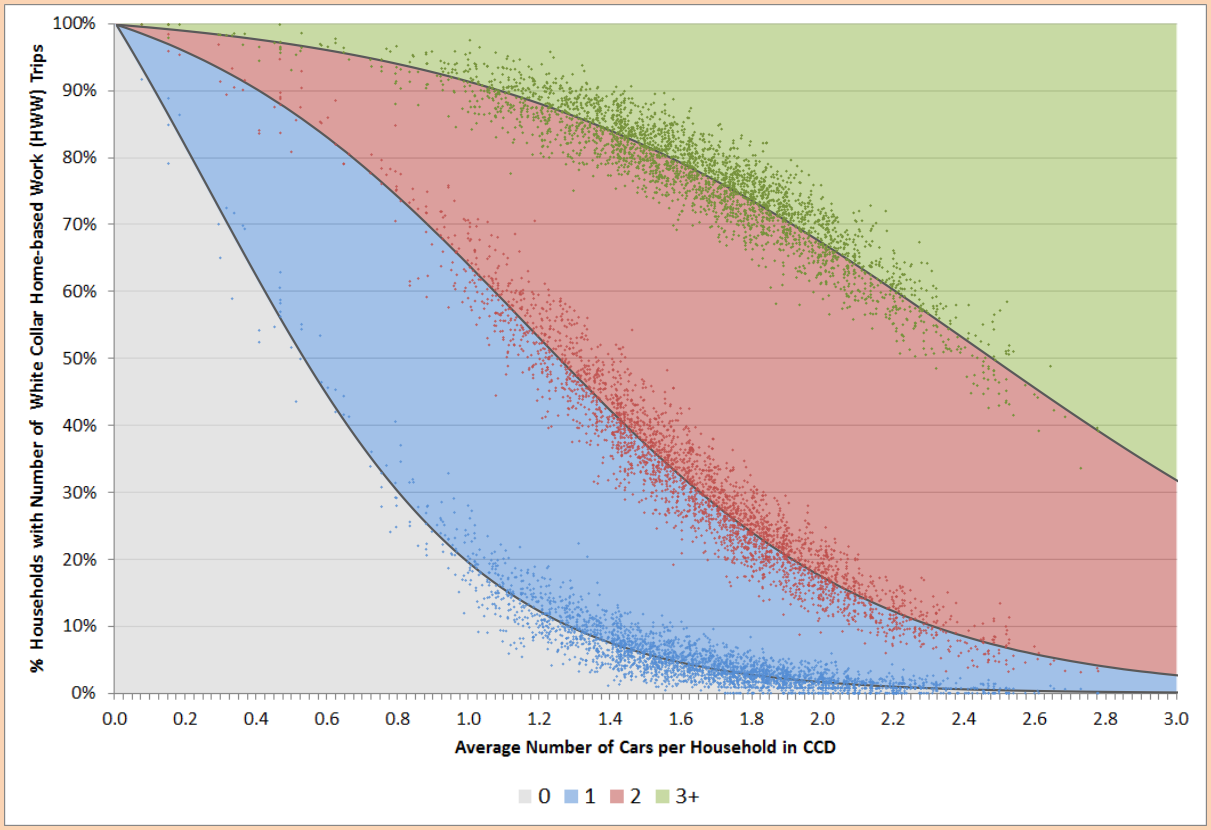


Figure 15 – Segmentation of Home Based Work (White Collar) Trips

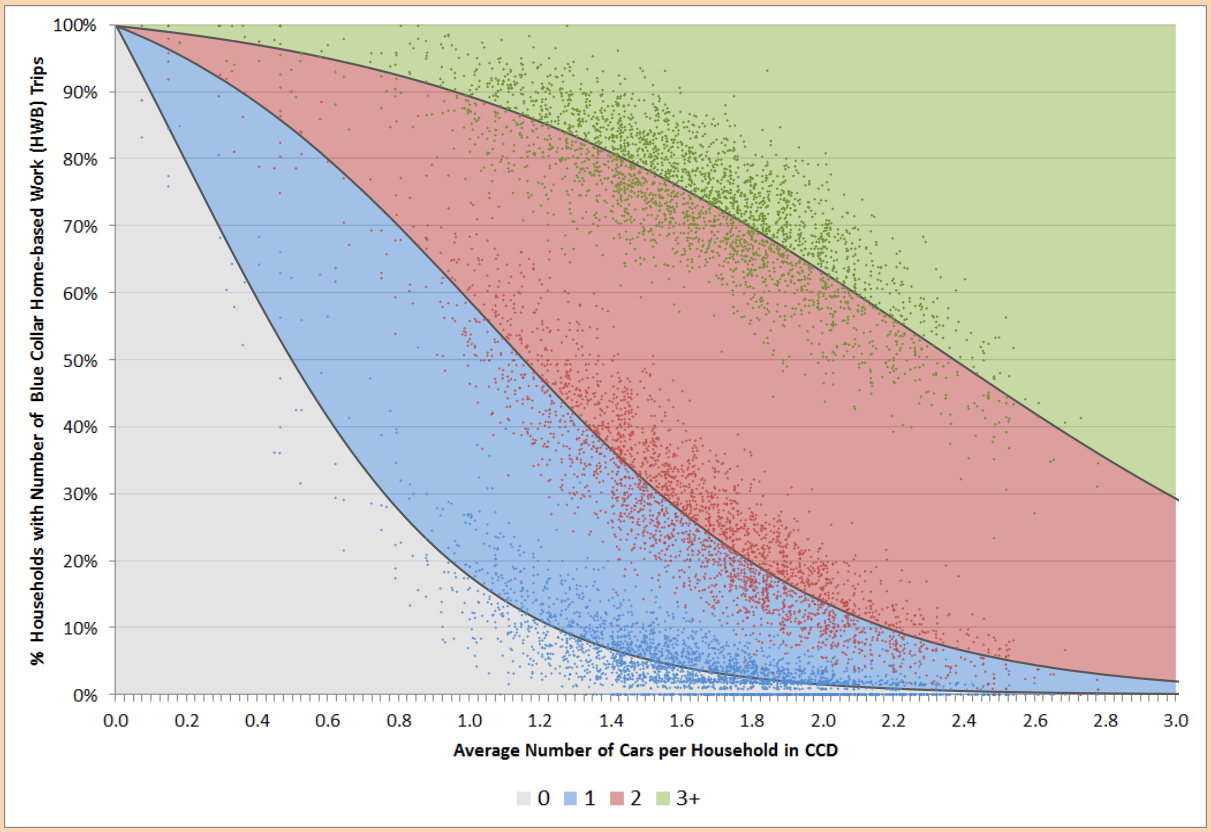


Figure 16 - Segmentation of Home Based Work (Blue Collar) Trips

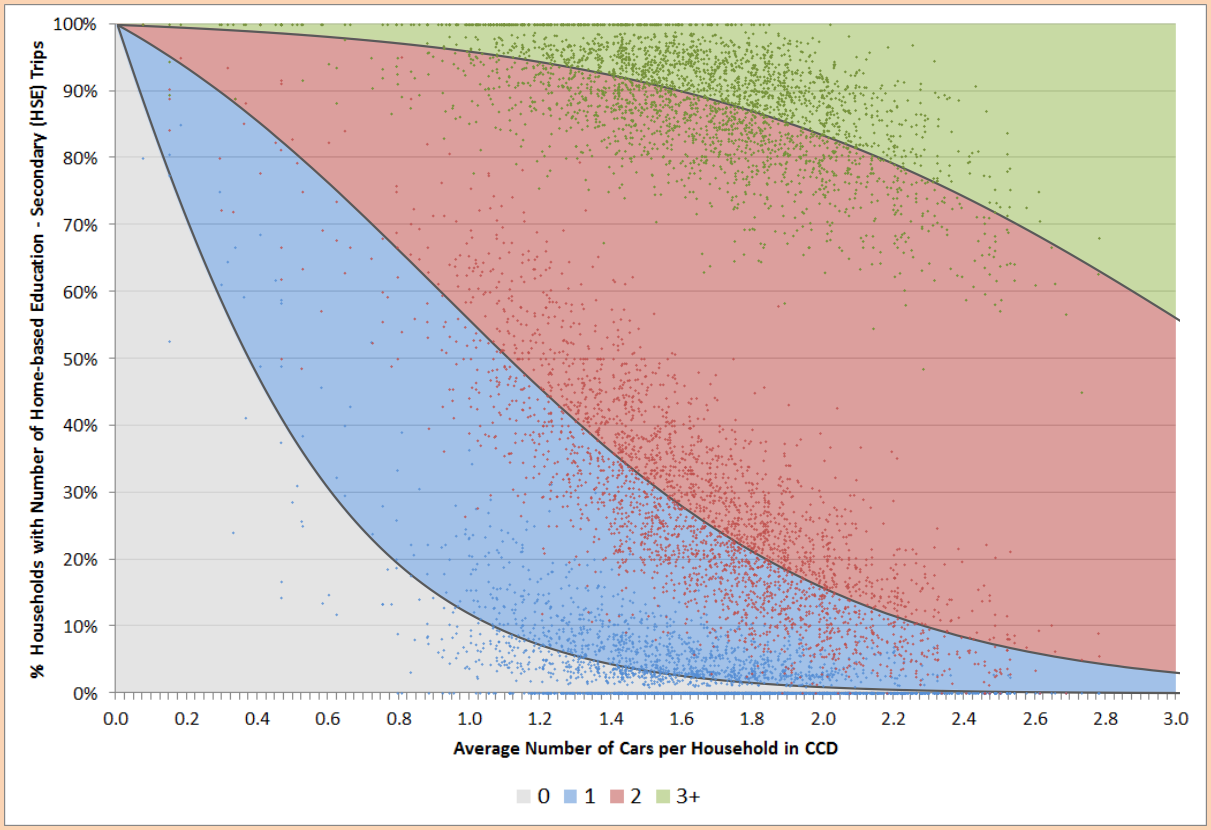


Figure 17 - Segmentation of Home Based Education (Secondary) Trips

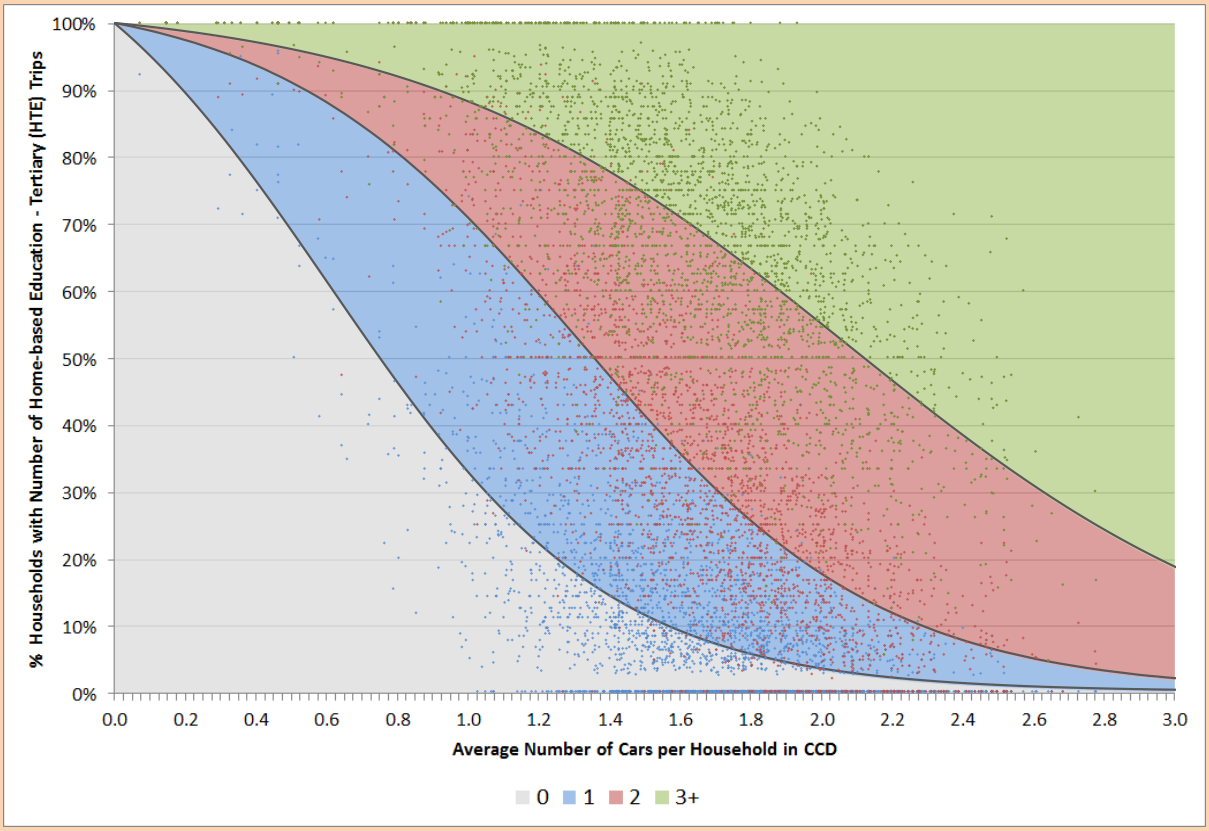


Figure 18 - Segmentation of Home Based Education (Tertiary) Trips

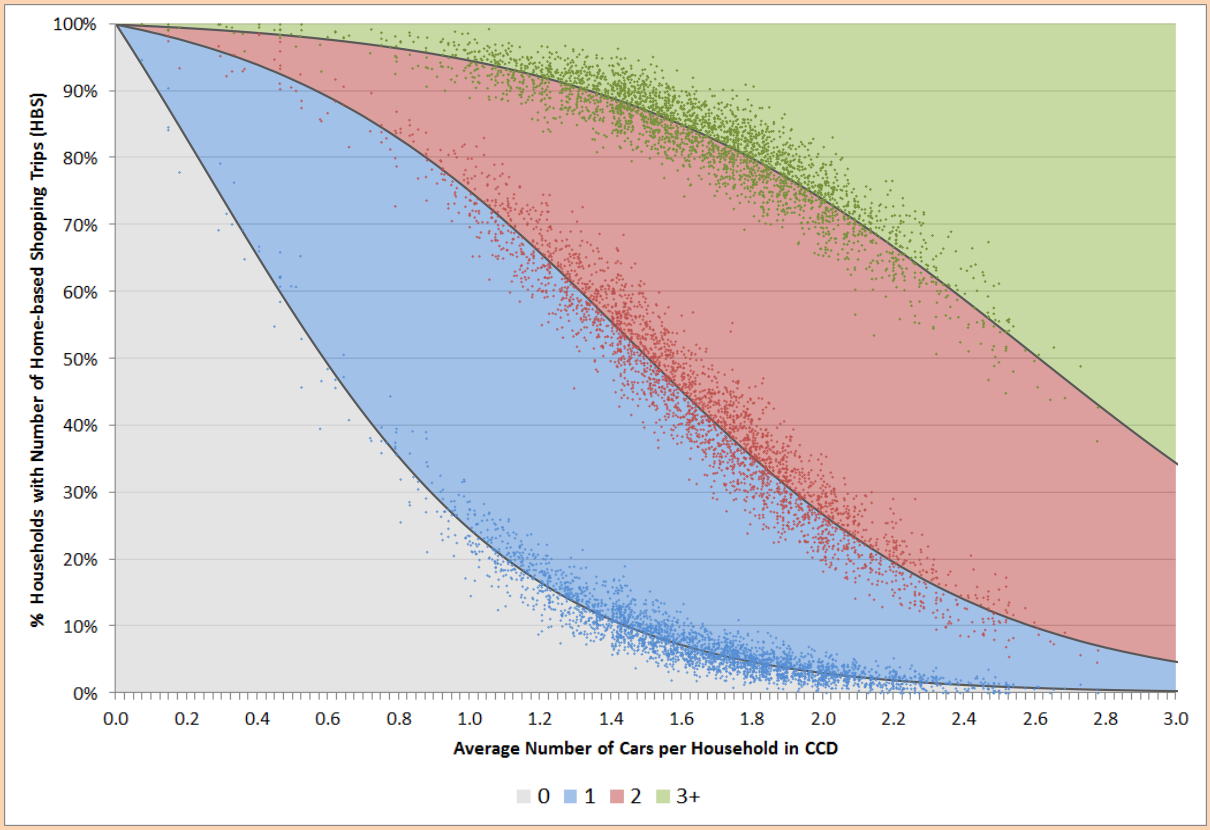


Figure 19 - Segmentation of Home Based Shopping Trips

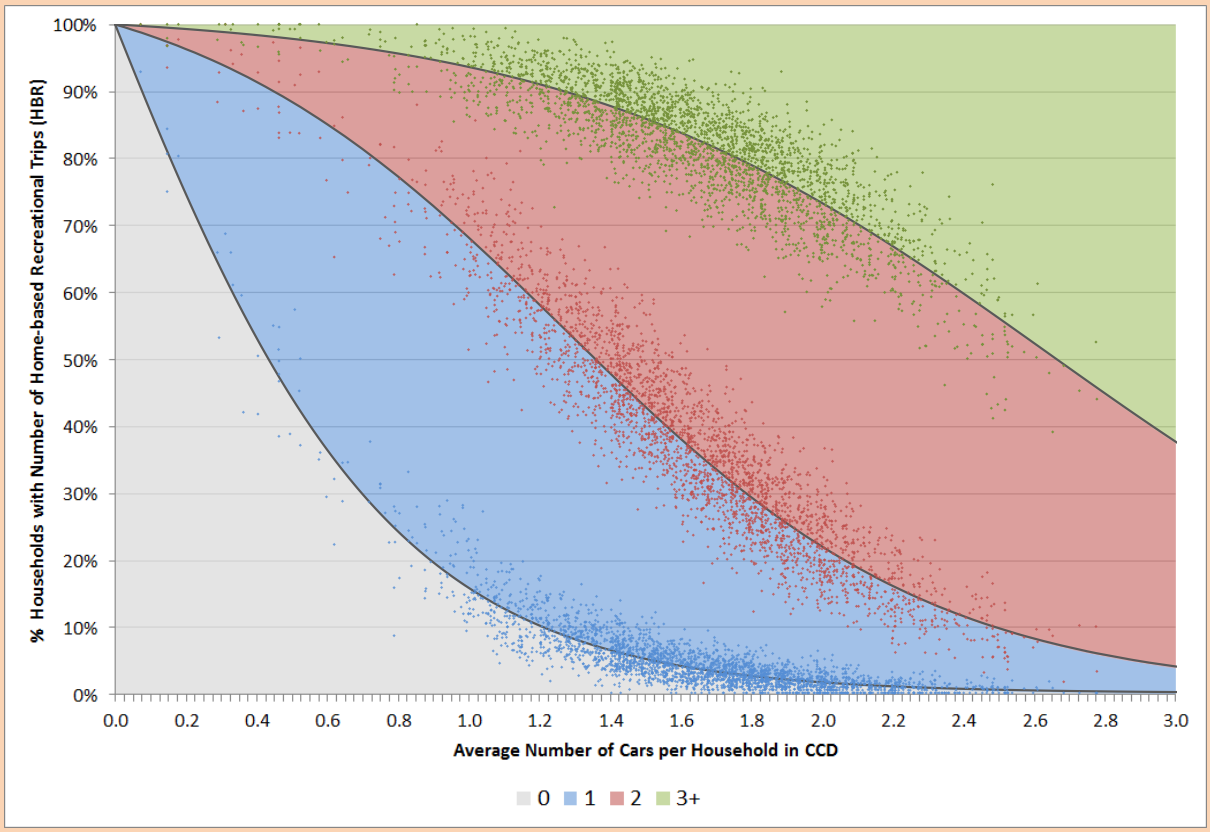


Figure 20 - Segmentation of Home Based Recreational Trips

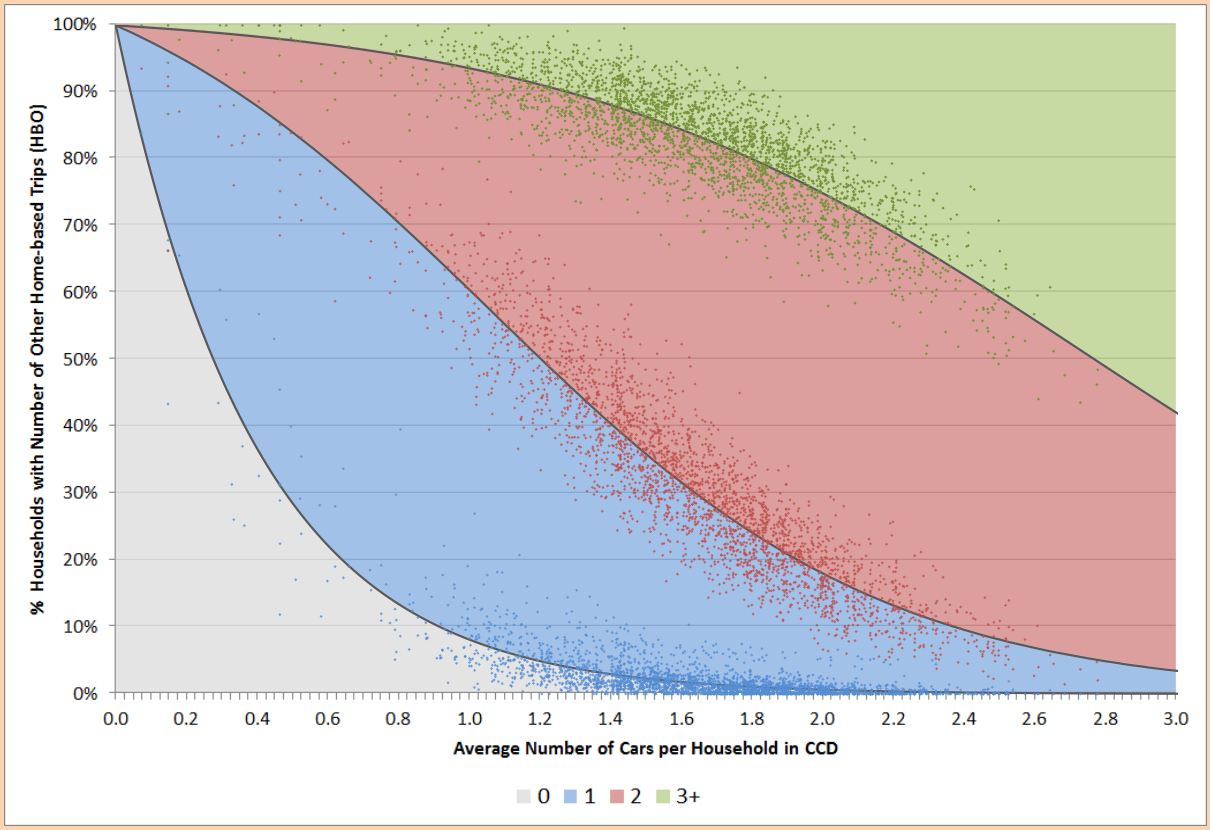


Figure 21 - Segmentation of Home Based Other Trips

4 Conclusions

The aim of both the Household and Travel Market Segmentation models is to construct robust distributions based on some average attribute value. In the case of Household Segmentation, we segment households according to 7 household variables, based on the average value of the variable:

- Number of White Collar workers (levels 0, 1, 2, 3+)
- Number of Blue Collar workers (levels 0, 1, 2, 3+)
- Number of dependants aged 0-17 (levels 0, 1, 2, 3+)
- Number of dependants aged 18-64 (levels 0, 1, 2, 3+)
- Number of dependants aged 65+ (levels 0, 1, 2+)
- Number of cars owned (levels 0, 1, 2, 3+)

In the case of Travel Market Segmentation, we segment *trips by the number of cars owned by the household making each trip.* This segmentation is performed on the basis of average household car ownership.

Both models have been estimated and validated using available data (Census and VISTA07), and implemented in Version 2.0.0 of the Zenith model.