Australasian Transport Research Forum 2013 Proceedings
2 - 4 October 2013, Brisbane, Australia
Publication website: http://www.patrec.org/atrf.aspx

Population Synthesis for Travel Demand Forecasting

Poh Ping Lim[1], David Gargett[1]

[1]Bureau of Infrastructure, Transport and Regional Economics, Department of Infrastructure and Transport, GPO Box 501, Canberra ACT 2601, Australia

Email for correspondence: pohping.lim@infrastructure.gov.au

Abstract

Population synthesis techniques are commonly used as viable alternatives to supplement the lack of availability and completeness of microdata collection for microsimulation modelling. The construction of a synthetic population set out in this paper is part of a broader research project that aims to develop an activity-based microsimulation model for travel demand forecasting, specifically for Australian capital cities. This paper describes the process of generating a synthetic baseline population for Sydney Greater Metropolitan (GMA) using 2006 Population Census. Microdata were created for households and individuals in Sydney GMA at Travel Zone (TZ) level and Censes Collection District (CCD) level. These synthetic data were benchmarked against aggregated census data to evaluate its representativeness.

1. Introduction

Activity-based microsimulation models for travel demand explicitly recognise that individuals and households are the actual decision makers, and that travel demand is derived from travellers' desire to participate in spatially dispersed activities (Axhausen and Gärling, 1992). It examines who, when, where, and how individuals travel by taking into account given characteristics of the households and their potential opportunities, the transport networks and various institutional constraints. The micro level decision making process in these models offer the capability to capture household interactions and their interdependencies that lead to joint activities and interrelated travel patterns. The modelling process hence provides a more realistic representation of behaviour in travel demand.

Given the disaggregate nature and conceptual advantages over the traditional four steps travel model, the microsimulation activity based approach can potentially offer a more responsive analytical tool for evaluating a wide variety of transportation policy initiatives. These models have been used for urban and transportation planning in many cities overseas, particularly in the United States. However, activity based microsimulation models have yet to be explored in much depth in Australia.

The construction of a synthetic population set out in this paper is part of a broader research project that aims to develop an activity-based microsimulation model for travel demand forecasting, specifically for Australian capital cities. An activity-based microsimulation model simulates activity-travel patterns for each and every individual or household of a study area. The nature of the simulation is such that the simulated pattern reflects the activity and network distribution of each individual and household, while the demographic structure remains to be true to the actual population in the study area.

The fundamental data requirement to operate a microsimulation model is to obtain details of individual or household attributes for the entire population of the study area. Information at this level of details is usually collected in a population census but has not been made available in the public domain due to privacy and confidentiality reasons. Population synthesis techniques are normally used as viable alternatives to supplement the lack of availability and completeness of microdata collection for microsimulation modelling.

This paper describes in details the process of developing a synthetic population base for Sydney Greater Metropolitan (GMA) using 2006 Population Census data. Microdata were created for households and individuals in Sydney GMA at Travel Zone (TZ) and Census

Collection District (CCD) level uses the Iterative Proportional Updating (IPU) algorithm. These synthetic data were benchmarked against aggregated census data to evaluate its representativeness. The preliminary validation results are included in this paper.

2. Population synthesis procedures

2.1 Synthesis procedure

The process of population synthesis generally involves expanding a sample drawn from a population to a full set of synthetic population, such that the generated synthetic population conforms as much as possible to the actual population at various aggregation levels. Population synthesis procedures are algorithms that applied to a sample data and its aggregated population data in order to grow a synthetic population that is statistically representative of the actual population data. The generated synthetic population basically represents one possible set of “best estimates” of the actual population, given the input information (Ryan et al, 2009).

Different algorithms used in the process may affect the quality of a synthetic population differently.

In this study, the synthetic population was created using **Synthetic reconstruction** method. Generally, this method involves two principal stages: fitting and generation (Bowman, 2009).

The fitting stage fits a disaggregated sample data to aggregated subtotals (also referred as marginal totals) in the population data. Most of the synthetic reconstruction methods are based on the Iterative Proportional Fitting (IPF) procedure. The IPF procedure begins by selecting a relevant set of attribute variables (also referred as control variables) from a population. The procedure uses the disaggregated sample data as “seeds” to create observations by estimating multi- way joint distributions, such that the generated distributions are collectively consistent with the cross tabulations provided by the actual population data.

The generation stage expands or grows the disaggregated sample data to a full population of the study area using joint distribution probabilities produced in the fitting stage.

Generally, the IPF procedure retains two important properties. Firstly, the number of observations created synthetically in a given category of the selected control variables matches the corresponding aggregated subtotals in the population data. Secondly, the correlation structure of the seed is retained (Müller and Axhausen, 2011). This procedure, with some variations, has been used in many integrated transport models that generate a synthetic population. The IPF technique is well documented in recent literature such as Beckman et al (1996), Frick and Axhausen (2004), Pritchard and Miller (2009); Srinivasan and Ma (2009).

However, the traditional IPF procedure can only match the observed and simulated control variables either based on household joint-distributions or individual joint-distributions. There is no mechanism to ensure that both household and individual distributions are closely matched. To overcome these limitations, different techniques have been proposed by Guo and Bhat (2007), Arentze et al (2007), Ye et al (2009) and Müller and Axhausen (2011) to estimate joint distributions of the control variables at both household and person level.

In this study, the Iterative Proportional Update (IPU) procedure was used for the population synthesis process. The IPU procedure is an extended version of the IPF procedure, developed by Ye et al in 2009. The procedure was designed to improve fit of household level distributions and person level distributions simultaneously. The basic idea behind the IPU procedure is that household weights are adjusted such that both household and individual level distributions can be matched as closely as possible. The IPF procedure is first applied to household level and then to person level to produce two separate and independent set of weights. The joint seed distributions of both household and person control variables are used to adjust and assign weights for each household/person type iteratively until marginal totals at both household and person level are closely matched to the known marginal distribution (Pendyala R M et al., 2009). In depth explanation and examples of the IPU procedure can be found in Ye et al (2009).

2.2 Data preparation for population synthesis

2.2.1 Data sources

Two main data sources were used in this study. They were the 1% Confidentialised Unit Record Files (CURFs) from the Australian Bureau of Statistics (ABS) 2006 population census and selected aggregated census data from the ABS Census Table Builder.

Study area

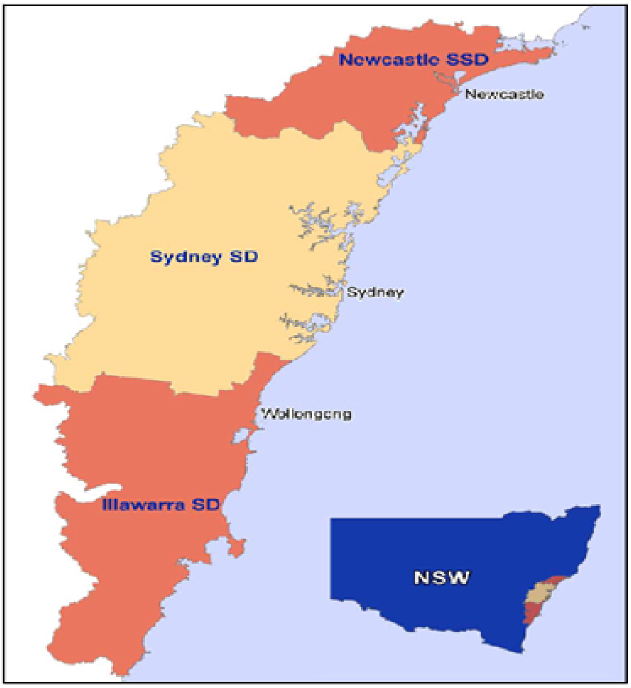
Sydney GMA was selected as the study area to align with the geographical coverage of the Household Travel Survey (HTS) by the Bureau of Transport Statistics (BTS) in New South Wales (Figure 1). HTS will provide crucial travel data of Sydney GMA for the second stage of this research project. It will be used in conjunction with the synthetic population to simulate travel patterns and existing demand conditions.

Based on the Australian Standard Geographical Classification (ASGC) 2006, there were 1953 Travel Zones (TZs) and 8374 Census Collection Districts (CCDs) in Sydney GMA. The household and person counts from the two main datasets are shown below:

Study Area: Sydney GMA

Benchmark data:	**Household count**	**Person count**
Census 2006	1864085	4895261
1% CURF 2006	19813	51801

Figure 1 Sydney Greater Metropolitan Area



Source: TDC, 2008

Record linkage and balancing census total counts

The data preparation for the population synthesis in this study comprises three steps:

The first step is to link geographical classifications between the 1% CURF 2006 and aggregated census data extracted from the Census Table Builder. Although the geographical areas in the 1% CURF and the Census data are both classified based on the Australian Standard Geographical Classification (ASGC), the 1% CURF was coded using Statistical Region (SR) whereas the census data were coded using Statistical Division (SD), Statistical Sub Division (SSD), Statistical Local Area (SLA) and CCD. Hence, it is necessary to ensure that the geographical areas covers in the census data is aligned and mapped to the geographical areas in the 1% CURF.

The second step of the data preparation is to link selected control variables between the two data sources. Generally, it is expected that the categories of some variables in the 1 % CURF are truncated compare to the census data extracted from the ABS Table Builder. It is crucial that these categories are being regrouped in the same way in both datasets. Some of these categories need to be regrouped or collapsed, either from CURF1% or aggregated census data, to ensure that both datasets contain variables of the same names with synchronised categories.

Finally, all sub totals and totals in the census aggregate data from the Census Table Builder are required to be made consistent. The existence of inconsistencies across sub totals and totals of the census data is due to the fact that ABS has introduced the confidentiality process to the Census Table Builder. The confidentiality process is undertaken to avoid revealing information that may allow for identification of particular individuals, families, household or dwellings. Consequently, total counts for selected control variables may vary. The process of overcoming these data inconsistencies are known as Census data 'balancing' (Chin and Harding, 2006). Census data balancing is considered to be one of the most time consuming process in the creation of small-area weights (Chin and Harding, 2006). The imbalance of census totals lead to an imbalance of marginal totals. The synthesis process will not be initiated if there are any inconsistencies in total counts or sub totals of the control variables.

In this study, the process of balancing the census subtotals was undertaken for every CCD and TZ in the study area. Minimal adjustments and redistributions were made during this process to achieve consistencies across subtotals and totals of each selected control variables in each geographical area. A program was created to handle and resolve the entire balancing process. In a nutshell, the process basically involved collating and rearranging all selected information (from the Census Builder Table) in a way that margins for all selected control variables can be adjusted across all geographical areas simultaneously.

2.2.2 Software Implementation

Many software applications for population synthesis have been developed over the years. Some are standalone software, such as PopSynWin by Auld et al. (2008) and PopGen by Arizona State University (2009). Some have been integrated to a model system (such as population synthesiser built in TRANSIMS by Beckman et al. (1996), ILUTE by Pritchard (2008); FSUMTS by Srinivasan and Ma (2009), CEMDAP by Guo and Baht(2007), and ALBATROSS by Arentze et al.(2007).

In this study, PopGen has been used to create synthetic populations at CCD and TZ level for Sydney GMA. To date, PopGen is the only standalone software that uses IPU algorithm. PopGen was developed in 2009 by Karthik Kondari and Bhargava Sana from the School of Engineering in Arizona State University.

As PopGen was primarily developed and designed to be used in the United States, the built-in graphic interface feature is not applicable for other countries and the specifications of the actual selected study area outside US is not possible for now. Adaptations have to be made for all input data as the designated data structure and format is understandably designed in accordance to Public Use Microdata Area (PUMA) from the US Census Bureau. It was necessary to realign and fit the geographical classification used in the ABS Census to those used in PUMA. Table 1 is a list of control variables selected and used in PopGen.

Table 1 Control variables used in PopGen

Control Variables at household level		Control Variables at person level	
HHCD	**Household Composition**	**AGEP**	**Age**
HHCD1	One family household	AGEP1	0-4 years
HHCD2	Two or more family household	AGEP2	5-9 years
HHCD3	Lone person household	AGEP3	10-14 years
HHCD4	Group household	AGEP4	15-19 years
HHCD5	Other groups	AGEP5	20-24 years
		AGEP6	25–29 years
STRD	**Dwelling Structure**	AGEP7	30–34 years
STRD1	Separate house	AGEP8	35–39 years
STRD2	Semi-detached, row	AGEP9	40–44 years
STRD3	Flat, unit or apartment	AGEP10	45–49 years
STRD4	Other dwelling	AGEP11	50–54 years
STRD5	Other groups	AGEP12	55–59 years
		AGEP13	60–64 years
NPRD	**Number of people (Derived)**	AGEP14	65–69 years
NPRD1	1 person	AGEP15	70–74 years
NPRD2	2 persons	AGEP16	75–79 years
NPRD3	3 persons	AGEP17	80–84 years
NPRD4	4 persons	AGEP18	85 years and over
NPRD5	5 persons		
NPRD6	6 persons	**SEXP**	**Sex**
NPRD7	7 persons	SEXP1	Male
NPRD8	8 or more persons	SEXP2	Female
VEHD	**Number of Motor Vehicles**	**LSFP**	**Labour Force Status**
VEHD1	None	LSFP1	Employed
VEHD2	1 motor vehicle	LSFP2	Unemployed
VEHD3	2 motor vehicles	LSFP3	Not in the labour force
VEHD4	3 motor vehicles	LSFP4	Other groups
VEHD5	4 or more motor vehicles		
VEHD6	Other groups		
Total Number of Household Type constraints = 1200		Total Number of Household Type constraints = 144	

3. Comparing the synthetic population to census counts

Overall, four synthetic population datasets were generated for Sydney GMA using Census data 2006:

- Household estimates at TZ level
- Person estimates at TZ level
- Household estimates at CCD level
- Person estimates at CCD level

The created synthetic data was validated against the aggregated census data extracted from the ABS Table Builder at TZ and CCD level. This section shows the performance statistics of all synthetic population data generated for the Sydney GMA. Table 2 gives an overview of differences in percentages between synthesised and actual census data at household and person level. According to the actual census data, there were 4,895,261 people lived in 1,864,085 households in Sydney GMA in 2006. The synthetic population generated produced a virtual match in terms of households for both syntheses run at TZ and CCD level. The number of synthetic persons generated was equal to 5,070,358 at TZ level, a discrepancy of 3.45 per cent compare to the actual number of persons in the population. Similarly, there were 4,990,319 synthetic persons generated at CCD level, a 1.9 per cent difference relative to the actual number of persons in the population.

Table 2 Overall comparison of synthesised and actual population of Sydney GMA in 2006

	TZ			CCD		
	Synthesised	**Actual**	**Difference (%)**	**Synthesised**	**Actual**	**Difference (%)**
Households	1864085	1864085	0.00	1864082	1864085	0.00
Persons	5070358	4895261	3.45	4990319	4895261	1.90

Figure 2 and Figure 3 show distributions of average δ values across geographies at TZ and CCD level for which synthetic estimates were generated. The δ value measures the average absolute deviation between the final household/person weighted sample computed and the corresponding known household/person marginal totals. Distributions of δ values for TZ and CCD were positively skewed with long tail. Average δ values for most geographical areas were concentrated on the lower ranges of the axis at both TZ level and CD level.

Figure 2 Distribution of δ value at TZ level

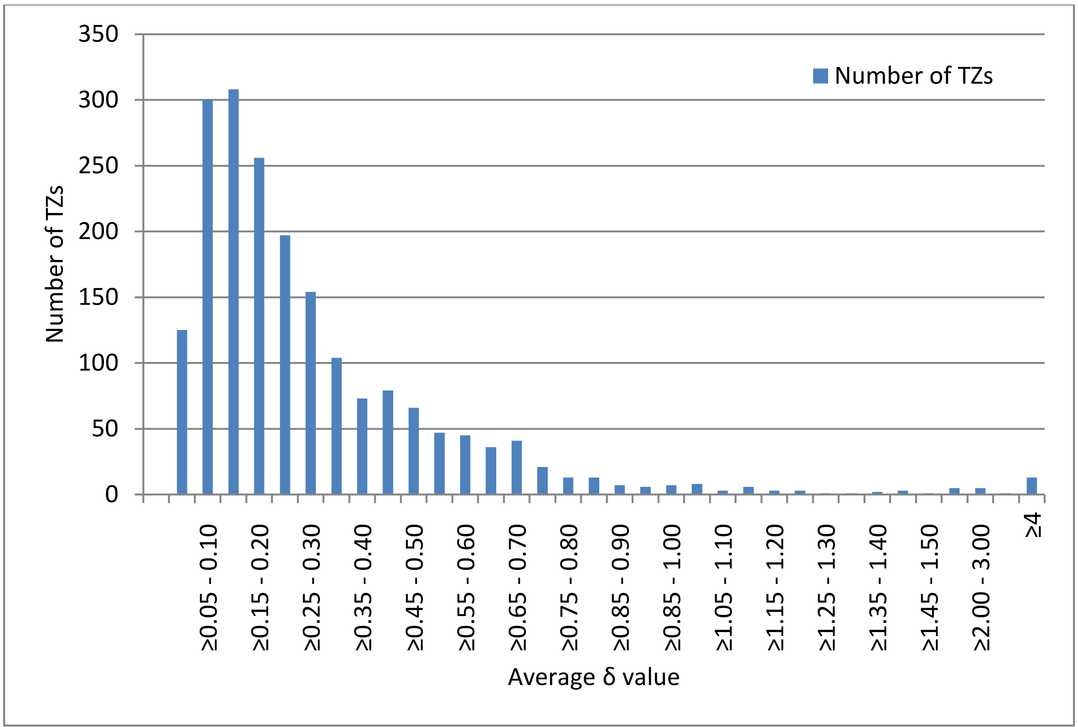
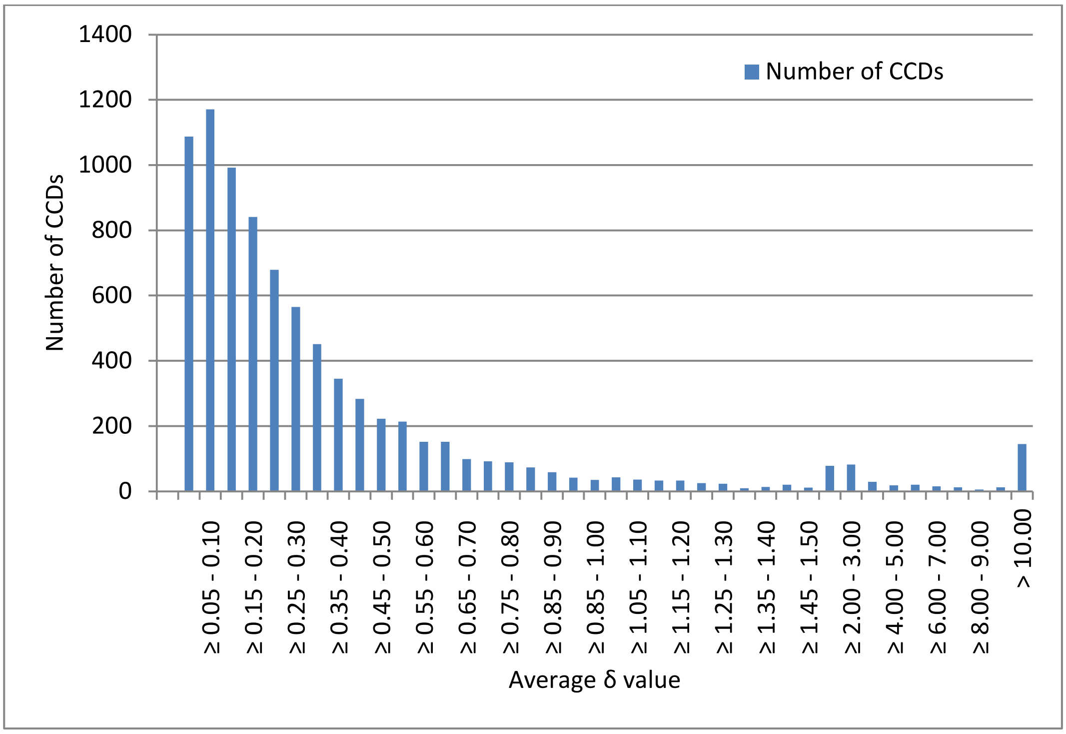


Figure 3 Distribution of δ value at CCD level



The next four sections summarise the synthetic estimates for Sydney GMA by household and person attributes at TZ and CCD level. Distributions of these attributes were compared between the synthesised and actual data. As shown, the synthetic results are generally closely matched with the actual census data.

3.1 Household estimates at TZ level

Figure 4 Household estimates at TZ level

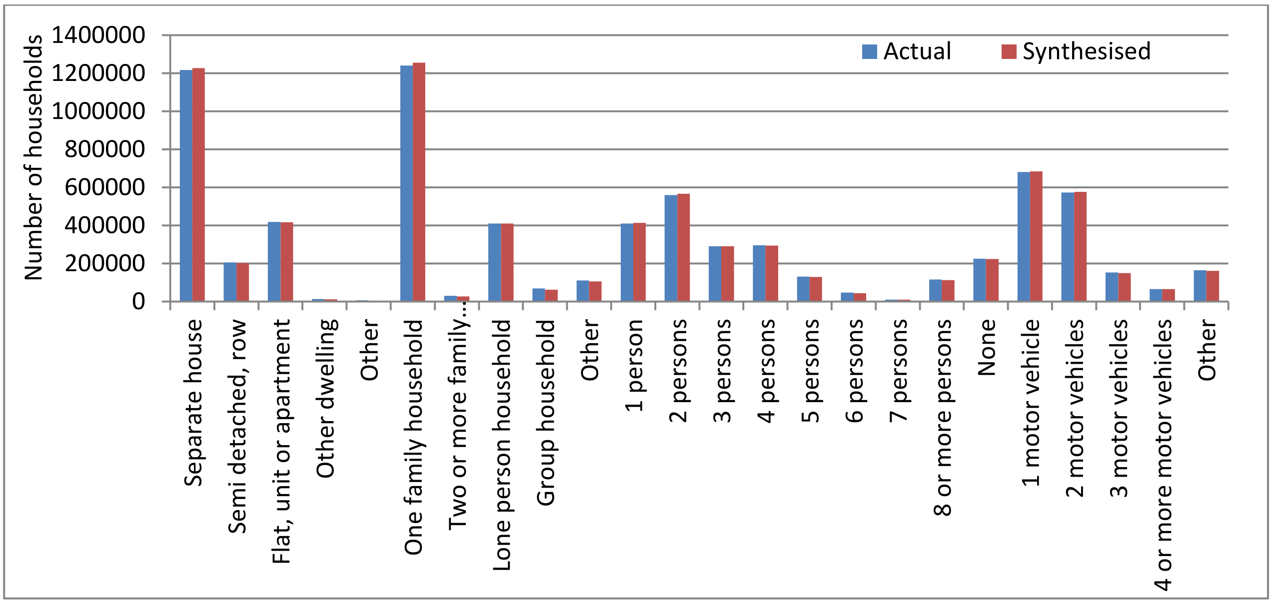


Table 3 Distribution of actual and synthesised population for Sydney GMA at TZ level, by household attributes 2006

	Actual	Synthesised	Difference
	%	%	% point
Dwelling Structure			
Separate house	65.34	65.87	0.53
Semi-detached, row	11.05	10.83	-0.22
Flat, unit or apartment	22.51	22.41	-0.09
Other dwelling	0.75	0.65	-0.11
Other	0.35	0.25	-0.10
Household Composition			
One family household	66.62	67.35	0.73
Two or more family household	1.67	1.46	-0.21
Lone person household	22.00	22.06	0.07
Group household	3.76	3.39	-0.37
Other	5.96	5.74	-0.22
Number of Persons Usually Resident in Dwelling			
1 person	22.00	22.21	0.21
2 persons	30.06	30.38	0.32
3 persons	15.61	15.59	-0.01
4 persons	15.87	15.83	-0.04
5 persons	7.08	7.00	-0.08
6 persons	2.53	2.40	-0.13
7 persons	0.57	0.55	-0.01
8 or more persons	6.29	6.03	-0.26
Number of Motor Vehicles			
None	12.10	12.02	-0.08
1 motor vehicle	36.51	36.73	0.22
2 motor vehicles	30.80	30.95	0.15
3 motor vehicles	8.20	8.09	-0.11
4 or more motor vehicles	3.58	3.52	-0.06
Other	8.82	8.68	-0.13

3.2 Person estimates at TZ level

Figure 5 Person estimates at TZ level

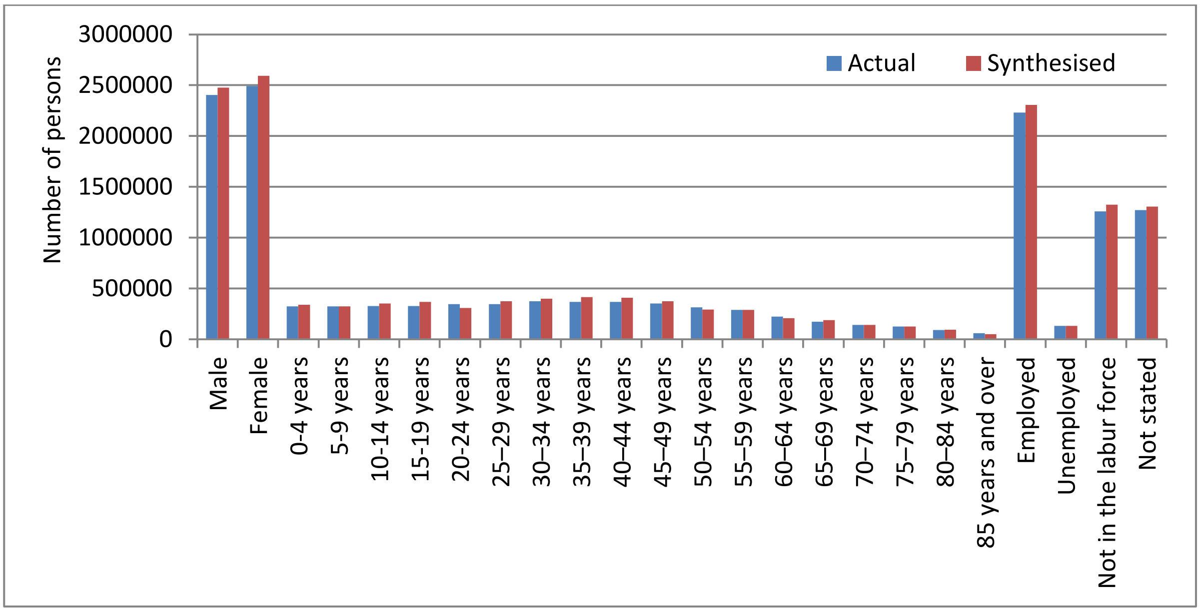


Table 4 Distribution of actual and synthesised population for Sydney GMA at TZ level, by person attributes 2006

	Actual	Synthesised	Difference
	%	%	% point
Gender	49.09	48.85	-0.25
Male	50.91	51.15	0.25
Female			
Age			
0-4 years	6.61	6.71	0.10
5-9 years	6.64	6.40	-0.23
10-14 years	6.70	6.98	0.28
15-19 years	6.67	7.25	0.58
20-24 years	7.11	6.10	-1.00
25–29 years	7.11	7.38	0.27
30–34 years	7.66	7.88	0.22
35–39 years	7.57	8.19	0.62
40–44 years	7.54	8.09	0.56
45–49 years	7.23	7.38	0.15
50–54 years	6.45	5.79	-0.65
55–59 years	5.95	5.75	-0.20
60–64 years	4.59	4.12	-0.47
65–69 years	3.58	3.76	0.18
70–74 years	2.94	2.79	-0.15
75–79 years	2.59	2.52	-0.07
80–84 years	1.85	1.87	0.01

Table 4 continue

	Actual	Synthesised	Difference
85 years and over	1.23	1.04	-0.19
Labour Force Status			
Employed	45.60	45.50	-0.10
Unemployed	2.72	2.62	-0.11
Not in the labour force	25.69	26.14	0.45
Not stated	25.99	25.74	-0.25

3.3 Household estimates at CCD level

Figure 6 Household estimates at CCD level

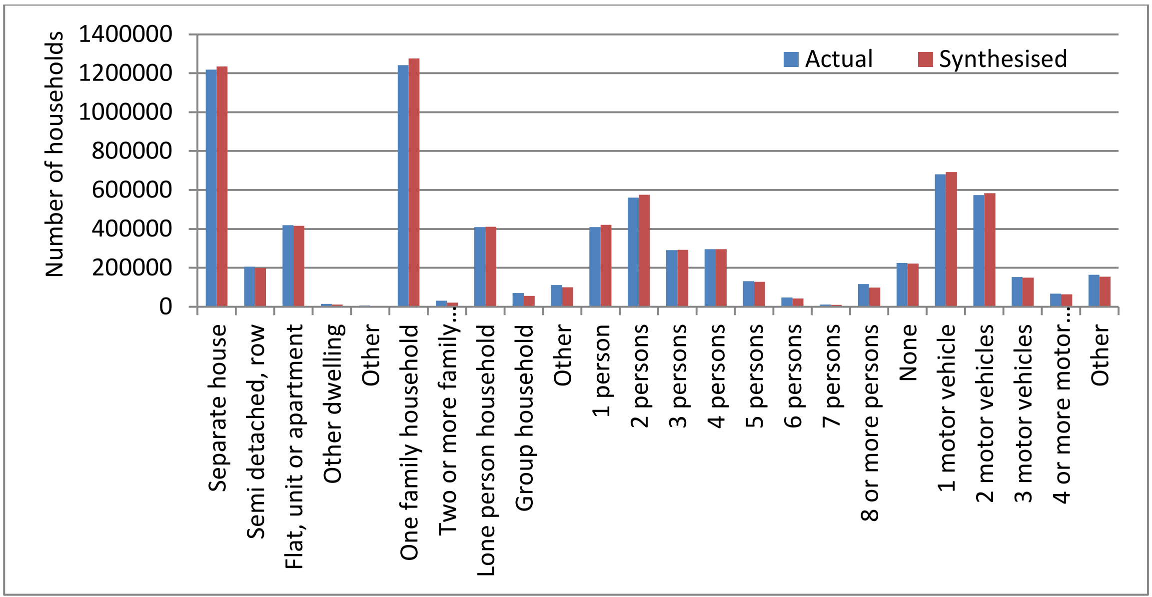


Table 5 Distribution of actual and synthesised population for Sydney GMA at CCD level, by household attributes 2006

	Actual	Synthesised	Difference
	%	%	% point
Dwelling Structure			
Separate house	65.30	66.20	0.90
Semi-detached, row	11.10	10.70	-0.40
Flat, unit or apartment	22.50	22.30	-0.20
Other dwelling	0.80	0.60	-0.20
Other	0.30	0.20	-0.10
Household Composition			
One family household	66.60	68.40	1.80
Two or more family household	1.70	1.10	-0.60
Lone person household	22.00	22.00	0.00

Table 5 continue

	Actual	Synthesised	Difference
Group household	3.80	3.00	-0.80
Other	6.00	5.40	-0.60
Number of Persons Usually Resident in Dwelling			
1 person	22.00	22.60	0.60
2 persons	30.10	30.90	0.80
3 persons	15.60	15.70	0.10
4 persons	15.90	15.90	0.00
5 persons	7.10	6.90	-0.20
6 persons	2.50	2.20	-0.30
7 persons	0.60	0.50	-0.10
8 or more persons	6.30	5.30	-1.00
Number of Motor Vehicles			
None	12.10	11.90	-0.20
1 motor vehicle	36.50	37.20	0.70
2 motor vehicles	30.80	31.30	0.50
3 motor vehicles	8.20	8.00	-0.20
4 or more motor vehicles	3.60	3.40	-0.20
Other	8.80	8.20	-0.60

3.4 Person estimates at CCD level

Figure 7 Person estimates at CCD level

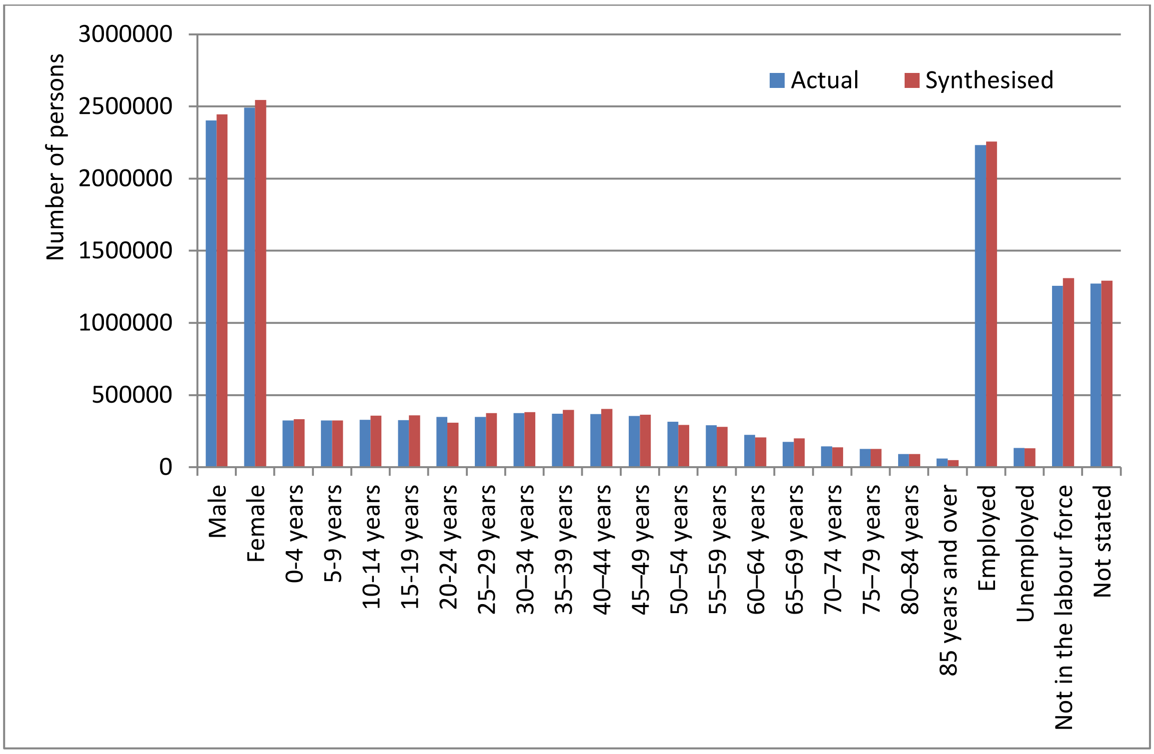


Table 6 Distribution of actual and synthesised population for Sydney GMA at CCD level, by person attributes 2006

	Actual	Synthesised	Difference
	%	%	% point
Gender	49.09	49.01	-0.09
Male	50.91	50.99	0.09
Female			
Age			
0-4 years	6.61	6.67	0.06
5-9 years	6.64	6.51	-0.13
10-14 years	6.70	7.16	0.47
15-19 years	6.67	7.20	0.53
20-24 years	7.11	6.20	-0.91
25–29 years	7.11	7.49	0.38
30–34 years	7.66	7.65	-0.01
35–39 years	7.57	7.94	0.38
40–44 years	7.54	8.11	0.57
45–49 years	7.23	7.30	0.07
50–54 years	6.45	5.88	-0.56
55–59 years	5.95	5.60	-0.35
60–64 years	4.59	4.13	-0.46
65–69 years	3.58	4.00	0.42
70–74 years	2.94	2.77	-0.17
75–79 years	2.59	2.56	-0.03
80–84 years	1.85	1.83	-0.02
85 years and over	1.23	0.99	-0.24
Labour Force Status			
Employed	45.60	45.23	-0.37
Unemployed	2.72	2.61	-0.11
Not in the labour force	25.69	26.25	0.56
Not stated	25.99	25.91	-0.08

4. Summary and conclusions

It is a challenging task to generate synthetic population data for travel demand modelling by geographies as small as TZs and CCDs. However, the fusion of disaggregated geographic information and socio demographic profiles is crucial to travel demand forecasting. It provides an important platform to model small area to area travel demand patterns and to explore the underlying factors behind travel activities.

There were a couple issues in the population synthesis process that need to be investigated further. Firstly, the common issue of 'zero cell problems'. The sample data generally does not contain sufficient information in domain specific or small geographies to provide adequate statistical precision to some of the estimates. Demographic group that appear in the population as represented by the aggregate data may not be represented in the sample data. In PopGen, the 'zero cell' correction issue is addressed by borrowing prior information for zero cells from the census data subject to an upper limit. This may allow convergence but the process does not necessary correct the zero cell problems.

Secondly, discrepancies in total counts between the actual census data and the synthetic data at some person level remain prominent after the IPU procedure. As shown in the synthetic results at TZ level, the effort of consolidating sparse categories to fewer geographical zones has slightly enhanced the performance at household level distributions but has not shown significant improvement for the performance of person level distributions.

Further validations of the synthetic population are necessary as the representativeness of the dataset is critical to the accuracy of subsequent simulation outcomes in activity based travel demand forecasting. In the next stage of the project, these datasets will be used in conjunction with the Sydney Household survey to generate household activity patterns.

References

Arentze, T A, Timmermans, HJ and Hofman, F (2007) Population Synthesis for Microsimulating Travel Behavior, *Transportation Research Record*, 2014 (11) 85–91.

Axhausen, K and Garling, T (1992) Activity based approaches to travel analysis: Conceptual frameworks, models and research problems, *Transport Reviews 12*: 324-341.

Auld, J, A K Mohammadian and K Wies (2008) Population synthesis with control category optimization, *paper presented at the 10th International Conference on Application of Advanced Technologies in Transportation*, Athens, Greece, May 2008.

Auld, J., A. K. Mohammadian and K Wies (2010) An efficient methodology for generating synthetic populations with multiple control levels*, paper presented at the 89th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2010.

Beckman, R J, Baggerly, K A and McKay, M D (1996) Creating synthetic baseline populations, *Transportation Research Part A: Policy and Practice*, 30 (6) 415–429.

Bhat, C R and Koppelman, F S (1999) Activity-Based Modeling for Travel Demand, *Handbook of Transportation Science*, R.W. Hall (ed.) , Kluwer Academic Publisher.

Bhat, C R, Guo, JY, Srinivasan, S, and Sivakumar A 2003, *CEMDAP User's Manual*, Product 4080-P2, prepared for the Texas Department of Transportation, October.

Bowman, J L (2004) *A comparison of population synthesizers used in microsimulation modelso f activity and travel demand*, http://jbowman.net/papers/2004.Bowman. Comparison_of_PopSyns.pdf, accessed on 29/07/2010.

Bowman, J L (2009) Population synthesizers, *Traffic Engineering and Control*, 49 (9) 342.

Chin, F S. and Harding, A (2006) *Regional Dimensions: Creating Synthetic Small-area Microdata and Spatial Microsimulation Models*, Technical Paper no.33, NATSEM, University of Canberra.

Frick, M and Axhausen, K W (2004) Generating synthetic populations using IPF and Monte Carlo techniques: Some new results*, paper presented at the 4th Swiss Transport Research Conference, Ascona*, March 2004.

Kitamura, R, Chen C, Pendyala, R M and Narayanan, R (2000), Micro-simulation of daily activity travel patterns for travel demand forecasting, *Transportation*, 27:25-51.

McNally, M G, and Recker, WW (2000) *Activity based models of travel behavior: Evolution and implementation.* Elsevier Science.

Miller, E J (1997) Microsimulation and activity based forecasting*, Activity based Travel Forecasting Conference Proceedings*. Travel Model Improvement Program (TMIP), U.S. Department of Transportation, Washington, D.C.

Moeckel, R, Spiekermann, K, and Wegener, M (2003)*,* Creating a synthetic population, *Paper presented at the 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM),* Sendai, Japan, May 2003.

Müller, K and Axhausen K W (2011), Population synthesis for microsimulation: State of the art, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Pendyala, R M, Bhat C R, Goulias K G,Paleti R, Konduri K, Sidharthan R,Christian KP (2013) *SimAGENT Population Synthesis* Report submitted to South California. Association of Governments, Santa Barbara.

Petersen, E and Vuk, G (2006) Comparing a conventional travel demand model to an activity-based travel demand model: A case study of Copenhagen, *paper present at the European Transport Conference*, Strasbourg.

Pritchard, DR and Miller, EJ (2009) Advances in agent population synthesis and application in an integrated land use and transportation model, *paper presented at the 88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2009.

Ryan, J, Maoh, H and Kanaroglou, P (2009) Population synthesis: Comparing the major techniques using a small, complete population of firms', *Geographical Analysis*, 41 (2) 181–203.

Shiftan, Y and Ben Akira M S (2003) Activity-Based Modeling as a Tool for Better Understanding Travel Behaviour', *paper presented on the 10th International Conference on Travel Behaviour Research*, Lucerne, 10-15th of August 2003.

Srinivasan, S. et al. (2009), Synthetic population generation: A heuristic data-fitting approachand validations, *paper presented at the 12th International Conference on Travel Behaviour Research (IATBR)*, Jaipur, December 2009.

Srinivasan, S. and L. Ma (2009) Synthetic population generation: A heuristic data-fitting approach and validations, *paper presented at the 12th International Conference on Travel Behaviour Research (IATBR)*, Jaipur, December 2009.

Srinivasan, S, L Ma and K Yathindra (2008) *Procedure for forecasting household characteristics for input to travel-demand models*, Final Report, TRC-FDOT-64011-2008, Transportation Research Center, University of Florida. http://www.fsutmsonline.

Xu, M, Taylor, MAP (2004) Microsimulation modelling of synthetic population, presented at the *27th Australasian Transport Research Forum*, 29 September-1 October 2004, Adelaide.

Ye, X, Konduri, K, Ram P, Sana, B and Waddell, P (2009) A Methodology to Match Distributions of Both Household and Person Attributes in the Generation of Synthetic Populations. *Presented at the 88th Annual Meetings of the Transportation Research Board.*