PROJECTING SMALL AREA STATISTICS WITH AUSTRALIAN SPATIAL MICROSIMULATION MODEL (SPATIALMSM)

Yogi Vidyattama

Research Fellow, National Centre for Social and Economic Modelling (NATSEM), University of Canberra.

Robert Tanton

A/g Research Director of the Social Inclusion and Small Area Modelling Team, National Centre for Social and Economic Modelling (NATSEM), University of Canberra.

ABSTRACT: "Think Global, Act Local" has become a theme for development planning of governments around the world. This is partly due to the increasing recognition of the importance of planning at a small area level. As a consequence, there is a need to derive estimates of socio-economic variables for local areas, and project these into future. Regional scientists have been involved in both the small area estimation and the application of regional estimates to Government policy. This paper will describe a new technique to project small area socio-economic statistics into the future using a spatial microsimulation model. Spatial microsimulation models are a new form of microsimulation models that allow small area estimates of socio-economic variables to be derived from survey data, and allow scenario modelling using survey microdata. This paper extends the spatial microsimulation methodology by adding a projection technique that allows projections of the microdata to be derived. The paper applies this method to project variables that target service delivery populations for Australian State Governments. T he spatial microsimulation method used also allows some scenario modeling, and the paper will calculate projections of service delivery populations after a scenario of increasing unemployment as a result of the global financial crisis.

1. INTRODUCTION

In the past decade, the need to analyse local or regional economies has brought dynamic spatial microsimulation into the forefront of microsimulation research. There is an increasing recognition of the importance of regional economies in terms of sub-national economies and the way they evolve over time (Neary, 2001). This has made small area statistics as well as projections for small areas increasingly crucial. The increasing need of many governments in the world to plan their economy at a regional level has also increased the need for small area statistics and their projections.

The strong demand for small area information by planning agencies, especially State and Territory governments in the case of Australia, has mainly focused on the characteristics of individuals and households and the small area impact of possible policy changes. There are several reasons for this. First, such information is required, for example, by those government agencies with responsibility for allocating scarce resources to where they are most needed – ranging from the most effective placement of child care or aged care services to

disability programs and services targeted towards youth-at-risk. Second, governments often need accurate information about the degree to which deprivation or disadvantage is concentrated in particular places, to inform social policy formation more generally. Third, an ability to estimate the spatial impact of a policy before the policy change is introduced helps to prevent the emergence of unintended small area consequences.

Despite this great need for small area statistics for planning purposes, the data can be very hard to obtain. National censuses are typically conducted relatively infrequently and their extensive geographic detail comes at the price of containing only a limited range of information about households. On the other hand, surveys obtain much richer information, but are designed for national, or at most, state level estimation. They are therefore unsuitable for directly estimating statistics for small areas due to small sample sizes in small areas (Heady et al., 2003). Therefore, various techniques have been developed to achieve small area estimates from sample surveys. Spatial microsimulation techniques are among the techniques used to estimate small area statistics. (see Rahman, 2008 for a review of the literature). The spatial microsimulation technique essentially reweights survey data to match new small area benchmarks from the Census

Dynamic microsimulation allows the user to predict regional economic and demographic conditions in the future as well as predict the impact of policy at a regional level. Although the development of the dynamic microsimulation model was started a half century ago by Guy Orcutt in 1957, the development of dynamic microsimulation for small areas is fairly recent. This is mainly because the use of microsimulation in a spatial context is somewhat rare (Birkin et al., 1996). Birkin and Clarke (1988, 1989) and Williamson (1992) are among the first microsimulation applications that involve spatial estimation. SVERIGE, which was built in 1996 in Sweden, is considered to be the first dynamic spatial microsimulation model that covers the entire nation (Vencatasawmy et al., 1999; Holm et al 2001). The model was developed based on CORSIM, a dynamic microsimulation model to estimate new indicators of wealth in the USA (Caldwell 1990). At the time SVERIGE was built, there was another dynamic spatial microsimulation that was being built in the Netherlands (Hooimeijer, 1996). Other dynamic spatial microsimulation models built in recent years include SimBritain (Ballas et al., 2005a), SMILE (Ballas et al., 2005b) and an agent based spatial microsimulation model (Wu et al., 2008).

In general, there are many methodologies that have been developed to make a microsimulation model (including a spatial microsimulation model) dynamic. These methodologies have also been used to produce more sophisticated and more accurate projections from the model. These methodologies can be categorised into fully dynamic and semi or pseudo dynamic microsimulation models. While fully dynamic models simulate the dynamic behaviour of the unit record in the survey data (or microdata), a semi or pseudo dynamic model projects the dynamic constraints from the census data (Caldwell, 1990). Other researchers have named the pseudo-dynamic approach static ageing (O'Donoghue, 2001). The static ageing method is traditionally used if the microsimulation model is a static model. Static ageing adjusts the benchmark

table to account for changes in the population structure, price structure (inflation), the distribution of income and to some extent changes in policy rules (O'Donoghue , 2001). In many cases these adjustments are based on national macro economics forecast (Eason, 1996; Gupta and Kapur, 1996).

This paper describes an effort to project small area statistics in Australia by employing an existing spatial microsimulation model for Australia (SpatialMSM). In particular, this paper shows how we have modified the Australian static spatial microsimulation model SpatialMSM to make it a pseudo dynamic microsimulation model.

Section two briefly discusses spatial microsimulation and overseas efforts to derive projections from spatial microsimulation models. Section three will introduce SpatialMSM, the Australian spatial microsimulation model, as a static spatial microsimulation model for Australia, while section four describes the projection methodology we have developed and its reliability. Section five contains conclusions.

2. PROJECTIONS USING SPATIAL MICROSIMULATION

Spatial microsimulation involves creating synthetic spatial microdata.[1] Some of the early research in this field was undertaken by geographers and concentrated upon whether it was possible to create small area specific microdata from the UK Census one per cent sample (Williamson et al, 1998; Voas and Williamson, 2000; Williamson, 2001). While various approaches to reconstructing spatially detailed microdata have been trialled, including data fusion and synthetic reconstruction (Voas and Wiliamson, 2000, p. 349), the more successful endeavours essentially involve methods of reweighting the original sample survey data to match small area population targets from a relevant Census. Ballas et al (2006a, p. 65) explain these techniques 'involve the merging of census and survey data to simulate a population of individuals within households (for different geographic units), whose characteristics are as close to the real population as it is possible to estimate'.

Once synthetic household microdata have been created for each small area, then it becomes feasible to use this microdata for microsimulation modelling. Microsimulation models were initially developed within the discipline of economics (Orcutt et al, 1986) and have today become very widely used by governments across the developed world for analysis of the fine-grained distributional impact of possible changes in government programs (Harding, 1996; Gupta and Kapur, 2000; Mitton et al, 2000; Harding and Gupta, 2007b). However, importantly, the overwhelming majority of these microsimulation models have been national models, constructed on top of national sample survey

---

[1] Unit record data (alternatively termed 'microdata') usually consist of thousands of individual records of persons, families or households in a computer readable format. Such microdata are the essential building block for microsimulation models, which in the past two decades have revolutionised the quality of information available to policy makers about the likely distributional impact of policy reforms that they are contemplating (Harding and Gupta, 2007a).

microdata and predicting the distributional impact of policy change for an entire country, rather than for a small region within a country.

A new development during the past decade has been the construction of spatial microsimulation models, constructed using the synthetic spatial microdata bases described earlier. This rapidly growing field now includes simulation of the small area impact of changes in income taxes and cash transfers (Chin et al, 2005; Harding et al. 2009b); development of small area measures of poverty and housing stress (Tanton et al, 2009; McNamara et al, 2007); small area modelling of Activities of Daily Living Status and need for different types of care (Lymer et al, 2006, 2008a, 2008b); development of the SimObesity model to examine small area obesity among children (Procter, 2007); small area health-related conditions (Ballas et al, 2006a); the socio-economic impacts of major job gain or loss at the local level (Ballas et al, 2006b) and a range of other applications (Ballas et al, 2005a, 2005b; Clarke 1996).

A further development has been the attempt to 'age' the spatial microsimulation databases forward through time, so as to provide projections. As noted in Harding and Gupta (2007a), a conceptual distinction can be drawn here between models that undertake 'static ageing' (such as reweighting the small area dataset to future population projections) and those that attempt 'dynamic ageing', which involves updating the characteristics of the micro-units through time.

As outlined in the introduction, there are a number of dynamic microsimulation models already in existence (SVERIGE, CORSIM, SMILE). There are also examples of pseudo-dynamic models in the UK, which are not fully dynamic in that they do not model individual life experiences like mortality, fertility and migration (as SVERIGE and SMILE do); but reweight to projections of Census tables, so use static ageing. Examples of these models include SimBritain (Ballas et al., 2005a).

SVERIGE uses the pattern of emigration, immigration, employment and earnings, education, leaving home, divorce, cohabitation and marriage, as well as mortality and fertility as the dynamic individual behaviours in the model. The Monte Carlo simulation picks individuals in the Microdata to experience any of the above behaviours based on simple probabilities and hence updates the individual characteristics in the microdata. So central to creating projections in this model are accurate probabilities of each behaviour. In SVERIGE, these probabilities are obtained using either probabilities from past experience or estimated logistic regression equations.

SMILE is built as both a static and dynamic spatial microsimulation model (Ballas et al., 2005b). It is constructed to estimate and project small area statistics in Ireland. The model starts as a static model using an iterative proportional fitting (IPF) method to spatially disaggregate the aggregate microdata. Once this has been done, the demographic processes of mortality, fertility and migration are simulated. The mortality process is simulated by using the probability of death based on age, gender and location while the probability of birth is simulated based on age, marital status and location. The simulation of the migration process uses random sampling from calculated

migration probabilities derived from the 1991 and 1996 Census of Population. These data provide migration probabilities from one area to another by age, gender and location.

SimBritain (Ballas et al., 2005a) is a spatial microsimulation model for Britain's small areas. Unlike SVERIGE and SMILE, SimBritain is constructed as a pseudo dynamic microsimulation model. The model projects benchmark tables from 2001 to 2011 and 2021 using the long term trend of each small area based on data from the UK 1971, 1981, and 1991 census.' The benchmark projections are calculated using a logistic model of the changing population proportion in each category of each benchmark table. After all the 6 benchmark tables in SimBritain are projected, the microdata are reweighted to the projections, and new weights are calculated for each household or person on the microdata.

3. PROJECTING SMALL AREAS STATISTICS IN AUSTRALIA

SpatialMSM is a spatial microsimulation model that has been developed to estimate small area statistics in Australia. The model has been under development for several years, initially reweighting a household expenditure survey to 2001 Census small area benchmarks (see Chin et al. 2005; 2006; Chin and Harding, 2006, 2007 and, for documentation of the earliest efforts, see Melhuish et al 2002). Later versions of the modelling have reweighted ABS income surveys to 2001 Census benchmarks (Tanton et al, 2009), while the version described in this paper utilises the latest 2006 Census benchmarks.

Besides estimating small area statistics, this model has also been linked to a static microsimulation model in Australia called STINMOD to estimate small area impacts of policy change. The model is also used by various service delivery agencies to derive small area estimates of groups that will require services from the service providers. The general method has also been used to develop a small area spatial microsimulation model for projecting customer service needs, CUSP (Phillips, 2007); develop HOUSEMOD for examining the impact of changes in housing assistance (McNamara et al, 2007); and to develop CAREMOD for assessing small area care needs (Lymer et al., 2006).

The SpatialMSM model employs an Australian Bureau of Statistics' reweighting program called GREGWT (Tanton et al, 2009). The GREGWT algorithm uses a regression technique to create initial weights for the Microdata and then because the optimisation process is constrained to having no weights less than 0, it iterates until the new weights produce an overall characteristic that is close to the constraints or benchmarks for a small area. The general method is outlined in more detail in Lymer et al. (2008b) and Chin et al. (2006).

3.1 SPATIALMSM/08C

The version of the spatial microsimulation model used for this paper is called SpatialMSM/08C. This version of the modelling has been designed to derive results for Statistical Local Areas (SLA) across Australia, using the 2006 Australian Standard Geographic Classification. This is done by reweighting households and individuals from the 2002-03 and 2003-04 Surveys of Income

and Housing to Statistical Local Area benchmarks from the 2006 Australian Census of Population and Housing (with all of the above being produced by the Australian Bureau of Statistics (ABS)).

The first step in producing the small area estimates involves combining information from two surveys – the 2002-03 and 2003-04 ABS Survey of Income and Housing (SIH) Confidentialised Unit Record Files (CURFs) – and the 2006 Australian Census of Population and Housing. This process uses GREGWT to reweight the national sample survey microdata files to the 2006 SLA Census tables, based on the 11 Census benchmarks shown in Table 1 below.

Table 1. Benchmark tables used in the reweighting algorithm

Number	Benchmark Table	Level
1	Age by sex by labour force status	Person
2	Total number of households by dwelling type (Occupied private dwelling/Non private dwelling)	Household
3	Tenure by weekly household rent	Household
4	Tenure by household type	Household
5	Dwelling structure by household family composition	Household
6	Number of adults usually resident in household	Household
7	Number of children usually resident in household	Household
8	Monthly household mortgage by weekly household income	Household
9	Persons in non-private dwelling	Person
10	Tenure type by weekly household income	Household
11	Weekly household rent by weekly household income	Household

Note: Most Benchmark Tables contain the total number of persons or households in occupied private dwellings (OPD) except for Table 2 and Table 9. These tables include people in non-private dwellings. People in non-private dwellings include people in prisons, hospitals, aged care facilities, etc.

Source: ABS Census Population and Housing 2006.

Given that the two national sample surveys and the census were conducted at different points in time, there are some adjustments needed before the reweighting process can start. The gross incomes from the surveys are uprated to 2006 dollar values, using changes in average weekly earnings to make the income values in both SIH years comparable to gross income values from the Census. The weekly household rent and mortgage on the surveys are also uprated using the changes in the housing component of the ABS Consumer Price Index (ABS 2008a).

The Statistical Local Area (SLA) is the spatial unit used in this paper. The SLA is one of the standard spatial units described in the Australian Standard Geographic Classification 2006 (ABS 2007). There were two main reasons why the SLA was chosen as the unit of analysis in this study. First, the SLA is the

smallest unit in the ASGC where there are not substantial issues with confidentiality, as occur with Census Collection Districts. (The ABS applies a confidentialising process to table cells with a small cell size.) Second, SLAs cover the whole of Australia (as opposed to Local Government Areas which do not cover areas with no local government) and also cover contiguous areas (unlike some postcodes).

The reweighting process in SpatialMSM uses an iterative constrained optimisation technique to calculate weights to produce the SLA level data that are closest to the Census Benchmarks. The procedure applies a generalised regression procedure outlined in Bell (2000) in a SAS macro developed within the ABS called GREGWT. The SpatialMSM model uses this process to create a synthetic household microdata file for each Statistical Local Area (SLA) in Australia, containing a set of synthetic household weights which replicate, as closely as possible, the characteristics of the real households living within each small area in Australia.

Because the reweighting process is an iterative process, there are areas where the procedure cannot find a solution (called non-convergence). The original GREGWT criteria for non-convergence is whether the maximum number of iterations (as specified by the user) was reached and a solution was not found. For SpatialMSM, the number of iterations was set to 30. After some experimenting, the original criteria from GREGWT was found to be too strict, since for some areas, the population estimates using the weights were still reasonable when GREGWT showed that the procedure had not converged. Therefore, another measure has been used in determining the reliability of the weights. This measure is the total absolute error (TAE) from all the benchmarks. This measure was developed by Paul Williamson for a combinatorial optimisation reweighting method (see Williamson et al, 1998). The TAE will be 0 if we can match the benchmarks perfectly, and will increase as the estimation process fails to meet the benchmarks. This will be related to the population of the area being estimated; so for an area with a population of 100 people, a TAE of 50 is bad; but for an area with a population of 10,000 people a TAE of 50 is good. So the criteria used in this paper is that if the TAE divided by the population of the area is greater than 1 then the area is dropped from any future analysis.

The model SpatialMSM/08C has been used to produce weights for 1214 SLAs and failed to produce reliable weights (so the TAE was greater than one) for 138 SLAs. Most of the areas where the TAE was greater than one were industrial areas, office areas or military bases with very low population counts. As a result, the proportion of people living in these SLAs is very small (Table 2). Only 0.7 percent of the total Australian population in 2006 were lost in the reweighting process.

While the results look acceptable for most states and territories, it must be noted that estimates for one quarter of the population in the Northern Territory had a high TAE - and thus small area estimates for the Northern Territory from SpatialMSM/08C should be treated very cautiously. (The Northern Territory contains many SLAs where a high proportion of the population are indigenous

Australians. Such households are not well represented in the sampling frame for the national ABS sample surveys that were reweighted, so the reweighting process may struggle to find an acceptable solution.)

Table 2. Number of SLAs dropped due to failed accuracy criteria in SpatialMSM/08C

State/ Territory	SLAs with failed TAE	Total SLAs	Per cent of SLAs with failed TAE (%)	Per cent of all persons living in SLAs with failed TAE (%)
NSW	2	200	1.0	0.4
VIC	4	210	1.9	0.0
QLD	43	479	9.0	0.8
SA	7	128	5.5	0.4
WA	17	156	10.9	0.9
TAS	1	44	2.3	0.1
NT	48	96	50.0	25.2
ACT	16	109	14.7	1.0
Australia	138	1422	9.7	0.7

Source: SpatialMSM/08C applied to 2002/03 and 2003/04 SIH CURF.

4. USING SPATIALMSM FOR PROJECTIONS OF SMALL AREA STATISTICS

In a prototype version of the modelling, a simple static ageing procedure was adopted, which essentially involves reweighting the data for each small area to population projections for each small area. This is similar conceptually to the approach followed in SimBritain (Ballas et al. 2005a) and in earlier work on projecting consumer characteristics out to 2020 in Australia (Harding and Gupta. 2007b). However, in this simple method, the 11 benchmarks are not projected using long-term trends, as in SimBritain. The main reason why, for example, the long-term trend away from home ownership and towards private rental for younger generations has not been simulated (Tanton et al. 2008, p. 26) is that such a technique requires data about such long-term trends *at the small area* level. This is difficult to achieve, especially because the changing boundaries of small areas makes the establishment of long-term trends by SLA a challenging task. (This is will illustrated, for example, in Vu et al. 2008, where they describe some of the challenges faced when trying to make the 2001 and 2006 Census SLA results comparable).

Another, more complex, approach is to project each one of the benchmarks, and then reweight to these new projections. This is the approach used by SimBritain, and also outlined in this paper for SpatialMSM. The approach used to project the benchmark tables leverages directly off the customised projections prepared for the Australian Government Department of Health and Ageing (DOHA) by the Australian Bureau of Statistics (ABS) (http://www.health.gov.au/internet/main/publishing.nsf/Content/ageing-stats-

lapp.htm). These population projections contain age by sex projections for each SLA in Australia until 2027 using the base assumption that has been described in the explanatory notes for the data, available from the DOHA website, and further discussed in ABS (2008b).

Note that the population projections from the DOHA have been produced using the cohort component method with the following assumption. The national fertility rate will decline gradually to 1.8 babies per woman in 2021, the life expectancy will increase to 85-88 in 2055, while migration is based on the historical and trend data. The population projections exclude 7 SLAs, being off-shore and migratory areas where no population projections are supplied. Therefore these SLAs will not be in our projections.

4.1 Projection Process

As described above, one of the first steps in the creation of SpatialMSM/08C essentially involves reweighting the two income survey sample files to benchmark tables from the 2006 Census. Creating the out years versions of the database again involves reweighting - but this time to newly created estimated benchmark tables for *future* given years.

One of the advantages of reweighting to benchmark tables in future years is that the projected benchmark tables can use a very rough estimate in the first stage, and then the method for projecting each benchmark table can be refined in the future, and the weights easily recalculated using the more refined benchmark tables. The method used in this paper to get the initial projections of the benchmark tables uses a logistic regression model based on age by sex by labour force status projections, but in the future any of the benchmark tables could be refined and new weights calculated.

The first constraint or benchmark table that is projected is the Labour Force by Age by Sex benchmark table, which has been projected up to 2027. To project this database, the SLA level population projections from DOHA are combined with projections of labour force status used in the Australian Commonwealth Treasury's 2007 *Inter Generational Report* (IGR) (Treasury 2007). The long run historical trend was also used in the report to project the participation rates for men and women of different ages. This incorporates the changing composition of the labour force in Australia, especially with more women participating in the labour force.

Our initial problem with the DOHA SLA level population projections is that they are only available by age and sex, and not by labour force status. The projection of age by sex by labour force status is undertaken in two steps. The first is to take the DOHA age by sex by SLA projections for 2007 (so the year after our benchmark table) and use the labour force by age/sex by SLA splits from the 2006 Census data to apportion labour force status onto the 2007 age/sex population projections. The second step is to use the percentage point change in the *national* projections of labour force status by age by sex from the Commonwealth Treasury's IGR 2007 report to adjust the proportion of persons in each labour force category for every SLA. It should be noted here that the national growth trend has been applied to each SLA, in the absence of any SLA-

specific labour force projections.

In this first attempt at projecting the benchmarks, the labour force by age by sex table plays an important role in the projections of all the other benchmarks since it is the exogenous variable used to project the other benchmark tables. The projections for all the other benchmark tables are calculated using the relationship between the benchmark table and the labour force by age by sex table in the base year (2006). The coefficients used to project all the other benchmark tables are estimated using a log linear model:

$$Ln(PopBC) = f\left(\sum_{i=0 \, j=1 \, k=1}^{i=5 \, j=6 \, k=2} \beta_{ijk} Ln(PopLF_i Age_j Sx_k)\right) \tag{1}$$

Where $PopBC$ is the number of population in each benchmark table category while $PopLF_i Age_j Sx_k$ is the population in labour force status i, age j, and sex k. The estimation is done using a cross section regression with every SLA in Australia as an observation. Given that the estimate of β_{ijk} in equation (1) is the growth elasticity of the population in the benchmark table to the population in labour force status i, age j, and sex k, the population growth in each benchmark table can be projected as:

$$\frac{\Delta PopBC_{2006-T}}{PopBC_{2006}} = \sum_{i=0 \, j=1 \, k=1}^{i=5 \, j=6 \, k=2} \beta_{ijk} \frac{\Delta PopLF_i Age_j Sx_{k \, 2006-T}}{PopLF_i Age_j Sx_{k \, 2006}} \tag{2}$$

The estimation in equation (2) will give us the estimated growth and hence the estimated number of every category's population in the benchmark tables for any year into the future. Note that all the financial data has been kept in 2006 prices, so we haven't inflated rents, mortgages, incomes, etc. What we are projecting is the number of people in each income category; or the number of people in each rent category. So the categories stay the same each year; only the number of people in each category changes.

To derive reasonable estimates from Equation 2, the total number of people or households in each benchmark table must be the same. In many cases (due to the ABS' randomisation rule), these totals are not the same. Therefore, the number of people or households in each table is adjusted so the totals are the same across all benchmark tables. This adjustment process takes one table as having the correct number, and then adjusts all the other tables so they match this first table. In this paper, the priority is the same as it is in Table 1; so there is an assumption that benchmark table 1 has the correct total for number of people; and benchmark table 2 has the correct number for total number of households. All other tables are then adjusted to match the totals in these tables.

As in the base year (2006), the reweighting process uses an iterative constrained optimisation technique to calculate the weights for every household in the microdata for every projected year. One of the problems with using this technique is the loss of estimates for some SLAs because the iterative process failed to find the optimal solution given the constraints from the 11 benchmark tables.

The results from the reweighting process for the projected benchmarks shows that the further the model is projecting out, the more SLAs fail to converge. In

the base year of SpatialMSM/08C, there are 138 out of 1422 SLAs that did not converge. The number of non converging SLAs increases to 157 out of 1415 SLAs in the 2010 projection, and increases further to 208 SLAs and 236 SLAs in the 2020 and 2027 projections, respectively. Table 3 shows that besides the Australian Capital Territory and Northern Territory, most of the additional SLAs that fail to fulfil the TAE criteria are non capital city SLAs.

Table 3. Number of SLAs dropped due to failed TAE in the projections

Major Statistical Region (MSR)	SLAs with failed TAE in SpatialMSM/08c	Total SLAs Projected	SLAs with failed TAE in 2010 projection	SLAs with failed TAE in 2020 projection	SLAs with failed TAE in 2027 projection
Sydney	1	64	0	0	0
NSW-Balance of State	1	135	2	10	15
Melbourne	0	79	2	2	2
VIC-Balance of State	4	130	7	14	25
Brisbane	3	215	7	6	8
QLD-Balance of State	40	263	40	48	46
Adelaide	0	55	0	0	0
SA-Balance of State	7	72	10	18	20
Perth	2	37	2	1	2
WA-Balance of State	15	118	17	24	27
Hobart	0	8	1	1	1
TAS-Balance of State	1	35	2	3	3
Darwin	6	41	6	10	12
NT-Balance of State	42	54	43	44	43
Canberra	15	108	17	26	31
ACT-Balance of State	1	1	1	1	1
Australia	138	1415	157	208	236

Source: SpatialMSM/08C projections

Losing 236 of the 1415 SLAs in the 2027 projection is still considered as acceptable for the purposes of this study, since these SLAs only contain 2.8 per cent of the whole population (Table 4). It should be noted, however, that around one-quarter to one-third of the Australian Capital Territory and Northern Territory populations live in SLAs which fail our TAE test in 2027, so projections for the two territories must be treated with caution. A special note, however, needs to be given to Queensland that has substantially more SLAs than New South Wales and Victoria. It was notable that around 18 per cent of the SLAs outside Brisbane failed the TAE test in the projections. This requires further investigation and may be related to the relative high (5.1 percent) Census undercount outside Brisbane in 2006.

Table 4. Number of SLAs dropped due to failed TAE criteria in the 2027 projection

State/ Territory	SLAs with failed TAE	Total SLAs	Per cent of SLAs with failed TAE (%)	Per cent of all persons living in SLAs with failed TAE (%)
NSW	15	199	7.5	1.6
VIC	27	209	12.9	2.6
QLD	54	478	11.3	2.3
SA	20	127	15.7	3.4
WA	29	155	18.7	1.6
TAS	4	43	9.3	2.5
NT	55	95	57.9	32.5
ACT	32	109	29.4	24.7
Australia	236	1415	16.7	2.8

Source: SpatialMSM/08C projections

4.2 Reliability of the Projections

After the weights for future years are produced, the next step is to check the reliability of the estimation using this set of future weights. The validation process is the step that is commonly used to check the reliability of the spatial microsimulation modelling.

There are two sources of model error in our projections. One comes from the projections of each benchmark table; so it is to do with the reliability of the coefficient β_{ijk} in Equation 1. The second source of error is in the generalised regression routine that will reweight the survey data to the projected benchmarks.

In terms of the first source of model error, if the Age by Sex by Labour Force projections are not very good at estimating our other benchmarks, then the estimated weights for the projections will not be accurate and the projections will be unreliable.

The estimate of the size of the errors in the forecasting of the benchmarks can be looked at using the coefficient of determination (R^2) of the regression process that produces the elasticity coefficients (Equation 1). This figure will show how much variation in the benchmark table in the base year can be explained by the age by sex by labour force structure. As the regression was done separately, each category in each benchmark table has it's own R^2. However, to simplify the analysis the means of the R^2 in the benchmark tables will be presented. The range of R^2 values will also be given to give a better idea as to the reliability.

Looking at Table 5, the R^2 indicate that most of the variation in the original tables can be explained by the Age by Sex by Labour force status table. This means that projections of these benchmark tables using a coefficient calculated in the base year, while not perfect, would be reasonable as a first attempt at

projecting the base microdata. Further work could enhance these projections, and one option may be to introduce some historical time series where the projections are particularly bad (as has been done for SimBritain – see Ballas et al, 2005a), but for most of the benchmarks, the age by sex by labour force status table explained on average more than 70 percent of the variation in the other tables. However, there are 3 tables where the average R^2 was below 70 percent, which are tenure by weekly household rent, monthly household mortgage by weekly household income, and weekly household rent by weekly household income. These would be the first tables that further work could be conducted on getting better projections.

Table 5. R^2 for benchmarks used in the reweighting algorithm

Table No.	Benchmark Table	Lowest R^2	Highest R^2	Mean R^2
2	Total number of households by dwelling type (Occupied private dwelling/Non private dwelling)	0.542	0.993	0.767
3	Tenure by weekly household rent	0.424	0.862	0.635
4	Tenure by household type	0.516	0.984	0.826
5	Dwelling structure by household family composition	0.386	0.975	0.706
6	Number of adults usually resident in household	0.952	0.995	0.971
7	Number of kids usually resident in household	0.957	0.997	0.977
8	Monthly household mortgage by weekly household income	0.176	0.928	0.643
9	Persons in non-private dwelling	0.295	0.719	0.420
10	Tenure type by weekly household income	0.428	0.977	0.760
11	Weekly household rent by weekly household income	0.136	0.825	0.598

Source: authors' calculations

In conclusion, on the basis of the R^2 for the model in Equation 1, it is considered that the projected benchmarks were reliable enough to use in the reweighting process.

The second set of validation tests check the accuracy of the estimated projections against a projected variable that is not benchmarked, but is available from the small area projections we have. In our case, the number of children aged 3 and 4 years is not benchmarked (we benchmark the number of children

aged 0 – 17 years), can be estimated from our model, and is available from the age/sex projections.

One of indicators of accuracy that has been developed for the validation process uses called the measure of accuracy (Miranti et al, 2008). This is essentially the dispersion of the estimated SLAs around the more reliable number from ABS publication or administrative data where the definition used has exactly the same definition. So measure of accuracy or MA is calculated as:

$$MA = 1 - \frac{\sum (y_{est} - y_{ABS})^2}{\sum (y_{ABS} - \bar{y}_{ABS})^2} \qquad (3)$$

Where

$MA = measure\ of\ accuracy$

$y_{est} = estimated\ number\ from\ spatial\ microsimulation$

$y_{ABS} = estimated\ number\ from\ the\ ABS$

$\bar{y}_{ABS} = mean\ estimates\ of\ the\ ABS'\ number$

The formula of this measurement is similar to the formula for the coefficient of determinant or R^2 in a regression model, which also calculates the dispersion of the estimated value from the regression to the actual data.

The measure of accuracy for the base year (2006) is 99.0 per cent for the number of children aged 3-4, so we get an excellent result for the base year. The measure of accuracy for the projection in 2027 is 95.1 per cent. This shows that our modelled projected data match very well to the DOHA population projections.

5. APPLYING THE PROJECTION AND SCENARIO BUILDING

As mentioned in the introduction, these spatial microsimulation projections are built to assist planning agencies such as government by providing information about the characteristics of individuals and households in certain small areas in the future. This information can then be used to anticipate the need for resource allocation for each small area in the future. Nevertheless, the information provided by these projections is based on strict assumptions about the long term projections of the benchmarks and maintaining the socio-economic structure and relationship that exists in 2006. These assumptions may not prevail and a good projection model should be ready to supply alternative future scenarios.

Building a new scenario for a projection is undertaken by altering the assumptions that are used in the base projection. The scenario built will adjust the future socio-economic conditions based on different assumptions about long term expectations or the socio-economic conditions in 2006.

5.1 Projections of the base scenario

Without changing any assumptions or data in the model, it can provide useful information for policy makers on projections of populations who may demand certain types of services in the future. For example, we would expect families

where there are young children (below school age) and where all parents are working to require childcare services. So an estimate of the number of children aged 3 – 4 where both parents are working may give policy makers in a State some idea on where to locate child care centres.

A researcher may assume that the number of children aged 3 – 4 is a reasonable proxy for the number of children aged 3 – 4 with all parents working. What this section shows is the danger of using these simple proxies.

An estimate of the number of children age 3 and 4 years who have all their parents working in 2027 is produced by applying the record unit data from the 2002-03 and 2003-04 SIH-CURFs to the projected small area weights from the reweighting process. The variable representing the number of children aged 3 and 4 in a household from the survey is combined with person level data on the employment status of all people in the household. This allows us to calculate the number of children aged 3 and 4 where all parents are employed. Given that the spatial microsimulation process calculates weights at a household level, the number of children aged 3-4 in a household where all parents are working is multiplied by the small area weight for each household.

The question that we are trying to answer for the service providers is the demand for child care services in each area. What we have from the DOHA population projections is projections of the number of children aged 3 – 4 in small areas, but not all families will require childcare services. The demand for child care services will also depend on who is working in the family.

When we estimate the number of children with all parents working, we find a correlation with the number of children aged 3 – 4 of 0.51 (see Figure 1). This does suggest that our spatial microsimulation results, which add the criteria of all parents working, make a significant difference – so the number of children aged 3 – 4 is not a very good proxy for the demand for childcare places in an area. Other variables such as labour force status and family structure play an important part in determining the number of children aged 3-4 with all parents working, and only the spatial microsimulation model can add these criteria to the projections.

Analysing the spatial pattern of children aged 3 – 4 with all parents working is another way of examining whether this new variable adds any further information. Figure 2 presents 4 maps. Map A and B show the growth in the projected number of children aged 3-4 years from the DOHA projections and the growth in the projection of children aged 3-4 years with all parents working from SpatialMSM. The classes in the map are distributed using natural breaks and the darkest colour shows the highest estimate of growth.

These maps show that there are several SLAs on the western outskirts of Sydney where the growth in the number of children aged 3-4 with all parents working is particularly high compared to the growth in the 3-4 year old population. Liverpool-West, Blacktown-South-West, and Fairfield-West are among those SLAs. These areas may have found a significant lack of childcare places in 2027 if the estimates of children aged 3 – 4 were used to show where future childcare places should be allocated. A further investigation of this issue is discussed in Harding et al (2009).
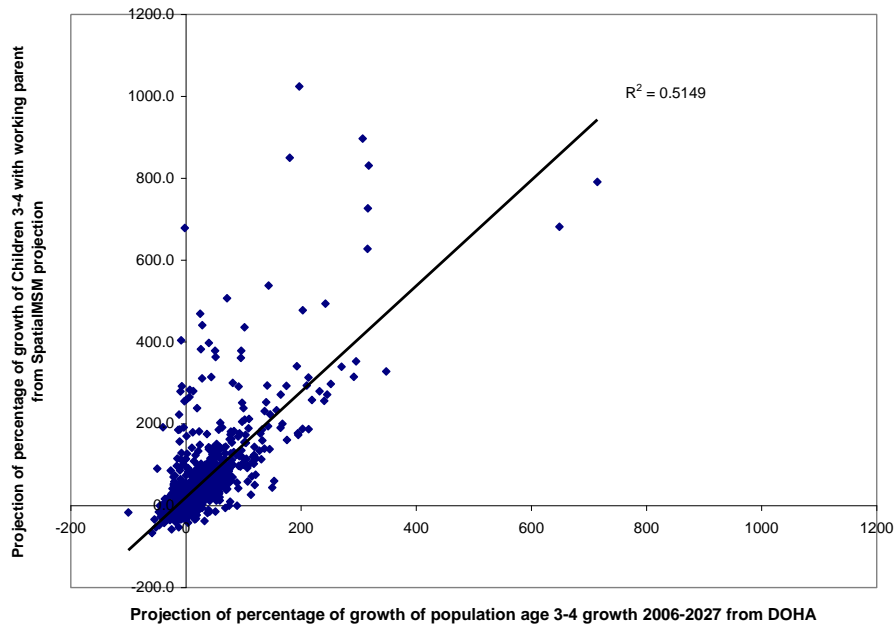
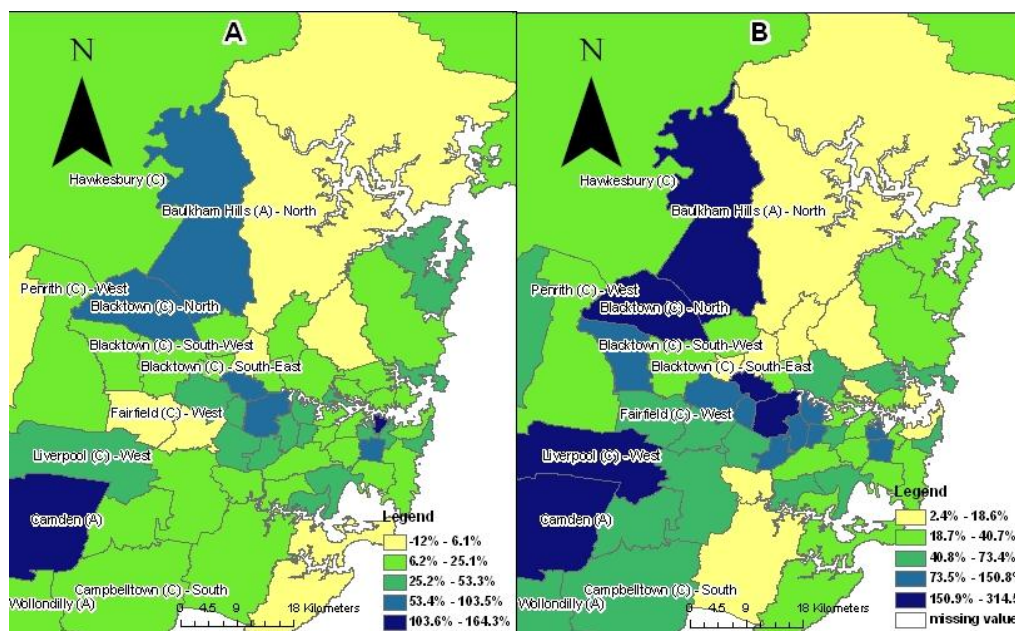**Source:** SpatialMSM Projections, DOHA Population projections

Figure 1. Comparison of projections of children aged 3 – 4 (DOHA) and projections of children aged 3 – 4 with all parents working (SpatialMSM)

5.2 Building a Scenario

The second type of analysis we can do with this microsimulation model is to change some of the assumptions, and build scenarios that then affect the final weights, and the projections. The base model will give an indication of what the future will be like given certain.assumptions. However, no one really knows what will happen in the future, and whether the conditions that become the basis for the base projections will prevail. Therefore, the ability to build a scenario to anticipate different assumptions allows planning agencies such as the government to formulate an alternative plan given different assumptions about the future.

Given that the projection methodology outlined in this paper is mainly built on the projection of benchmark tables, any new scenarios also have to be built by altering the benchmark tables. Looking back at Section 4.1 of this paper, it can be seen that there are two steps in projecting the benchmark tables. The first step is the logistic regression using age, sex and labour force status that projects the benchmark tables forward; and the second step is reweighting the survey data to the new Census benchmarks.

Notes: A: DOHA projection of Growth in 3-4 year old 2006-2027; B: SpatialMSM projection of growth in 3-4 year old with working parent 2006-2027

Source: DOHA Population projection, SpatialMSM projection.

Figure 2. Projected spatial pattern of children aged 3-4 years with working parents compared to children aged 3 – 4 years

As a consequence, any scenario has to be implemented in either of these two steps. There is a major difference between implementing a scenario in the first step and implementing one in the second step. The introduction of a new scenario in the first step means changing the labour force structure, age structure, sex structure or a combination of those variables. These changes will also affect all other benchmark tables since the projections for these tables are made using projections of Age, Sex and Labour Force Status (Figure 3). The changes made to the Age/Sex/Labour force status projections will flow through to each benchmark table through the logistic regression model shown in Section 4.1.

Introducing a change into the second step of the projection process involves identifying every table that could be affected by the proposed change, and then making changes to those tables. The example shown in this paper is a change to housing tenure, so modelling a trend out of purchasing houses and into private rental, possibly because house prices have increased or there is a societal shift away from purchasing houses in Australia and towards renting houses, due to labour mobility. This is only one scenario that could be modelled – in theory, any scenario can be modelled, but different scenarios will affect different tables, so some thought has to be put into which tables are affected, and how they are

affected.
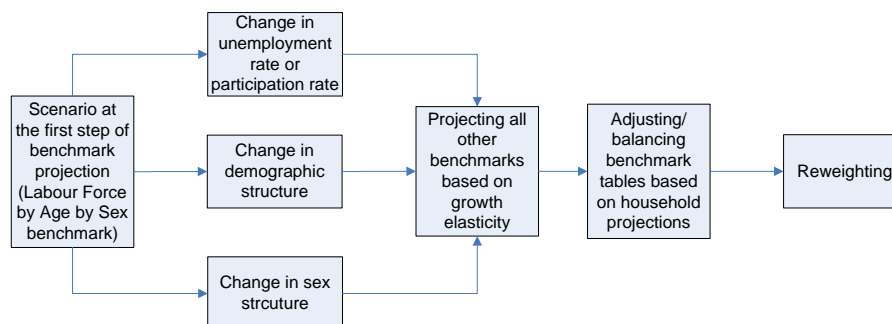
Figure 3. Building a scenario into the first step of the projection process

In this case, Figure 4 shows how this scenario can be built in the second step of the projection process. Looking at Figure 4, the proposed scenario means that the proportion of private renters in the "Tenure by Household Type" table, the "Tenure Type by Weekly Household income" and the "Tenure by Weekly Household Rent" tables should be increased. Because we don't want to change the total population, and we are modelling people moving from purchasing to renting, a change to the number of renters will increase the number of people paying rent in the "Weekly Household Rent by Weekly Household Income" table and decrease the number of purchasers in the "Monthly Household Mortgage by Weekly Household Income" table.

Note that we could also assume that 90 percent of the new renters were previously purchasers; and 10 per cent were previously some other tenure (like public housing or employer provided housing). So we don't have to assume that all the new renters were previously purchasers – we can make this scenario as complicated as we need to.

It can be seen that the effect of changing one variable can be quite complicated, and the changes need to be made explicitly to each of the benchmark tables, requiring some thinking about the secondary effects of any scenario. However, because we are making changes to each benchmark table, the scenarios can be as simple or as complicated as we need.

Because the reweighting algorithm is re-run, there will also be a different number of areas dropped due to not meeting the TAE criteria. The first group of scenario changes modelled implemented a change to the unemployment rate, implemented in the first step of the projection process. Three different scenarios for unemployment were used to test the stability of the model. One of these is the base scenario, where the change to unemployment is the national change as projected in the Inter-Generational Report. The second scenario introduces a two percentage point increase in unemployment for every SLA, while the third scenario uses the unemployment rates from 2006 (so the unemployment rate remains unchanged over the projection years).
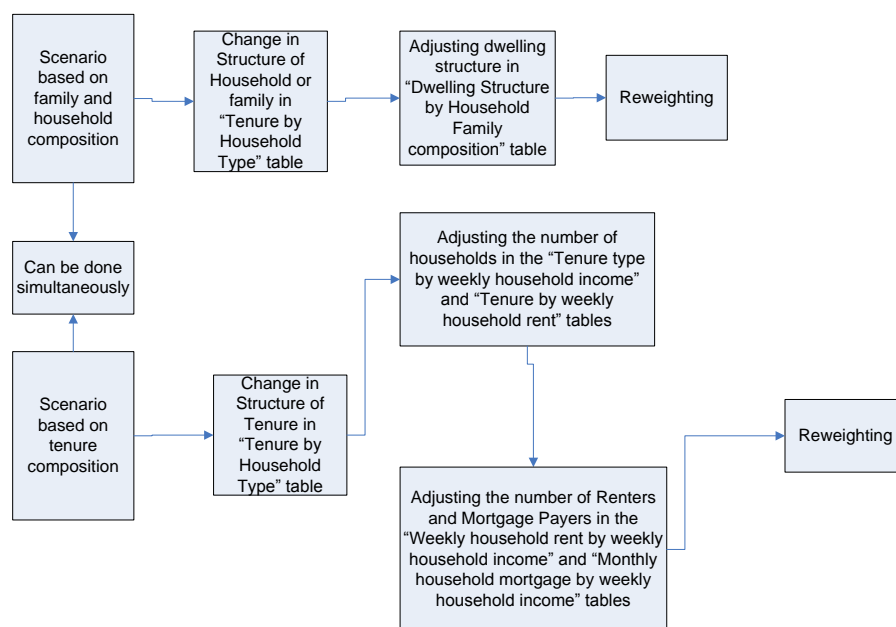
Figure 4. Building a scenario at the second step of the projection process

A change in unemployment is chosen for two reasons. First, the unemployment rate is a good indicator of whether the economy growing or shrinking, so changing the unemployment rate can allow us to simulate a better or worse economy out to the future. Second, changing unemployment impacts a number of benchmark tables, as shown in Figure 3, so any instability in the model should be clearly shown.

Table 6 shows that the model is more stable in the capital cities when a change is made to the unemployment rate. As can be seen, increasing the unemployment rate by two percentage points for all SLAs has caused more SLAs to fail the TAE test in NSW-Balance of State, Victoria-Balance of State and Queensland-Balance of State than in the capital cities, where a maximum of three additional SLAs failed the test. This may also confirm the earlier analysis that the projection model itself is not as stable in non capital city SLAs, as shown in Table 3 of Section 4.1. Furthermore, the scenario using the 2006 unemployment rate came up with slightly fewer SLAs that failed the TAE, which shows that the closer the scenario is to the 2006 data, the fewer the number of SLAs that will fail the TAE test. However, the difference between this scenario and using the IGR projected unemployment rates is small, so it may also be due to the fact that the IGR does not predict a major change in unemployment.

Table 6. Number of SLAs dropped due to failed accuracy criteria in 2027 projection

Major Statistical Region (MSR)	Total SLAs Projected	SLAs with failed TAE in 2027 (with IGR)	SLAs with failed TAE in 2027 with a 2 pct. point unemployment increase from the base	SLAs with failed TAE in 2027 if the 2006 unemployment rate is used
Sydney	64	0	1	1
NSW-Balance of State	135	15	35	12
Melbourne	79	2	3	4
VIC-Balance of State	130	25	38	20
Brisbane	215	8	11	8
QLD-Balance of State	263	46	62	47
Adelaide	55	0	0	0
SA-Balance of State	72	20	25	18
Perth	37	2	2	2
WA-Balance of State	118	27	33	24
Hobart	8	1	1	1
TAS-Balance of State	35	3	7	4
Darwin	41	12	12	11
NT-Balance of State	54	43	44	42
Canberra	108	31	30	33
ACT-Balance of State	1	1	1	1
Australia	1415	236	305	228

Source: SpatialMSM/08C projections

6. STRENGTHS, WEAKNESSES AND THE WAY FORWARD

Sections 2 to 5 of this paper have revealed how statistics for small areas in Australia can be estimated and projected using the SpatialMSM model. In this section, we sum up the strengths and weaknesses of this projection model. We will also look at the way forward for this projection model.

6.1 Strengths and Weaknesses

The main strength of this projection model is the ability to provide a picture of the household composition and future conditions according to assumptions given by other models, such as population projections from the Australian

Bureau of Statistics and labour force projections from the Inter-Generational Report. Demographic and labour force projections are the main determinants of the projections from SpatialMSM and there are models that can produce these projections using various assumptions about fertility, mortality, and migration for the population projections, and different economic projections for the unemployment rate. So it is easy to bring in new scenarios for population growth and a change in the labour force status. Nevertheless, it is also important to note that the interpretation and the performance of this model is highly dependent on the assumptions underlying the projections.

Another strength of this projection model is the possibility of altering any of the variables in the benchmark tables. Although initially projected by the growth in the population by age, sex and labour force status based on the elasticities in 2006, it is possible to alter any of the benchmark tables, with some care given to which tables are affected by the new scenario. The example used in this paper is a change to the housing tenure driven by people moving from home ownership to renting. The effect of the scenario on the benchmark tables can be as complicated as required.

The next strength of this model is the independence of each SLA in the model. This means that each SLA can have a scenario change applied separately and as long as the SLA does not fail the TAE criteria, then the model can provide projections for just that SLA. However, this feature is also one of the weaknesses of this projection method, as SLAs may interact through population movement, especially if unemployment rates are changed in one SLA; and this population movement is not modelled (although it could be in the future through a dynamic model).

The main weakness of the model is the fact that the projection relies on the relationship between the labour force by age by sex composition and the composition of the other benchmark tables based on the 2006 population census. This is a reasonable assumption if the model projects into the near future, but may be unreasonable for a long term projection. Any change in personal preferences could make this assumption invalid. For example, one change modelled in this paper is people preferring to rent instead of buying their own house in the future, due to labour mobility. As a result, even if there are no changes in the structure of the labour force by age by sex, the number of people who live in rental dwellings may still be increasing.

Not only are the benchmarks based on 2006 Census data projected forward, the survey data used is from 2002/03 and 2003/04. There is a strong possibility that these data do not represent individual households in the long term. Cassells and Harding (2007) show that the generation born between 1976 and 1991 (generation Y) has different characteristic to the previous generation in terms of working and having families, which are two variables we benchmark to.

Because this is a static microsimulation model, we are not ageing the population at all; we are just benchmarking this 2002/03 and 2003/04 data to future projections. So in 2002, this generation (Gen Y) is aged between 11 and 26. The characteristics in the benchmarks for this age group are projected into the future, and then people aged 11 to 26 in 2027 will be benchmarked to these

tables. So we are applying the GenY characteristics to people aged 11 to 26 in 2027. But people aged 11 to 26 in 2027 may be very different from the GenY group in 2002. This also works the other way. So the GenY group from 2002 will be aged between 36 and 51 in 2027, and their characteristics may be very different from people aged between 36 and 51 in 2002.

Again, the flexibility of this model means we could assume some other preferences for this group in 2027, and adjust the benchmark tables using some behavioural model; but we really have no information on what preferences these people will have in 2027. So using the preferences from 2006 may be the best information we have.

6.2 The Way Forward

One of the limitations of this model is that it is a static model, so there is no dynamic ageing process. Making the model more dynamic is a clear way forward. The simple static ageing procedure employed in this model utilises the correlation between the labour force by age by sex status and other socio-economic variables in 2006 to create projected benchmark tables for the model. The model then uses the unit record data from the ABS SIH 2002/03 and 2003/04 to populate the small area given these new projected benchmarks. By doing this, the model may fail to capture any trend that changes the relationship between labour force by age by sex and other socio-economic variables in the future. Furthermore, using unit record data with 2002-2004 characteristics may also give false projections if there is a generational trend that alters the characteristics of households in the future. In theory, this affects any projection model – nobody really knows how these generational changes will develop in the future.

There are two steps that we think may improve this model. The first may be to capture and induce a long term trend in the benchmark table projection process. While this would give a more accurate picture of any change over time (for instance, a long term move from purchasing to renting that could be carried on in the projections), there are problems with it. One is that it would be done for every SLA, so it may just be picking up a local short term trend that may not continue into the future. This could be ameliorated by looking at national trends and applying these trends to the small areas; however, we are then ignoring local effects. So there is a balance between these two that would need to be considered before implementing a time trend into the projected benchmarks. The other problem is that the benchmark tables are from the Australian Census, which is conducted every five years; and the small areas and the data definitions change every Census. So creating a comparable time series of the benchmark tables using Census data is going to be difficult.

The alternative is to use a simple shift share for the growth projections such as used in SimBritain (Ballas et al, 2005a). This uses linear exponential proportional smoothing of 1971, 1981, and 1991 data to project the constraint for 2001, 2011 and 2021. Again, this would be difficult to apply for each SLA in Australia because in every Census, the SLA boundaries and some data definitions can change, but it may be possible to project State aggregates and

then redistribute the projected proportions back to SLAs.

The second way to improve this model is to update the unit record data so they become more representative of future conditions.   This could be implemented by making the model a dynamic microsimulation model, so individually updating the characteristics of each individual and family contained within the model for each time period. In Australia, the Australian Population and Policy Simulation model (APPSIM) has been developed to update unit record characteristics (Cassells et al., 2007). However, such dynamic population microsimulation models involve a very high degree of complexity and cost (Harding, 2007). In addition, there would also be problems getting appropriate longitudinal data to estimate the relevant transition probabilities at a small area level.

So there are significant barriers to either of these modifications, although not insurmountable.   Further, these two modifications do not have to be applied simultaneously. In the model as it currently stands, it is possible to apply the first change without having the unit record data updated, and continue to use the model with the limitation that the underlying survey datasets are not updated.

Converting the model to a dynamic model is a much larger step, as the dynamic process needs to be modelled for every SLA.   However, using this process, projections could be derived without using the reweighting process to age the population, as the population is dynamically aged.   The reweighting process could still be used for aligning the dynamically aged survey data to external benchmarks from the Census, but the change made would be minimal. So the totals would match the Census benchmarks due to the alignment; but the relationships between variables may have changed because of the dynamic ageing process.

7. CONCLUSIONS

This paper has given an overview of a model that can address not only the need for small area information for the present, but also for the future. In the past decade, this need has become more and more apparent as planning agencies in Australia (such as its local and federal governments) need to focus on service delivery for local areas given the characteristics of individuals and households in those areas. The paper started with the current spatial microsimulation model in Australia named SpatialMSM/08C and then described the first attempt to develop projections from this model.

A static ageing process is the approach taken in developing the projection model given the very high degree of complexity, cost and data requirements in building a fully dynamic microsimulation model. The static ageing model is undertaken by employing the currently available population and labour force projections to estimate the various constraint tables used in SpatialMSM/08C. The model then uses the reweighting process in SpatialMSM/08C to reweight the microdata or unit record data according to the projected constraints.

As this paper has shown, the model has been able to produce information for small area planning into the future with a reasonable degree of reliability. The model is also able to take some simple scenarios to model some changes in the

future, and seems to be most reliable for capital cities. Nevertheless, the static ageing approach that the model uses means that it is difficult to model any behavioural change, without identifying the effect of the behavioural change and implementing this in the benchmark tables. Further, while we have not tested this, we expect that any large changes in the characteristics of the society in the future will be difficult to estimate, as the large changes in the benchmarks will mean the reweighting process will fail to find reasonable weights for a high proportion of areas.

This has led to some potential improvements to the model that we have considered, and two steps have been identified. The first is to explicitly acknowledge the long term trend of socio-economic changes in society while the second step is to use a dynamic microsimulation method to update the unit record data into the future. Both these steps have problems that would need to be resolved, but the problems are not insurmountable and could be the subject of future research.

ACKNOWLEDGMENTS

REFERENCES

ABS (2007) Australian Standard Geographical Classification (ASGC), 1216.0, Australian Bureau of Statistics.

ABS (2008a) Consumer Price Index, Australia, December 2007, Table 13: CPI Groups, Sub-Groups and Expenditure Class, Index numbers by capital city, 6401.0, Australian Bureau of Statistics.

ABS (2008b) Population Projections, Australia 2006 to 2101, TABLE B9. Cat. No. 3222.0 http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3222.02006%20to%202101?OpenDocument

Ballas D., Clarke G., Dorling D., Rigby J., Wheeler B. (2006a) Using geographical information systems and spatial microsimulation for the analysis of health inequalities. *Health Informatics Journal*, 12, pp. 65-79

Ballas, D., Clarke, G. and Dewhurst, J. (2006b) Modelling the Socio-Economic Impacts of Major Job Loss or Gain at the Local level: A Spatial Microsimulation Framework. *Spatial Economic Analysis*, 1(1) pp. 127-146

Ballas, D., Clarke, G. and Weimers (2005b) Building a Dynamic Spatial Microsimulation Model for Ireland. *Population, Space and Place*, 11, pp. 157-172

Ballas, D., Rossiter, D., Thomas, B., Clarke, G. and Dorling, D. (2005a) Geography Matters: Simulating the Local Impacts of National Social Policies. Joseph Rowntree Foundation: York.

Bell, P. (2000) *GREGWT and TABLE macros - Users guide*, Unpublished, Australian Bureau of Statistics.

Birkin M, Clarke G., Clarke M. (1996) Urban and regional modelling at the microscale in Clarke (eds.), *Microsimulation for Urban and Regional Policy Analysis*. Pion: London, pp. 10–27.

Birkin M, and Clarke M. (1988) SYNTHESIS – a synthetic spatial information system for urban and regional analysis: methods and examples. *Environment and Planning A,* 20, pp. 1645–1671

Birkin M, and Clarke M. (1989) The generation of individual and household incomes at the small area level using Synthesis. *Regional Studies,* 23, pp. 535–548.

Caldwell S. B. (1990) Static, dynamic and mixed microsimulation, Dept. of Sociology, Cornell University, Ithaca, New York.

Cassells, R. and Harding, A. (2007) Generation whY?, *AMP NATSEM Income and Wealth Report* , Issue 17.

Cassells, R., Kelly, S., and Harding, A. (2007) Problems and Prospects for Dynamic Microsimulation: A Review and Lessons for APPSIM, Online Discussion Paper - DP63, available at http://www.canberra.edu.au/centres/natsem/publications?sq_content_src=%2BdXJsPWh0dHAlM0ElMkYlMkZ6aWJvLndpbi5jYW5iZXJyYS5lZHUuYXUlMkZuYXRzZW0lMkZpbmRleC5waHAlM0Ztb2RlJTNEcHVibGljYXRpb24lMjZwdWJsaWNhdGlvbiUzRDk0MyZhbGw9MQ%3D%3D

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

Chin, S.F. and Harding, A. (2007) SpatialMSM – NATSEM's Small Area Household Model of Australia. In Harding, A and Gupta, A. (eds) *Modelling Our Future: Population Ageing, Health and Aged Care*, International Symposia in Economic Theory and Econometrics. North Holland: Amsterdam.

Chin, S.F., Harding, A. and Bill, A. (2006) Regional Dimensions: Preparation of the 1998-99 Household Expenditure Survey for Reweighting to Small-area Benchmarks, Technical Paper no. 34, NATSEM, University of Canberra, Canberra.

Chin, S.F., Harding, A., Lloyd, R., McNamara, J., Phillips, B. and Vu, Q.N. (2005) Spatial microsimulation using synthetic small-area estimates of income, tax and social security benefits. *Australasian Journal of Regional Studies*, 11(3), pp. 303-336

Clarke, G. (1996), *Microsimulation for Urban and Regional Policy Analysis*, (ed) , Pion: London.

Eason, R. (1996) Microsimulation for direct taxes and fiscal policy in the United Kingdom. In A Harding (Eds.), *Microsimulation and Public Policy*. North Holland: Amsterdam.

Gupta, A. and Kapur, V. (2000) *Microsimulation in Government Policy and Forecasting*. North Holland: Amsterdam.

Gupta, A. and Kapur, V. (1996) Microsimulation Modelling Experience at the Canadian Department of Finance. In A Harding (Eds.), *Microsimulation and Public Policy*. North Holland: Amsterdam.

Harding, A and Gupta, A. (2007a) Introduction and Overview. In Harding, A and Gupta, A. (Eds), *Modelling Our Future: Population Ageing, Social Security and Taxation* , International Symposia in Economic Theory and Econometrics. North Holland: Amsterdam.

Harding, A and Gupta, A. (2007b) *Modelling Our Future: Population Ageing, Social Security and Taxation.* International Symposia in Economic Theory and Econometrics. North Holland: Amsterdam.

Harding, A. (2007) Challenges and Opportunities of Dynamic Microsimulation Modelling. Plenary paper presented to the 1st General Conference of the International Microsimulation Association, Vienna, 20-22 August.

Harding, A., 1996, *Microsimulation and Public Policy*, Contributions to Economic Analysis Series, North Holland, Amsterdam.

Harding A., Vidyattama, Y., Tanton, R. (2009) Population Ageing and the Needs-Based Planning of Government Services: An application of spatial microsimulation in Australia', 2nd General Conference of the International Microsimulation Association, Ottawa, Canada, 8-10 June.

Harding A., Vu, Q. N., Tanton, R., Vidyattama, Y. (2009b) Improving work incentives for mothers: the national and geographic impact of liberalising the Family Tax Benefit income test. *The Economic Record*, 85 (Special Issue), pp. 48 - 58

Heady, P., Clarke, G.P., Brown, G., Ellis, K., Heasman, D., Hennell, S., Longhurst, J., Mitchell, B. (2003) Model-based small area estimation series no. 2: small area estimation project report, UK, Office for National Statistics.

Holm, E., Holme, K., Makila, K., Kauppi, M.M. and Mortvik, G. (2001) The SVERIGE spatial microsimulation model - content, validation, and example applications, Spatial Modelling Centre, Umeå University, Kiruna.

Hooimeijer P. (1996) A life-course approach to urban dynamics: state of the art in and research design for the Netherlands. In Clarke G. (eds.) *Microsimulation for Urban and Regional Policy Analysis*. Pion: London; pp. 28–63.

Lymer, S., Brown, L., Harding, A., Yap, M., Chin, SF. and Leicester, S. (2006) Development of CareMod/05, Technical paper no. 32, NATSEM, University of Canberra, Canberra.

Lymer, S., Brown, L., Yap, M. and Harding, A. (2008a) Regional disability estimates for New South Wales in 2001 using spatial microsimulation. *Applied Spatial Analysis and Policy*, 1(2), pp. 99-116.

Lymer, S., Brown, L., Harding, A. and Yap, M. (2008b) Predicting the need for aged care services at the small area level: the CAREMOD spatial microsimulation model, *International Journal of Microsimulation*.

McNamara, J., Tanton, R. and Phillips, B. (2007) The regional impact of housing costs and assistance on financial disadvantage: final report, Australian Housing and Urban Research Institute, Melbourne, Australia.

Melhuish, T., Blake, M. and Day, S. (2002) An evaluation of synthetic household populations for Census collection districts created using Spatial Microsimulation techniques, 26th Australian and New Zealand Regional Science Association International (ANZRSAI) Annual Conference, Gold Coast, Queensland, Australia, 29 September - 2 October.

Miranti, R., McNamara, J., Tanton, R. and Harding, A. (2008) Poverty at the local level: National and small area poverty estimates by family type for Australia in 2006, paper presented at the Creating Socio-economic Data for Small Areas: Methods and Outcomes Workshop, University of Canberra, Canberra.

Mitton, L., Sutherland, H. and Weeks, M. (2000). *Microsimulation Modelling for Policy Analysis*. Cambridge University Press: Cambridge

Neary, J.P. (2001) Of hype and hyperbola: introducing the new economic geography. *Journal of Economic Literature* 39 (2), pp. 536–61.

O'Donoghue, C. (2001) Dynamic microsimulation: a methodological survey. *Brazilian Electronic Journal of Economics*, 4(2) [on-line journal]. Paper available on-line from: http://www.microsimulation.org/IMA/BEJE/BEJE_4_2_2.pdf.

Orcutt G. (1957) A new type of socio-economic system. *Review of Economics and Statistics*, 58, pp 773-797.

Orcutt, G., Merz, J. and Quinke, H. (1986). *Microanalytic Simulation Models to Support Social and Financial Policy.* North-Holland: Amsterdam.

Phillips, B. (2007). Customer Service Projection Model (CuSP): A Regional Microsimulation Model of Centrelink Customers. In Gupta, A. and Harding,

A. (eds.), *Modelling Our Future: Population Ageing, Health and Aged Care*, International Symposia in Economic Theory and Econometrics, North Holland: Amsterdam.

Procter, K. (2007) How where we live influences obesity: a geo-demographic classification of obesogenic environments using spatial microsimulation modelling. Paper presented at the American Association of Geographers, San Francisco, 17-21 April.

Rahman, A. (2008) A review of small area estimation problems and methodological developments, online discussion paper, Online Discussion Paper   -   DP66 (http://www.canberra.edu.au/centres/natsem/publications?sq_content_src=% 2BdXJsPWh0dHAlM0ElMkYlMkZ6aWJvLndpbi5jYW5iZXJyYS5lZHUuY XUlMkZuYXRzZW0lMkZpbmRleC5waAlM0Ztb2RlJTNEcHVibGljYXR pb24lMjZwdWJsaWNhdGlvbiUzRDExNDImYWxsPTE%3D).

Tanton, R., Nepal, B., and Harding, A. (2008) Wherever I Lay My Debt, That's My Home: Trends in Housing Affordability and Housing Stress, 1995-96 to 2005-06, *AMP.NATSEM Income and Wealth Repor*t , Issue 19.

Tanton, R. McNamara, J. Harding, A. and Morrison, T. (2009) Small Area Poverty Estimates for Australia's Eastern Seaboard in 2006. In A Zaidi, A Harding and P Williamson, *New Frontiers in Microsimulation Modelling*, Ashgate: London , pp. 79-96

Treasury (2007) *Intergenerational Report 2007*, Australian Commonwealth Department   of   Treasury,   Canberra. http://www.treasury.gov.au/igr/IGR2007.asp

Vencatasawmy, C.P., Holm E., Rephann T., Esko J., Swan N, Ohman M., Astrom M., Alfredsson E., Holme K. and Siikavaara J. (1999) Building a spatial microsimulation model, SMC Internal Discussion Paper. Spatial Modelling Centre, Umeå University, Kiruna.

Voas, D. and Williamson, P. (2000) An Evaluation of the Combinatorial Optimisation Approach to the Creation of Synthetic Microdata. *International Journal of Population Geography*, 6, pp 349-366

Williamson, P., Birkin, B., and Rees, P.H. (1998) The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*, pp. 785-816.

Williamson, P. (1992) Community care policies for the elderly: a microsimulation approach, Unpublished PhD thesis, School of Geography, University of Leeds, Leeds.

Williamson, P. (2001) A Comparison of Synthetic Reconstruction and Combinatorial Optimisation Approaches to the Creation of Small-Area Microdata, Working Paper 2001/2, Population Microdata Unit, Department of Geography, University of Liverpool, Liverpool.

Wu, B.M., Birkin, M.H. and Rees, P.H. (2008) A spatial microsimulation model with student agents, *Computers. Environment and Urban Systems*, 32(6), pp. 440-453.