Small-area Estimation using a generalised regression procedure

A THESIS SUBMITTED AS PART OF A PHD BY PUBLICATION AT THE

UNIVERSITY OF CANBERRA

BY

ROBERT TANTON

MAY 2011

ABSTRACT

This thesis describes a process of small area estimation which has been used in Australia to develop estimates and projections of poverty and housing stress for small areas, and which has been linked to a Tax/Transfer microsimulation model to estimate the effects of a policy change for small areas.

The thesis consists of a literature review in the area of spatial microsimulation; the five journal articles as they are published; and then a concluding chapter. The model is described in detail in one journal article, and a number of applications are presented in three other journal articles. These articles also show how the model has been validated. A fifth article shows how the model has been tested and further developed in a number of ways.

The thesis concludes that the model developed is a useful addition to a microsimulation modelling toolkit, especially given the need to investigate social characteristics for small areas. The method can be used by Government's and researchers to derive not only small area estimates of a number of variables, but also small area effects of a policy change; and small area cross-tabulations for a number of variables. This flexibility of the spatial microsimulation means it is a very powerful approach to small area estimation, which can traditionally only provide point estimates for certain variables.

TABLE OF CONTENTS

INTRODUCTION

This thesis describes the application in Australia of a method for deriving small-area estimates for a number of variables. Small-area estimation methods are prominent in the Statistical literature (Pfeffermann, 2002), and most of the methods derive a point estimate for one variable – so, for instance, crop areas (Battese, Harter & Fuller, 1988) or poverty (Elbers, J. Lanjouw & P. Lanjouw, 2002). While these methods use some complex statistical methods, and the Elbers method has now been incorporated into a program called PovMap, the methods are limited to deriving point estimates of one variable only.

Spatial microsimulation techniques are an alternative to these methods that have been developed over the last few years. Microsimulation is a technique designed by Orcutt (Orcutt, 1957) that models characteristics for individuals. These characteristics can then be adjusted and aggregated to either cross tabulations or different areas. This microsimulation method has recently been applied to geographical analysis, and has been called spatial microsimulation.

The spatial microsimulation method outlined in the articles that make up this thesis is one that has been used to estimate a number of different variables (including poverty, housing stress and superannuation). It can be used to derive projections for small areas, and can be linked to other models to derive small-area estimates of policy change. It can also be used to derive cross tabulations, for instance, poverty by family type.

The method outlined in this thesis has been developed over a number of years at NATSEM, with the author being involved in considerable development of the model by adding a link to a tax/transfer microsimulation model, making the model more efficient, adding a number of new benchmarks to the model and assessing the reliability of the model. These are all documented in the articles that are being submitted for this thesis.

The research questions covered by this thesis are:

- Can a generalised regression method be used efficiently for spatial microsimulation;
- Can this model then be used for studying the small-area effects of a policy change;
- How accurate is this model, and does the model continue to be accurate as the complexity of the modelling increases?

The five articles in this thesis have all been either published in reputable peer-reviewed journals, or are about to be published in the case of one article. They show the development of the model in a number of ways, and represent a body of work that has taken spatial microsimulation forward in Australia and internationally.

This thesis consists of eight chapters. Chapter one is a literature review on the spatial microsimulation methods which has not been published before. Chapter 2 shows how the five articles have contributed to a significant development in the spatial microsimulation literature. The five articles are then included as separate chapters (Chapter 3 – 7), and then a concluding chapter (Chapter 8) sums up the contribution of the body of work and outlines some future directions.

CHAPTER 1: LITERATURE REVIEW

Like normal microsimulation models, spatial microsimulation methods can be either static or dynamic.

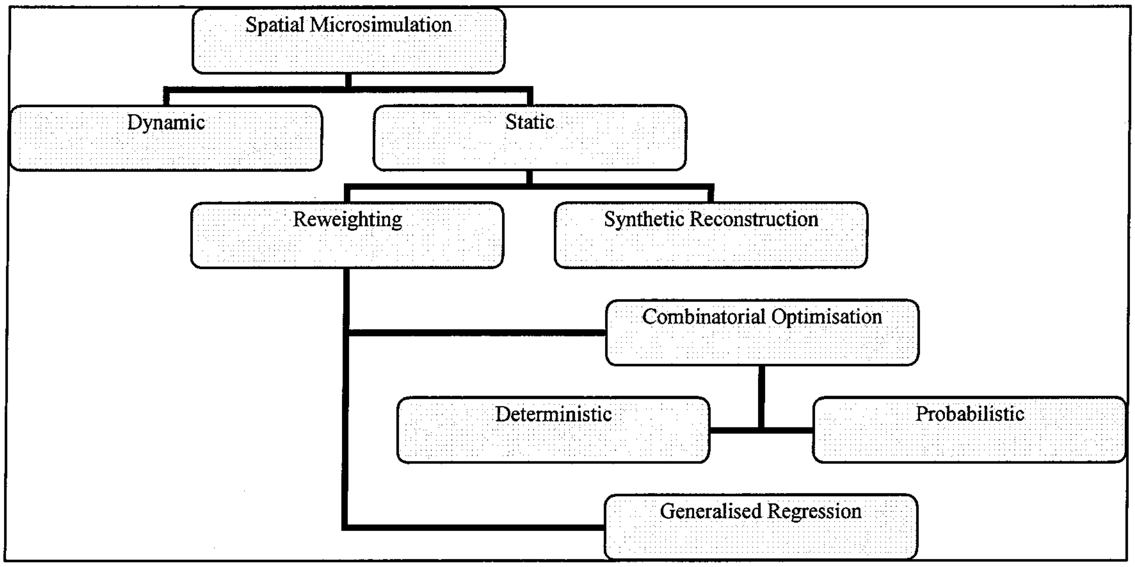
Dynamic microsimulation models take a base dataset, age this dataset over time, and model certain life events including marriage, deaths and births. They usually use probabilities to model these life events. They are similar to dynamic microsimulation models in that they model change over time, rather than a static model which only estimates the so called 'day after' effects (the effects the day after a policy has been implemented).

Static microsimulation models do not model life events, so the proportion of people married (and in fact, the people married on the base dataset) does not change. A static microsimulation model recalculates certain attributes, like incomes or eligibility for legal-aid services, using the unit record data on the base dataset. Static microsimulation models are usually used for rules based systems, where it is easy to use a rule to calculate pension eligibility or the amount of tax a household should pay based on a set of rules.

Both these types of models are also relevant for spatial microsimulation models, and this literature review will use these two categories to review the literature. Within each of these two types of models, there are then different methods to calculate the results. Under a static model, these two methods are a reweighting method and a synthetic reconstruction method.

These different methods are shown in Figure 1, and this section broadly outlines each of these methods and a number of models using each of the methods.

Figure 1: Methods of spatial microsimulation



HISTORY OF SPATIAL MICROSIMULATION

This section provides a brief history of spatial microsimulation, and the methods used. The rest of this chapter describes the methods in more detail, provides a comparison of the methods, validation and some applications.

While some early modelling attempts could be considered spatial microsimulation (Hagerstrand, 1952), one of the first spatial microsimulation models was a model for health-care planning developed by Clarke et al. (Clarke, Forte, Spowage & Wilson, 1984). The model, called HIPS (Health Information and Planning System), was developed for the British Health District Authorities. The model generated an initial population from aggregate data for each location. The demographics of this initial population were then updated each year.

Clarke was also involved in other papers on spatial microsimulation modelling, including a spatial microsimulation model developed with Birkin called Synthesis which used an iterative proportional fit method, as described later in this thesis (Birkin & Clarke, 1988; 1989; Clarke & Holm, 1987; Clarke & Wilson, 1985).

The next development in spatial microsimulation was a model developed by Clarke et al., which was developed to estimate demand for water (Clarke, Kashti, Mcdonald & Williamson, 1997; Williamson, Birkin & Rees, 1998). This model used a mathematical method called Combinatorial Optimisation, also described later in this thesis.

Around this period, there was a considerable amount of work being done using spatial microsimulation. It was being used for looking at regional changes in household incomes (Caldwell, Clarke & Keister, 1998); for population projections (Van Imhoff & Post, 1998), and for estimating household attributes (Ballas & Clarke, 1999; Ballas, Clarke & Turton, 1999; Williamson, Birkin & Rees, 1998). All these models were static spatial microsimulation models, and used either an iterative proportional fit method or a probabilistic combinatorial optimization method.

Other static spatial microsimulation models using combinatorial optimization include the SMILE model from Ireland (Ballas, Clarke & Wiemers, 2005).

Another method being used for spatial microsimulation is a deterministic reweighting method using a generalised regression method to reweight the survey weights provided on many survey files to small-area benchmarks. This method has been pioneered by the National Centre for Social and Economic Modelling in Australia (Harding, Lloyd, Bill & King, 2003; Tanton, 2007).

The final method described is a deterministic combinatorial optimization method (Procter, Clarke, Ransley & Cade, 2008).

In terms of dynamic microsimulation models, the first dynamic spatial microsimulation model was created by staff at the Spatial Modelling Centre in Sweden in 1999 (Vencatasawmy, Holm, Rephann, Esko, Swan, Öhman, Åström, Alfredsson, Holme & Siikavaara, 1999). This model is called SVERIGE. In Britain, Birkin has also developed a dynamic spatial microsimulation model called Moses (Birkin, Wu & Rees, 2009). Also, SimBritain has been developed by Ballas et al. as a dynamic microsimulation model for Britain (Ballas, Clarke, Dorling & Rossiter, 2007).

DYNAMIC SPATIAL MICROSIMULATION

Dynamic Spatial Microsimulation is one of the most complex forms of spatial microsimulation, and the most data intensive. It requires raw data for each of the small areas as a starting point for the modelling. These raw data are then updated using probabilities derived from other sources. For the best results, these probabilities also need to be available for each small area, although probabilities for larger areas could be applied to the smaller areas if there is not much spatial variability in the raw data being updated. For instance, birth rates do not vary much for very small areas, so some aggregation could be used.

There are a number of examples of dynamic spatial microsimulation models, and the examples shown here are SVERIGE in Sweden, MOSES and SIMBritain in Britain, SMILE in Ireland and CORSIM in the US.

The SVERIGE model (Rephann, 2004; Vencatasawmy et al., 1999) uses longitudinal socio-economic information on every resident in Sweden from 1985 - 1995 with co-ordinates accurate to 100 m. This is a very powerful longitudinal dataset of all Swedes, and allows for very complex modelling.

The SVERIGE model has 10 modules, each with a set of rules that determine the occurrence of specific events in a person's life. Events are generated through deterministic models of behaviour and a monte carlo simulation. These behaviours are functions of individual, household and regional socio-economic characteristics.

For example, the mortality module uses two sets of mortality equations, one for those under 25; and one for those over 25. For those under 25, historical mortality rates by age and sex are used to decide whether a person dies. For those aged over 25, a regression model is used to calculate the probability that the person will die in that year. This regression model includes age, marital status, family earnings, education level, sex and whether working.

There are ten modules in SVERIGE:

- Fertility
- Education
- Employment and Earnings
- Cohabitation and Marriage
- Divorce/Dehabitation
- Leaving home
- Immigration
- Emigration
- Internal migration
- Mortality

Each of these modules is run using either a sample of the full population (to test a scenario); or the full population of Sweden (to minimise the risk of error). The sequence in which the modules are applied is the same as the list above.

The SVERIGE model is used for policy analysis in Sweden.

The next dynamic spatial microsimulation model is MOSES, by Birkin et al. in the UK (Birkin, Wu & Rees, 2009). This model starts with a synthetic database of everyone in the UK, so it uses a Population Reconstruction Model (Birkin, Turner & Wu, 2006) that provides considerable levels of spatial detail. There are a number of demographic processes modelled, including Birth, Death, Marriage, Household Formation, Health, Migration and Housing. All these are modelled using transition probabilities calculated from the Census, ONS Vital Statistics (for Births and Deaths) and the British Household Panel Survey. People are also aged forward for each year modelled. It has been used in the UK for demographic and health projections. The main difference between MOSES and SVERIGE is that MOSES uses a synthetic dataset of everyone in the UK, whereas SVERIGE uses a geocoded dataset of everyone in Sweden. This means that, in theory, SVERIGE will provide more accurate modelling.

Another dynamic spatial microsimulation model is called SimBritain (Ballas, Clarke, Dorling, Eyre, Thomas & Rossiter, 2005; Ballas, Clarke, Dorling, Rigby & Wheeler, 2006a; Ballas, Clarke, Dorling & Rossiter, 2007). This model is both a static and a dynamic spatial microsimulation model that uses the British Household Panel Survey as the base dataset and Small Area Statistics (SAS) tables from the British Census. This section will only describe the dynamic model.

The method used is a probabilistic synthetic reconstruction which uses Iterative Proportional Fitting (IPF) to generate a vector of individual characteristics on the basis of a joint probability distribution from the SAS tables. The method is a reweighting technique.

Once this base dataset for each area is created, the population for the future is calculated. This is where the dynamic element of this model is introduced. Mortality and fertility are based on location-specific probabilities. Fertility is a function of age, marital status and location. Monte-Carlo sampling against the fertility probabilities of each female is used to determine which females give birth, and if a birth occurs then a new individual with age 0, sex determined probabilistically, single, and social class and location that of the mother is created.

Migration can also be modelled using a probabilistic method, but this has not been implemented in SimBritain.

SimBritain is not as comprehensive a dynamic model as SVERIGE. Only fertility and mortality are modelled, and migration is not modelled yet. SimBritain also uses a synthetic dataset of the UK population, as does MOSES, whereas SVERIGE uses actual data on everyone in Sweden.

Another dynamic microsimulation model is a version of SMILE (Spatial Microsimulation model for Ireland), which is a model for Ireland (Ballas, Clarke & Commins, 2001; Ballas, Clarke & Wiemers, 2005; 2006b; Hynes, O'Donoghue, Morrissey & Clarke, 2009; Hynes, Morrissey & O'Donoghue, 2006). This model uses the Life-Cycle Analysis Model (O'Donoghue, Lennon & Hynes, 2009), which is a framework for dynamic microsimulation models that includes how data are stored in a relational database, the processes for ageing, birth, death, and migration.

Two processes are used in SMILE, a static spatial microsimulation model to create a base population for each area and a dynamic ageing process for this base population. The static model used for the dynamic version of SMILE is an IPF method, which is described below. The dynamic part of the model uses probabilities to model mortality, fertility and internal migration, similar to SVERIGE.

SMILE has been developed by the Rural Economy Research Centre of Teagasc, based in Galway. Teagasc have used SMILE to model methane emissions in Ireland, and the implications of the Common Agricultural Policy reforms on Ireland. The model is being further developed, and has a number of researchers at Teagasc using it.

The CORSIM model has been in development at Cornell University in the US since 1986 (Caldwell, Clarke & Keister, 1998; Caldwell & Keister, 1996). The CORSIM model incorporates 50 economic, demographic and social processes using about 900 stochastic equations and rule-based algorithms and 17 national microdata files, so it is a huge model. The model also projects forward to 2030. The model is modularised with a number of modules including a wealth module (Caldwell & Keister, 1996). While the CORSIM model has been used extensively in the past, little has been published from the model recently.

It can be seen from this review that the number of dynamic spatial microsimulation models is limited. This is partly because these models are data intensive. There is only one model in this review (SVERIGE) that uses actual data from the whole population to do the microsimulation modelling. All other models create synthetic small-area data for the microsimulation modelling. Not only is record-unit data required for each small area being estimated, but transition probabilities for each small area are required to update the populations. This can mean up to 900 equations (as used in the CORSIM model), and each equation will require updating at some time. Without this regular updating, and regular funding to update the models, they can get out of date very quickly, and they become unusable over time.

Because of these problems, a much easier type of model is the static spatial microsimulation model. Some of these still require synthetic data (although some methods do not), but they do not require a set of equations to update the small-area datasets so there is no need for regular updating of 900 equations that age the records in some way.

Static Spatial Microsimulation Models

Static microsimulation models are far less complex than dynamic models. A static microsimulation model will only calculate the day-after effect of a modelled change. This effect assumes no behavioural change, and while projections can be incorporated into a static microsimulation model, they are usually incorporated in a fairly basic way using population projections from another source.

A static spatial microsimulation model usually has a static microsimulation model for a larger area behind it, so any spatial effects modelled will be for the day after the policy change; and no behavioural responses will be modelled.

As shown in Figure 1, there are two broad methods for static spatial microsimulation: reweighting and synthetic reconstruction. The reweighting method uses a number of different techniques to adjust survey weights from a national survey so that the sample represents small areas rather than national totals. The synthetic reconstruction method creates a new dataset for each small area, based on tables from a Census for each small area, so microdata from a national survey are not required.

Synthetic Reconstruction

The idea of synthetic reconstruction is to create an imaginary list of individuals and households where the characteristics of the individuals, when aggregated, will match known aggregates in the area being estimated. The starting point may be to create a population that matches a known Age/Sex distribution; and then this population might be adjusted to reflect a known labour force distribution; and then occupation or industry could be added. These characteristics are matched sequentially, rather than all at once.

There are a few different methods used for synthetic reconstruction (Clarke & Spowage, 2005; Williamson, 1996), but the main method used is Iterative Proportional Fit (Birkin & Clarke, 1988; 1989). The IPF method can either build up a synthetic dataset for a small area using Census tables, and create an entirely synthetic dataset; or it can use a national sample from a survey to select records to fill a particular small area, subject to constraint tables from the Census. The first method was used in early work by Birkin and Clarke; and the second method was a later development.

The first method is described in detail by Birkin and Clarke, and was used in the SYNTHESIS model (Birkin & Clarke, 1988). To generate a vector of individual characteristics in an area x = (x_1,x_2,...,x_m), a joint probability distribution needs to be created, p(x). As information is rarely available for the full joint distribution, it needs to be built up one attribute at a time, so the probability of the different attributes is conditionally dependent on existing (known) attributes:

$$p(x) = p(x_1) * p\left(\frac{x_2}{x_1}\right) * p\left(\frac{x_3}{x_2}, x_1\right) * \ldots * p\left(\frac{x_m}{x_{m-1}}, \ldots, x_1\right)$$

The problem is how to use as much information as possible for the right hand side of this equation. This means estimating the joint probability distribution p(x_1, x_2, x_3) subject to known joint probabilities p(x_1,x_2) and p(x_1,x_3).

If we let the first approximation of p(x_1, x_2, x_3) = 1/$N_1N_2N_3$, where N_1, N_2 and N_3 are the number of states attributed to x_1, x_2 and x_3, then the vector x can be adjusted by the known states:

$$p^2(x_1, x_2, x_3) = p^1(x_1, x_2, x_3) \frac{p(x_1, x_2)}{\sum_{x_1} p^1(x_1 x_2 x_3)}$$

$$p^3(x_1, x_2, x_3) = p^2(x_1, x_2, x_3) \frac{p(x_1, x_3)}{\sum_{x_1} p^2(x_1 x_2 x_3)}$$

These equations are iterated until the probabilities reach a certain acceptable limit. Williamson et al. use this same estimation methodology in early work (Clarke, Kashti, Mcdonald & Williamson, 1997; Williamson, Clarke & McDonald, 1996), and there are many other users of the IPF method, or variations on this method (Ballas, 2004; Ballas & Clarke, 2001; Ballas, Clarke & Wiemers, 2005). Williamson et al. also used this method to estimate water demand (Williamson, Mitchell & McDonald, 2002)

Models that use the IPF method include SimLeeds 1 and 2 (Ballas & Clarke, 1999; Ballas, Clarke & Dewhurst, 2006; Ballas, Clarke & Turton, 1999) and other versions of SimLeeds (Ballas, 2000)

One of the advantages of the IPF method is that it does not require a microdata set, as it creates a synthetic midrodata set. Many statistical agencies, including the Office of National Statistics (ONS) in the UK and the Australian Bureau of Statistics (ABS) in Australia provide microdata files of their surveys (called Confidentialised Unit Record Files, or CURF's in Australia). However, in many cases users need to provide reasons for using the data, and need to fill in forms to access the data. Further, the ABS has restrictions on being able to match the data with other datasets. In these cases, a synthetic dataset may be the best option to provide a generalised model that users do not need to gain access to official datasets to use.

Reweighting

The next stage of development was to use microdata from a survey as the base for the spatial microsimulation model. There are two broad methods of reweighting. One is to select individuals from the microdata, until the sample selected for the small area looks like the small-area totals from some other source (usually a Census). These individuals can be randomly selected, but with some intelligence in the selection algorithm – this is a probabilistic method. The other way to choose individuals is through a formula, called a deterministic method. One method that chooses individuals from the microdata is called combinatorial optimisation.

The other reweighting strategy is a generalised regression reweighting method. This is a deterministic algorithm, which uses an iterative regression routine to derive weights that best represent the small-area totals. Many countries use this routine to reweight samples from their surveys to national totals (Cai, Creedy & Kalb, 2004; Singh & Mohl, 1996).

(a) Combinatorial Optimisation (Probabilistic method)

One method developed by Williamson uses a process called Combinatorial Optimisation (CO). This process is a mathematical process that finds an optimal object from a finite set of objects. Applied to spatial microsimulation, the process is used to choose which records from a survey best represent a small area (Williamson, 1996). This CO method is a synthetic reconstruction method that uses a microdata set as the base for the method.

This method is an iterative approach which selects a combination of households from the microdata to reproduce, as closely as possible, the population in the small area. The process starts with a random selection from the microdata, and then looks at the effect of replacing one household. If the replacement improves the fit to some small-area benchmarks, then it is chosen; if not, the original household is replaced and another household is chosen to replace it. This process is repeated, with the aim of gradually improving the fit.

Because the process is iterative, a decision needs to be made as to when to stop the process. This could be time elapsed, number of iterations reached, or accuracy level reached. The approach used by most users of this technique is a level of accuracy called the Total Absolute Error, which is the difference between the estimated totals and the benchmark totals, squared (so it is the absolute error) and summed for all benchmark tables.

An assessment of this technique by Voas and Williamson (Voas & Williamson, 2000) found that the results were reasonable for any variables that were part of the set of constraint tables.

However, estimating cross-tabulations for variables that were not in the list of constraints resulted in a poor fit.

The worst case scenario for this method is that every single combination of households is assessed to find the best fit. This maximises the time taken for the procedure to run. Further developments of the combinatorial optimisation techniques built some intelligence into the searching for records to select from the microdata, rather than randomly selecting records (Williamson, Birkin & Rees, 1998).

The first technique for intelligently searching for records tested was a hill-climbing approach. This approach selects a combination of records to be replaced, and then selects one record to replace a record in the combination. This reduces the number of combinations to be tested. While this technique is faster than randomly selecting from the whole microdata file, the procedure can still get stuck in sub-optimal solutions. Better solutions may exist, but because of a replacement made earlier, the optimal solution will not be found.

In testing this method, Williamson et al. observed that the hill climbing routine was getting sub-optimal solutions (Williamson, Birkin & Rees, 1998).

The next technique tested to make the CO algorithm more efficient was called simulated annealing. This allows the algorithm to climb down from sub-optimal solutions by allowing changes to the combination being optimised even if they make the solution worse (in terms of the TAE).

The choice of whether or not to accept a worse replacement is determined by an equation from thermo-dynamics:

$$p(\delta E) = \exp\left(-\frac{\delta E}{T}\right)$$

Where δE is the potential increase in energy and T is the absolute temperature.

In applying this to the combinatorial optimisation technique, T is set to the maximum change in performance likely by replacing an old element with a new one, and δE is the increase in the TAE. As replacement elements are selected, any that make the fit better are accepted, and any that make the fit worse are accepted if $p(\delta E)$ is greater than or equal to a randomly generated number. Note that an important element of this formula is that smaller values of δE (change in TAE as a result of the replacement) lead to a greater likelihood of a change being made.

The main problem with this method is that T, the initial temperature, has to be set. This temperature is also reduced over time, to simulate the cooling process in thermo-dynamics. Williamson et al. suggest reducing T by 5% if the number of successful replacements carried out is in excess of some set maximum. So there are three parameters to set – initial temperature, number of swaps before reducing the temperature, and the extent of reduction made each time. Williamson et al. test a number of different parameters in their paper (Williamson, Birkin & Rees, 1998).

Williamson et al. found that this simulated annealing method performs much better than the hill climbing algorithms, but also noted that to obtain the best solution, the amount of backtracking needs to be as small as possible. This means setting the parameters of the simulated annealing algorithm to minimise the back tracking. Williamson et al. stated that an initial temperature of 10 and a reduction in temperature of 5% after 100 successful swaps provided optimum results.

This combinatorial optimisation with simulated annealing technique has also been used by Ballas in a static version of the SMILE spatial microsimulation model for Ireland (Hynes, O'Donoghue, Morrissey & Clarke, 2009; Hynes, Morrissey & O'Donoghue, 2006) and in a model called Micro-MaPPAS, which is an extension of SimLeeds (Ballas, Kingston, Stillwell & Jin, 2007).

The static SMILE model uses combinatorial optimisation with simulated annealing on the National Farm Survey and Census for Ireland. The model is a farm level spatial microsimulation model that has been used to estimate farm income (Hynes, Morrissey & O'Donoghue, 2006) and methane emissions (Hynes, O'Donoghue, Morrissey & Clarke, 2009).

Micro-MaPPAS is a spatial microsimulation model used as a planning support system to analyse local policy and planning procedures. This means that the spatial microsimulation model has been integrated with GIS to enhance local policy decisions. This also involved creating a graphical user-interface to allow policy makers to interrogate the datasets and create their own thematic maps. The datasets for each small area are created using a combinatorial optimisation procedure with simulated annealing.

The method as implemented by Williamson has been fully described in the POP91 Instruction Manual (Williamson, 2007).

Another application of the CO method with simulated annealing is SimCrime. This model has been developed using the British Crime Survey and Census data in the UK. The model has been developed by Kongmuang et al. (Kongmuang, Clarke, Evans & Jin, 2006).

Another method for intelligent sampling suggested by Williamson et al. was a genetic algorithm. These algorithms were developed to imitate nature evolving towards 'optimal' solutions through natural selection. This process assesses each combination for 'fitness', and the fittest chromosomes are chosen to be 'parents', generating a new set of 'child' solutions. A random element is introduced through 'mutation', which is a low level chance of mistranslation between parents in reproduction. This mutation is important to introduce diversity into the population of possible solutions.

The problem with this method is that there are a number of parameters that need to be determined, and testing by Williamson et al. found that the genetic algorithm procedures worked worse that the hill climbing and simulated annealing process. Having said this, Birkin et al. has used a GA algorithm to calculate the base population for their model MOSES (Birkin, Turner & Wu, 2006), and found that the method performed poorly in terms of the accuracy of the results.

(b) Combinatorial Optimisation (deterministic method)

One of the problems with the CO probabilistic method is that for each run of the model, a different result will be given. This is because the records from the microdata are randomly selected, and assuming a purely random selection of records from the microdata (so, in computing terms, a different seed is used each time for the random-number generator), a different set of households will be selected each time the procedure is run.

Another way to choose records from the microdata is a deterministic method, which uses a formula to choose the records. This then produces the same results each time the model is run.

The process is described in Ballas et al. (Ballas, Clarke, Dorling, Eyre, Thomas & Rossiter, 2005). The starting point for the reweighting is the weights provided in the sample microdata, which take into account the sample selection probability, and are adjusted for partial and full non response and any known bias. These weights are first scaled down to the known totals for a particular cross tabulation in the small area:

$$n_i = w_i * s_{ij} / m_{ij}$$

Where w_i is the original weight for the household, n_i is the new weight, s_{ij} is data for the small area from a cross-tabulation, and m_{ij} is the data from the microdata for that same cross-tabulation (eg, Age by Sex).

This same process is used to fit the individuals to another table, and this process is then iterated.

One of the perceived problems with this method is that many of the households used to populate the small area can come from out of the area. Research by Tanton et al. show that in most cases this isn't a problem, but for smaller cities in Australia it can be (Tanton & Vidyattama, 2009). Ballas overcame this problem by using geographical multipliers, so households from within the area had a higher weight than households outside the area. These weight multipliers were then used to adjust the final weights (Ballas, Clarke, Dorling, Eyre, Thomas & Rossiter, 2005).

This procedure gives non-integer weights, so Ballas then uses another procedure to force these weights to be integer so each record on the microdataset represents a whole number of people.

Ballas showed that there was considerable bias in this method for variables not used as constraints in the simulation, but swapping suitable individuals between small areas further reduced the error.

This method has also been used by Procter et al. (Procter, Clarke, Ransley & Cade, 2008) for estimating obesity, and Anderson (Anderson, 2007a; 2007b) for estimating poverty in England and Wales.

(c) Generalised Regression

The first proponent of this method was Melhuish at the National Centre for Social and Economic Modelling (NATSEM) at the University of Canberra, and it was later significantly developed by Tanton, including using the model for small-area policy analysis; adding significant validation to the model; and adding projections to the model (Melhuish, Blake & Day, 2002; Chin, Harding, Lloyd, McNamara, Phillips & Vu, 2005; Chin & Harding, 2006; 2007; Chin, Harding & Bill, 2006; Phillips, Chin & Harding, 2006; Tanton, McNamara, Harding & Morrison, 2009; Cassells, Harding, Miranti, Tanton & Mcnamara, 2010; Tanton, Vidyattama, Nepal & McNamara, 2011). This method is the basis of this thesis.

The method uses the same generalised regression reweighting method used to reweight Australian surveys to National and State benchmarks. In summary, the method starts with the weights provided on the microdata. In Australia, these have been adjusted for the sample design (clustering, stratification, oversampling). This initial weight is divided by the population of the area to provide a reasonable starting weight required for the generalised regression procedure.

The generalised regression procedure uses a regression model to calculate a new set of weights, given the constraints provided for each small area. These weights are limited to being positive weights only, which means the procedure must iterate a number of times. The process takes an initial weight from the survey, and continually adjusts this weight until reasonable results are achieved, or until a maximum number of iterations has been reached. A full description of the method is in Tanton et al. (Tanton, Vidyattama, Nepal & McNamara, 2011)

The model has been used to derive estimates of poverty (Harding, Lloyd, Bill & King, 2003; 2006; Phillips, Chin & Harding, 2006; Tanton, Harding & McNamara, 2010; Tanton, McNamara, Harding & Morrison, 2009; Tanton, 2011), housing stress (McNamara, Tanton & Phillips, 2006; Phillips, Chin & Harding, 2006; Tanton, McNamara, Harding & Morrison, 2009) and Wealth (Vidyattama, Cassells, Harding & McNamara, 2011).

One of the advantages of this method is that projections are very easy to create, either by inflating the weights; or inflating the benchmarks and reweighting to new benchmarks (Harding, Vidyattama & Tanton, 2009), This method for projecting has also been used for SimBritain by Ballas (Ballas, Clarke, Dorling & Rossiter, 2007).

The other use that has been made of this model is for policy analysis for small areas (Harding, Vu, Tanton & Vidyattama, 2009; Tanton, Vidyattama, McNamara, Vu & Harding, 2009)

There are a number of areas where the SpatialMSM model may fall down, and these have all been tested in a paper by Tanton and Vidyattama (Tanton & Vidyattama, 2009). This paper tested three different aspects of the model, to test whether the model was resilient to outside factors. The three factors tested were increasing the number of benchmarks; using a restricted sample for estimating some areas in Australia; and using univariate constraint tables rather than multivariate constraints.

The authors found that the model stood up well to this testing. The authors added a number of benchmarks, and found that adding another two benchmarks decreased the level of accuracy slightly, and increased slightly the number of SLA's failing an accuracy criterion. The advantage of the additional benchmarks was that the final dataset was more general – so it could now be used for estimating education outcomes or occupation, as these were the two new dataset added.

Using univariate benchmarks gave more usable areas, but with a reduced level of accuracy for these areas.

Using records from the area being estimated (for example, not using Sydney records to populate Canberra SLA's) did not have a huge effect on many areas, but did affect some smaller capital cities in Australia, so more accurate estimates were derived for Adelaide and Perth.

This is the most comprehensive testing done for a spatial microsimulation model, and shows that the SpatialMSM model is a robust model.

(d) Other Methods

Birkin et al. used an Iterative Proportional Sampling method, which appears to be a sampling and then reweighting method for their Population Reconstruction Model in MOSES. The procedure pulls a random sample for the Sample of Anonymised Records; constructs cross tabulations from this synthetic population and compares this to the actual populations for the small area; and then adjusts the weights upwards for attributes under-represented in the area and downwards for attributes over-represented. This process is iterated until acceptable results were achieved (Birkin, Turner & Wu, 2006).

A very early implementation of spatial microsimulation used a spatial-interaction model and allocated individuals using allocation models solved at an aggregate level (Clarke & Wilson, 1985). These allocation models included housing and labour market models, fed from national economic forecasts.

(e) Comparison of methods

Synthetic reconstruction and combinatorial optimisation methodologies for the creation of small-area synthetic microdata have been examined by Huang and Williamson (2001). They found that outputs from both methods can produce synthetic microdata that fit constraint tables very well (Huang & Williamson, 2001). However, the dispersion of the synthetic data has shown that the variability of datasets generated by combinatorial optimisation is much less than by synthetic reconstruction, at ED and ward levels. The main problem for the synthetic reconstruction is that a Monte Carlo solution is subject to sampling error which is likely to be more significant where the sample sizes are small.

The ordering of the conditional probabilities in the synthetic reconstruction method can also be a problem as synthetic reconstruction is a sequential procedure. Another drawback of synthetic reconstruction is that it is more complex and time consuming to program. The outputs of separate combinatorial optimisation runs are much less variable and much more reliable. Moreover, combinatorial optimisation allows much greater flexibility in selecting small area constraints. Huang and Williamson conclude that combinatorial optimisation is much better than synthetic reconstruction when used to generate a single set of synthetic microdata.

In a recent conference paper, Tanton, Williamson and Harding compared the CO and Generalised Regression methods (Tanton, Williamson & Harding, 2007). The main problems with the SpatialMSM model at the time was the number of areas that a solution could not be found for, whereas the CO method was able to nearly always get an estimate for an area. The generalised regression algorithm at the time was also much slower than the CO algorithm, although again much has now been done to make the generalised regression algorithm more efficient.

The CO method gave slightly better results compared to the generalised regression algorithm, in terms of measures of accuracy.

VALIDATING SPATIAL MICROSIMULATION MODELS

There are a number of ways to assess the accuracy of spatial microsimulation models, or validate them, and many of these methods are also used to assess geographic models. Voas and Williamson (Voas & Williamson, 2001) outline a number of methods for assessing geographic data, that could also be used to assess the accuracy of spatial microsimulation models. These include statistics tested against the chi-squared distribution, the normal Z score and information theory, including a simple total absolute error.

Another question that arises in testing the output from spatial microsimulation is whether we should be testing proportions or absolute numbers. The normal Z score is a test of proportions, but can be adapted to be a test of absolutes, as described in Voas and Williamson (2001).

The authors conclude that the simple TAE is a crude but quick means of assessing the accuracy of small-area statistics.

Tanton et al. have also done considerable work on validating the SpatialMSM model, and report three broad methods (Tanton, Vidyattama, Nepal & McNamara, 2011). These are to aggregate the small-area data up to State level and test against reliable survey data for each Australian state; run the SpatialMSM model using benchmarks for each State rather than small areas and compare to reliable State level data from a survey; and calculate an estimate of poverty that can be calculated from Census data, calculate this same estimate from the SpatialMSM model and compare the results. This last method used a Standard Error around Identity to test whether the estimates from the Census were the same as the estimates from the model. This index was also used by Ballas (Ballas, Clarke, Dorling & Rossiter, 2007).

The Standard Error around Identity is calculated as:

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

Where

$$\begin{aligned} SEI &= S\tan dard\ Error\ about\ Identity \\ y_{est} &= estimates\ of\ poverty\ rates\ from\ spatial\ microsimulation\ (gross\ income) \\ y_{ABS} &= estimates\ of\ poverty\ rates\ from\ the\ ABS \\ \bar{y}_{ABS} &= mean\ estimates\ of\ poverty\ rates\ from\ the\ ABS \end{aligned}$$

Tanton et al. show that all of these validation methods work well, testing different aspects of the model.

Rahman et al. have also published on validation techniques, and have developed a validation method that compares the small-area estimates from the spatial microsimulation model to reliable estimates from a Census (Rahman, Harding, Tanton & Liu, 2010). The statistical measure of the accuracy of the results is called the Z Index, and is calculated as:

$$Z_i = \frac{\hat{P}_{ij}^m - P_{i0}^m}{\sqrt{\dfrac{P_{i0}^m (1 - P_{i0}^m)}{\sum_j n_{ij}}}}$$

Where:

$\hat{P}_{ij}^m$ is the estimated proportion in category m for the j^{th} sample in the in the i^{th} small area

P_{i0}^m is the true proportion in that category from the Census for the j^{th} sample in the i^{th} small area and

n_{ij} is the estimated sample size for the j^{th} sample in the i^{th} small area.

Rahman et al. find that this test provided accurate estimates for 98.4 per cent of areas with a p-value of 0.01.

There has also been some comparison of a number of different methods, to test the reliability for the application being considered. For example, SimLeeds uses 4 different methods: SimLeeds1 uses a conditional probabilities from Census data and a synthetic unit record data; SimLeeds2 uses data from the Sample of Anonymised Records (SAR) and conditional probabilities from Census data; SimLeeds 3 and 4 use IPF (Ballas, Clarke & Turton, 1999).

REASONS FOR SPATIAL MICROSIMULATION

There are a number of reasons for creating a spatial microsimulation model. The simplest reason is that small-area estimates of a particular variable only available on a survey are required. This could be used by Government for planning purposes (Harding, Vidyattama & Tanton, 2009; McNamara, Gong, Miranti, Vidyattama, Tanton, Harding & Kendig, 2009) and by researchers for looking at spatial disadvantage (Miranti, McNamara, Tanton & Harding, 2010; Tanton, Harding & McNamara, 2010).

The next reason is that spatial microsimulation models can be linked to static tax/transfer microsimulation models to look at small-area effects of a planned policy change. This can be used for Government to analyse which areas will be affected most by a policy change (Harding, Vu, Tanton & Vidyattama, 2009; Tanton, Vidyattama, McNamara, Vu & Harding, 2009; Vu & Tanton, 2010).

The final reason is that the user may be interested in projections of a particular piece of data into the future. Spatial microsimulation models allow for easy creation of these projections from within the modelling framework (Ballas, Rossiter, Thomas, Clarke & Dorling, 2005; Vidyattama & Tanton, 2010).

All of the methods and models outlined in this chapter can be used for any of these reasons, and the final outcome required will determine the type of model used. For example, if accurate projections are required then a dynamic model may be preferred over a static model, as while static models can make projections, a dynamic model will provide greater flexibility over how these projections are calculated as it will allow births and deaths to be modelled. If only basic projections are required, then a static model will be much easier to create, and will provide acceptable results for policy modelling and data estimation.

CHAPTER 2: CONTRIBUTION OF THE PUBLISHED PAPERS

The five peer-reviewed journal articles that constitute the body of this thesis have been published in highly ranked journals. The papers progress the area of spatial microsimulation using a particular spatial microsimulation model, SpatialMSM. This is a microsimulation model for Australia, but the five articles provide significant contributions in the area of spatial microsimulation in general.

The published papers show how the model has developed over time and are presented in a sequence that shows this development and the significant contribution that these papers make to this field. The papers outline the method used, summarise how the method differs from other methods, show a number of applications of the method that have been developed, how the model has been validated, and how the method has been extended to include projections.

While the original method was developed before I worked at NATSEM, my work has driven the development, application and implementation of the model over the last four years. The main academic achievement of this work was further development of the original model, comparing the model to other methods, and extending the model to incorporate new benchmarks and validation methods. I also applied the model in a number of situations, and led the push to link the STINMOD Tax/Transfer microsimulation model to the SpatialMSM spatial microsimulation model to allow small-area effects of policy changes. I also significantly recoded the model to make it easier to run. All these contributions are written up in these five articles.

There are three types of articles included in this body of work: a method paper (Chapter 3, 'Small area estimation using a reweighting algorithm'), a paper outlining extensions to the base model that I led (Chapter 4, 'Pushing it to the Edge: Extending Generalised Regression as a Spatial Microsimulation Method'), two application papers (Chapter 5, 'Small area poverty estimates for Australia's Eastern Seaboard in 2006' and Chapter 6, 'Spatial microsimulation as a method for estimating different poverty rates in Australia'), and a policy simulation (Chapter 7, 'Old, Single and Poor').

METHODOLOGICAL PAPERS (CHAPTERS 3 AND 4)

These two papers describe the method used by the SpatialMSM model, how the results are validated, and outline some of the limitations of the model.

Small area estimation using a reweighting algorithm (Chapter 3)

This paper has been published in the Journal of the Royal Statistics Society Series A: Statistics in Society, an A* ranked journal from the Royal Statistical Society (Tanton, Vidyattama, Nepal & McNamara, 2011). The major contribution of this paper was to document the method, using statistical terminology and formulae. In terms of academic relevance, this paper is important, as it documents for other researchers how the SpatialMSM model works. This paper is highly relevant for other researchers looking to use the generalized regression method, as it is the first article to outline in detail the method in one place.

This article also outlines the validation that now takes place in terms of the SpatialMSM model. This validation is done in a number of different ways, but the most useful validation exercise is calculating a Standard Error around Identity (SEI). This method has been implemented by other authors, but is now a standard way to validate data from the SpatialMSM model. This paper also contains technical discussion about the validation, and comparisons between results from the SpatialMSM model for each State to results from the Survey data.

Pushing it to the edge: Extending generalized regression as a spatial microsimulation method (Chapter 4)

This paper outlines the technical limits of the generalized regression spatial microsimulation method, and describes important methodological developments that impact on the results from the model. It has been accepted for publication in the International Journal of Microsimulation (C rank journal) (Tanton & Vidyattama, 2009). The methodological developments include which observations to apply from the surveys being used for the modeling; and the effect of changing the number of benchmarks in the model. This paper is relevant to researchers assessing how the SpatialMSM method could be used, and the limitations of the method in some circumstances. This paper documents the robustness of the SpatialMSM model, and shows to the research community that the model can still provide reliable estimates with a number of new benchmarks.

This work has also led to additional research on how estimates can be derived for areas that are very different from other areas in Australia. In particular, we are now experimenting with using only Tasmanian households to derive better estimates for small areas in Tasmania. It has also led to estimates for areas where previously no estimates could be derived, as it has led to a better understanding of the trade-off between the number of benchmarks and the accuracy of the estimates, so that in other work we now reduce the number of benchmarks for some areas where we have previously not been able to derive estimates using all the benchmarks. This provides an estimate for the area, albeit one that is not as accurate as if all the benchmarks had been used.

This article has made a significant contribution to the thinking behind the SpatialMSM model, and how it can be extended and developed further.

These two methodological papers (Chapters 3 and 4) provide a body of work on the generalized regression method as applied to spatial microsimulation that provides readers with the technical detail of the model, and the limitations of the model.

APPLYING THE SPATIALMSM MODEL (CHAPTERS 5, 6 AND 7)

These three chapters show applications of the SpatialMSM model, with two of these papers applying the model to estimates of poverty and one paper showing an application to policy modelling.

The two papers detailing the application of the model to poverty estimation are relevant for poverty researchers. These papers use slightly different methods to estimate poverty, showing the flexibility of the spatial microsimulation approach. Once weights have been derived, they can be applied in a number of different ways to derive a number of different estimates.

Small area poverty estimates for Australia's Eastern Seaboard in 2006 (Chapter 5)

This book chapter (Tanton, McNamara, Harding & Morrison, 2009) uses poverty rates calculated as the proportion of people below the poverty line, where the poverty line is half the median household equivalised disposable income.

These poverty rates are the ones usually derived for Australia (Harding, Lloyd & Greenwell, 2000; Saunders, Hill & Bradbury, 2008), and so are highly relevant for Australian researchers looking for small-area poverty rates. The results show much higher poverty rates in rural and remote areas compared to capital cities, and these results have had significant implications for decision making by people working in these communities and for policy makers developing policies for these communities. This paper therefore highlights the powerful nature of the spatial microsimulation tool in the management of social issues.

Spatial microsimulation as a method for estimating different poverty rates in Australia (Chapter 6)

This paper estimating different types of poverty rates has been published in Population, Space and Place, an A rank publication (Tanton, 2011). It is highly relevant for international researchers, as it uses spatial microsimulation to estimate the depth of poverty, which is generally accepted by international researchers as a better measure than head count poverty rates (which don't take into account how far a person is in poverty). This paper uses the same weights as used for the previous paper, showing the power of the reweighting spatial microsimulation methodology in that once the weights are calculated, they can be used to estimate other similar variables.

Together, these two papers show that the SpatialMSM model is relevant for policy makers in Australia and for researchers. While national poverty rates are easy to derive using national surveys from the Australian Bureau of Statistics, we know that this national average hides pockets of poverty in small areas. For a policy maker, there is not much point developing policies to combat poverty and then implementing these policies in an area that does not have high poverty. Knowing where poverty exists is extremely valuable information for a Government policy maker, and allows them to target policies to particular areas.

Old, Single and Poor: Using Microsimulation and Microdata to Analyse Poverty and the Impact of Policy Change among Older Australians (Chapter 7)

This paper (Tanton, Vidyattama, McNamara, Vu & Harding, 2009) takes a major step in the spatial microsimulation process by doing two things: it shows how the spatial microsimulation weights can be applied to a static national microsimulation model to derive small-area estimates of a policy change, and it shows how the method can be used to derive poverty rates for a sub-population (in this case, single aged-pensioners).

To be able to apply the weights to a static tax/transfer microsimulation model, the only requirement is that both models need to use the same surveys, so that the weights from the spatial microsimulation process can be matched to the records in the Tax/Transfer microsimulation model.

This paper is a significant paper for policy makers in two ways. First of all it shows how spatial microsimulation can be used with a Tax/Transfer microsimulation model to see what areas are affected most. This is a major contribution to spatial microsimulation knowledge, as the link between a Tax/Transfer microsimulation model and a spatial microsimulation model has not been done previously. It provides a powerful tool to not only estimate the small-area effects of a policy change, but also to develop projections of the small-area effects.

This tool is of considerable value for policy analysts because, as highlighted earlier, if poverty is spatial then it is important to know where it is highest when implementing a policy. However, even more important for policy makers is to know where a policy is going to have the greatest effect, as this will show where they need to implement the proposed policy being modelled.

The other thing that this paper shows is how the spatial microsimulation method can be used to estimate poverty for sub-groups of the population. This is a particularly useful contribution, as it means that policy makers can see how a new policy is affecting different types of families in different areas, again allowing policy makers to target their policies to certain groups of people.

CHAPTER 3: SMALL AREA ESTIMATION USING A REWEIGHTING ALGORITHM

Chapter 3

This chapter has been removed due to copyright restrictions.

This chapter is available as:

Tanton, R., Vidyattama, Y., Nepal, B. & McNamara, J. (2011) Small area estimation using a reweighting algorithm. *Journal of the Royal Statistical Society: Series A*, vol.174, no. 4, pp.931-951

Links to this chapter:

Print	http://webpac.canberra.edu.au/record=b1625778~S4
Online subscribed content (UC community)	http://ezproxy.canberra.edu.au/login?url=http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2011.00690.x/abstract
Online general public	http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2011.00690.x/abstract
DOI	10.1111/j.1467-985X.2011.00690.x

CHAPTER 4: PUSHING IT TO THE EDGE: EXTENDING GENERALIZED REGRESSION AS A SPATIAL MICROSIMULATION METHOD

Chapter 4

This chapter is available as:

Tanton, R., Vidyattama, Y. (2010) Pushing It To The Edge: Extending Generalised Regression As A Spatial Microsimulation Method. *International Journal of Microsimulation*, vol.3, no. 2, pp.23-33.

Links to this chapter:

Print	http://webpac.canberra.edu.au/record=b1625778~S4
Online subscribed content (UC community)	
Online general public	http://ima.natsem.canberra.edu.au/IJM/V3_2/Volume%203%20Issue%202/2_IJM_48%20Proof.pdf
DOI	

CHAPTER 5: SMALL AREA POVERTY ESTIMATES FOR AUSTRALIA'S EASTERN SEABOARD IN 2006

Chapter 5

This chapter has been removed due to copyright restrictions.

This chapter is available as:

Tanton, R., McNamara, J., Harding, A. & Morrison, T. (2009) Small area poverty estimates for Australia's eastern seaboard in 2006. In Zaidi, A. [et al] (eds.) *New Frontiers in Microsimulation Modelling*. Surrey, Ashgate Publishing.

Links to this chapter:

UC Library Catalogue - Thesis	http://webpac.canberra.edu.au/record=b1625778~S4
UC Library Catalogue - Chapter	http://webpac.canberra.edu.au/record=b1493693~S4
NLA Trove Catalogue - Chapter	http://trove.nla.gov.au/work/35821703

CHAPTER 6: SPATIAL MICROSIMULATION AS A METHOD FOR ESTIMATING DIFFERENT POVERTY RATES IN AUSTRALIA

Chapter 6

This chapter has been removed due to copyright restrictions.

This chapter is available as:

Tanton, R. (2011) Spatial Microsimulation as a Method for Estimating Different Poverty Rates in Australia. *Population, Space and Place*, vol. 17, no. 3, pp. 225-235.

First published online 2009.

Links to this chapter:

Print	http://webpac.canberra.edu.au/record=b1625778~S4
Online subscribed content (UC community)	http://ezproxy.canberra.edu.au/login?url=http://onlinelibrary.wiley.com/doi/10.1002/psp.601/abstract
Online general public	http://onlinelibrary.wiley.com/doi/10.1002/psp.601/abstract
DOI	10.1002/psp.601

Chapter 7: Old, Single and Poor: Using Microsimulation and Microdata to Analyse Poverty and the Impact of Policy Change among Older Australians

Chapter 7

This chapter is available as:

Tanton, R., Vidyattama, Y., McNamara, J., Vu, Q. N. & Harding, A. (2009) Old, Single and Poor: Using Microsimulation and Microdata to Analyse Poverty and the Impact of Policy Change among Older Australians. *Economic Papers*, vol. 28, no. 2, pp. 102-120.

Links to this chapter:

Print	http://webpac.canberra.edu.au/record=b1625778~S4
Online subscribed content (UC community)	http://ezproxy.canberra.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=48044821&site=ehost-live
Online general public	http://onlinelibrary.wiley.com/doi/10.1111/j.1759-3441.2009.00022.x/abstract
DOI	10.1111/j.1759-3441.2009.00022.x

CHAPTER 8: CONCLUSION

This thesis provides an extensive literature review on the spatial microsimulation method. The literature review identifies the latest methods for static spatial microsimulation are reweighting methods, with SpatialMSM being one of these spatial microsimulation methods. This method has mainly been used in Australia by the National Centre for Social and Economic Modelling (NATSEM) at the University of Canberra, while the UK has tended to follow a combinatorial optimisation reweighting methodology developed by Williamson and colleagues.

The five papers included in this thesis extend the SpatialMSM model in a number of ways. The first paper (Chapter 3) is a method paper, and is the first time that the statistical formulae used for the SpatialMSM model have been described. This paper also describes the validation used for the SpaptialMSM model in detail.

The next paper (Chapter 4) extends the methodology, testing the limitations of the method and providing some conclusions based on this testing. The conclusions reached from this paper are that the model is robust and can provide reasonable estimates.

The next two papers (Chapters 5 and 6) are applications to estimating poverty rates, and are significant in that they highlight the flexibility of the SpatialMSM model. The same method for calculating the weights has been used for calculating both headcount poverty rates and the depth of poverty.

The final paper (Chapter 7) shows how the model can be applied to estimate the effects of policy for small areas. The significance of this paper is that the weights from the SpatialMSM model have been applied to the output from a static non-spatial microsimulation model, STINMOD, showing again the flexibility of the reweighting approach.

Problems and future directions

Several problems had been encountered with this type of research. The main problem was that estimates could not be derived for many areas due to the procedure not calculating accurate estimates. This has now been overcome in my work by reducing the number of benchmarks for areas where accurate estimates cannot be calculated.

The other problem was that for the projections, the benchmark tables were all based on a regression of one set of benchmarks against another set. For some tables, this may not give reasonable projections, but overall the projections being benchmarked seem to be reasonable.

Future directions for this work include deriving better estimates of the future benchmark tables, and deriving other cross-tabulations once other benchmarks are added. Adding additional benchmarks has already been tested in one of the papers for this thesis (Tanton & Vidyattama, 2009). Other future directions include making the model dynamic or adding some behavioural elements to the model, similar to what has been done for SimBritain (Ballas, Clarke, Dorling & Rossiter., 2005).

Conclusions

Overall these five articles have made a significant contribution to the intellectual field of spatial microsimulation, and the body of work that these articles represent has pushed the boundaries of spatial microsimulation. The papers provide for effective improvement in predictions that will assist decision making for management of social issues.

BIBLIOGRAPHY

Anderson, B. (2007a). *Creating small-area Income Estimates: spatial microsimulation modelling*. UK Department of Communities, Retrieved from http://www.communities.gov.uk/documents/communities/pdf/325286.pdf.

Anderson, B (2007b) *Creating small area income deprivation estimates for Wales: spatial microsimulation modelling*. Chimera Working Paper 2007-11, Colchester: University of Essex. Retrieved from http://www.communities.gov.uk/documents/communities/pdf/325286.pdf.

Ballas, D. (2000). GIS and microsimulation for local labour market analysis. *Computers, Environment and Urban Systems*, *24*(4), pp. 305-330. doi: 10.1016/S0198-9715(99)00051-4.

Ballas, D. (2004). Simulating trends in poverty and income inequality on the basis of 1991 and 2001 census data: a tale of two cities. *Area*, 36(2), pp. 146-163. doi: 10.1111/j.0004-0894.2004.00211.x.

Ballas, D & Clarke, G. (1999). Regional versus local multipliers of economic change? A microsimulation approach. *39th Regional Science Association (ERSA) Congress, University College*. Dublin: University of Leeds.

Ballas, D & Clarke, G. (2001). Modelling the local impacts of national social policies : a spatial microsimulation approach. *Environment and Planning*, *19*(1), pp. 587-606. doi: 10.1068/c0003.

Ballas, D, Clarke, G. & Commins, P. (2001). Building a dynamic spatial microsimulation model for the Irish rural economy. *The 12th European colloquium on Theoretical and Quantitative Geography*, Rouen, France.

Ballas, D, Clarke, G. & Dewhurst, J. (2006). 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. doi: 10.1080/17421770600697729.

Ballas, D, Clarke, G., Dorling, D., Eyre, H., Thomas, B. & Rossiter, D. (2005). SimBritain: a spatial microsimulation approach to population dynamics. *Population, Space and Place*, *11*(1), pp. 13-34. Retrieved from http://dx.doi.org/10.1002/psp.351.

Ballas, D, Clarke, G., Dorling, D., Rigby, J. & Wheeler, B. (2006). Using geographical information systems and spatial microsimulation for the analysis of health inequalities. *Health informatics journal*, *12*(1), pp. 65-79. doi: 10.1177/1460458206061217.

Ballas, D, Clarke, G., Dorling, D. & Rossiter, D. (2007). Using SimBritain to Model the Geographical Impact of National Government Policies. *Geographical Analysis*, 39(1), pp. 44-77. Retrieved from http://dx.doi.org/10.1111/j.1538-4632.2006.00695.x .

Ballas, D, Clarke, G. & Turton, I. (1999). Exploring microsimulation methodologies for the estimation of household attributes. *4th International conference on GeoComputation, Mary Washington College, Virginia, USA, 25-28 July 1999.* Virginia: University of Leeds.

Ballas, D, Clarke, G. & Wiemers, E. (2005). Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place*, 11(3), pp. 157-172. Retrieved from http://dx.doi.org/10.1002/psp.359.

Ballas, D, Clarke, G. & Wiemers, E. (2006). Spatial microsimulation for rural policy analysis in Ireland: The implications of CAP reforms for the national spatial strategy. *Journal of Rural Studies*, 22(3), pp. 367-378. doi: 10.1016/j.jrurstud.2006.01.002.

Ballas, D, Kingston, R., Stillwell, J. & Jin, J. (2007). Building a spatial microsimulation-based planning support system for local policy making. *Environment and Planning A*, 39(10), pp. 2482-2499. doi: 10.1068/a38441.

Ballas, D, Rossiter, D., Thomas, B., Clarke, G. & Dorling, D. (2005). *Geography matters: Simulating the local impacts of National social policies*. York: Joseph Rowntree Foundation.

Battese, G. E., Harter, R. M. & Fuller, W. A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, 83(401), pp. 28-36.

Birkin, M & Clarke, M. (1988). SYNTHESIS -- a synthetic spatial information system for urban and regional analysis: methods and examples. *Environment and Planning A*, 20(12), pp. 1645-1671. doi: 10.1068/a201645.

Birkin, M & Clarke, M. (1989). The Generation of Individual and Household Incomes at the Small Area Level using Synthesis. *Regional Studies: The Journal of the Regional Studies Association*, 23(6), pp. 535 - 548. Retrieved from http://dx.doi.org/10.1080/00343408912331345702 .

Birkin, M, Turner, A. & Wu, B. (2006). A Synthetic Demographic Model of the UK Population : Methods , Progress and Problems. *Proceedings of the Second international conference on e-Social Science, NCESS, Manchester*.

Birkin, M, Wu, B. & Rees, P. H. (2009). Moses: Dynamic spatial microsimulation with Demographic Interactions. In A. Zaidi, A Harding & P Williamson (Eds.), *New frontiers in Microsimulation Modelling,* pp. 53 - 77. Ashgate.

Cai, L., Creedy, J. & Kalb, G. (2004). *Reweighting the Survey of Income and Housing Costs for Tax Microsimulation Modelling*. Retrieved from http://www.melbourneinstitute.com/downloads/reports/46_reweighting_fin_report.pdf.

Caldwell, S B, Clarke, G P & Keister, L A. (1998). Modelling regional changes in US household income and wealth: a research agenda. *Environment and Planning C: Government and Policy*, 16(6), pp. 707-722. doi: 10.1068/c160707.

Caldwell, S.B. & Keister, L.A. (1996). Wealth in America: family stock ownership and accumulation, 1960–1995. In G. Clarke (Ed.), *Microsimulation for Urban and Regional Policy Analysis,* pp. 88 - 116. London: Pion.

Cassells, R., Harding, A, Miranti, R, Tanton, R & McNamara, J. (2010). Spatial Microsimulation : Preparation of Sample Survey and Census Data for SpatialMSM / 08 and SpatialMSM/09. *NATSEM Technical Report 36*, Canberra: NATSEM

Chin, S. F. & Harding, A. (2006). Regional Dimensions: Creating synthetic small-area microdata and spatial microsimulation models. *NATSEM Technical Paper 33*, Canberra: NATSEM.

Chin, S. F. & Harding, A. (2007). SpatialMSM - NATSEM's small area household model for Australia. In A. Gupta & A. Harding (Eds.), *Modelling our future: Population ageing health and aged care,* pp. 563-566. Oxford: Elsevier.

Chin, S. F., Harding, A & Bill, A. (2006). Regional dimensions: Preparation of 1998-99 HES for reweighting to small area benchmarks. *NATSEM Technical Paper 34*, Canberra: NATSEM.

Chin, S. F., Harding, A, Lloyd, R., McNamara, J, Phillips, B & Vu, Q. (2005). Spatial microsimulation using synthetic small-area estimates of income, tax and social sequrity benefits. *Australian Journal of Regional Studies*, 11(3), pp. 303-335. Retrieved from http://www.anzrsai.org/system/files/f8/f4/f20/f21/o74//6Chin and Others final.pdf.

Clarke, G. P., Kashti, A., McDonald, A. & Williamson, P. (1997). Estimating Small Area Demand for Water: A New Methodology. *Water and Environment Journal*, 11(3), pp. 186-192. doi: 10.1111/j.1747-6593.1997.tb00114.x.

Clarke, M. & Holm, E. (1987). Microsimulation Methods in Spatial Analysis and Planning. *Geografiska Annaler Series B*, 69(2), pp. 145-164.

Clarke, M. & Spowage, M. (2005). Integrated models for public policy analysis: An example of the practical use of simulation models in health care planning. *Papers in Regional Science*, 55(1), pp. 25-45. doi: 10.1111/j.1435-5597.1984.tb00825.x.

Clarke, M. & Wilson, A. (1985). The Dynamics of Urban Spatial Structure: The Progress of a Research Programme. *Transactions of the Institute of British Geographers*, 10(4), pp. 427 - 451. Retrieved from http://www.jstor.org/pss/621890.

Clarke, M., Forte, P., Spowage, M. & Wilson, A. (1984). A strategic planning simulation model of a district health service system: the in-patient component and results. In W. van Elmeren, R. Engelbrecht & C. D. Flagle (Eds.), *Systems Science in Health Care*. Berlin: Springer Verlag.

Elbers, C., Lanjouw, J. & Lanjouw, P. (2002). *Micro-level estimation of welfare*, World Bank Policy Research Working Paper, WPS2911.

Hagerstrand, T. (1952). *The propagation of innovation waves.* Retrieved from http://catalogue.nla.gov.au/Record/4364370.

Harding, A, Lloyd, R., Bill, A. & King, A. (2003). Assessing poverty and inequality at a detailed regional level: New advances in spatial microsimulation. *Inequality, Poverty and Human well-being, Helsinki, Finland, 30 - 31 May 2003*. Helsinki, Finland.

Harding, A, Lloyd, R., Bill, A. & King, A. (2006). Assessing poverty and inequality at a detailed regional level: New advances in spatial microsimulation. In M. McGillivray & M. Clarke (Eds.), *Understanding Human Well-being,* pp. 239-261. Helsinki: United Nations University Press.

Harding, A, Lloyd, R. & Greenwell, H. (2000). *Financial Disadvantage in Australia 1990 to 2000: The persistence of poverty in a decade of growth.* NATSEM/Smith Family Report, Canberra, Australia.

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..

Harding, A, Vu, Q., Tanton, R & Vidyattama, Y. (2009). Improving Work Incentives and Incomes for Parents: The National and Geographic Impact of Liberalising the Family Tax Benefit Income Test. *Economic Record*, 85(s1), S48-S58. Retrieved from http://dx.doi.org/10.1111/j.1475-4932.2009.00588.x .

Huang, Z. & Williamson, P. (2001). A Comparison of Synthetic Reconstruction and Combinatorial Optimisation Approaches to the Creation of Small-Area Microdata. *Working paper 2001/02, Department of Geography, University of Liverpool.*

Hynes, S, O'Donoghue, C, Morrissey, K & Clarke, G. (2009). A spatial micro-simulation analysis of methane emissions from Irish agriculture. *Ecological Complexity*, *6*(2), pp. 135-146. doi: 10.1016/j.ecocom.2008.10.014.

Hynes, S, Morrissey, K & O'Donoghue, C. (2006). Building a Static Farm Level Spatial Microsimulation Model: Statistically Matching the Irish National Farm Survey to the Irish Census of Agriculture. *46th Congress of the European Regional Science Association*. Volos, Greece.

Kongmuang, C., Clarke, G., Evans, A., & Jin, J. (2006). SimCrime : A Spatial Microsimulation Model for the Analysing of Crime in Leeds. *School of Geography Working Paper 06/1, University of Leeds*. Retrieved from http://eprints.whiterose.ac.uk/4982/1/SimCrime_WorkingPaper_version1.1.pdf.

McNamara, J, Gong, C., Miranti, R, Vidyattama, Y, Tanton, R, Harding, A & Kendig, H. (2009). Two Worlds of Ageing: Spatial Microsimulation Estimates of Small Area Advantage and Disadvantage Among Older Australians. *2nd General Conference of the International Microsimulation Association*. Ottawa, Canada.

McNamara, J, Tanton, R & Phillips, B. (2006). *The regional impact of housing costs and assistance on financial disadvantage*. AHURI Final Report No. 109.

Melhuish, A., Blake, M. & Day, S. (2002). An Evaluation of Synthetic Household Populations for Census Collection Districts Created Using Spatial Microsimulation Techniques, *Paper prepared for the 26th Australia & New Zealand Regional Science Association International (ANZRSAI) Annual Conference, Gold Coast, Queensland, Australia, 29 September – 2 October 2002.*

Miranti, R, McNamara, J, Tanton, R & Harding, A. (2010). Poverty at the Local Level: National and Small Area Poverty Estimates by Family Type for Australia in 2006. *Applied Spatial Analysis and Policy*. doi: 10.1007/s12061-010-9049-1.

Orcutt, G. H. (1957). A New Type of Socio-Economic System. *The Review of Economics and Statistics*, 39(2), pp. 116 - 123. doi: 10.2307/1928528.

O'Donoghue, C., Lennon, J. & Hynes, S. (2009). The Life-Cycle Income Analysis Model (LIAM): A Study of a Flexible Dynamic Microsimulation Modelling Computing Framework. *International Journal of Microsimulation*, *2*(1), pp. 16-31. Retrieved from http://www.microsimulation.org/IJM/V2_1/IJM_2_1_2.pdf.

Pfeffermann, D. (2002). Small Area Estimation-New Developments and Directions. *International Statistical Review*, 70(1), pp. 125-143. doi: 10.1111/j.1751-5823.2002.tb00352.x.

Phillips, B, Chin, S & Harding, A. (2006). Housing Stress Today : Estimates for Statistical Local Areas in 2005. *Australian Consortium for Social and Political Research Incorporated Conference, Sydney, 10-13 December 2006*

Procter, K, Clarke, G, Ransley, J & Cade, J. (2008). Micro-level analysis of childhood obesity, diet, physical activity, residential socioeconomic and social capital variables: where are the obesogenic environments in Leeds? *Area*, 40(3), pp. 323-340. doi: 10.1111/j.1475-4762.2008.00822.x.

Rahman, A., Harding, A, Tanton, R & Liu, S. (2010). Simulating the characteristics of populations at the small area level : New validation techniques for a spatial microsimulation model in Australia. *Proceedings of the Social Statistics Section of* *the American Statistical Association Conference, Vancouver, Canada* (pp. 2022-2036).

Rephann, T. (2004). Economic-Demographic Effects of Immigration: Results from a Dynamic Spatial Microsimulation Model. *International Regional Science Review*, 27(4), pp. 379-410. doi: 10.1177/0160017604267628.

Saunders, P., Hill, T. & Bradbury, B. (2008). Poverty in Australia: Sensitivity analysis and recent trends. *Social Policy Research Centre Report 4/08, Sydney.* Retrieved from http://www.sprc.unsw.edu.au/media/File/Report4_08_Poverty_in_Australia.pdf.

Singh, A & Mohl, C. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, 22, pp. 107 - 115.

Tanton, R., Williamson, P. & Harding, A. (2007). Comparing two methods of reweighting a survey file to small area data: generalised regression and combinatorial optimisation. *First General Conference of the International Microsimulation Association, Vienna, 20 - 22 August 2007.*

Tanton, R. (2011). Spatial microsimulation as a method for estimating different poverty rates in Australia. *Population, Space and Place*, 17(3), pp. 222 - 235. doi: 10.1002/psp.601.

Tanton, R, Harding, A & McNamara, J. (2010). Urban and Rural Estimates of Poverty: Recent Advances in Spatial Microsimulation in Australia. *Geographical Research,* 48(1), 52-64. doi: 10.1111/j.1745-5871.2009.00615.x.

Tanton, R, McNamara, J, Harding, A & Morrison, T. (2009). Rich suburbs, Poor suburbs? Small area poverty estimates for Australia's Eastern Seaboard in 2006. In A. Zaidi, A Harding & P Williamson (Eds.), *New Frontiers in Microsimulation Modelling*. Vienna, Austria: Ashgate.

Tanton, R & Vidyattama, Y. (2009). Pushing them to the edge: An assessment of spatialmicrosimulation methods. *2nd General Conference of the International Microsimulation Association*. Ottawa, Canada.

Tanton, R, Vidyattama, Y, McNamara, J, Vu, Q. & Harding, A. (2009). Old, Single and Poor: Using Microsimulation and Microdata to Analyse Poverty and the Impact of Policy Change among Older Australians. *Economic Papers: A journal of applied economics and policy*, 28(2), pp. 102-120. doi: 10.1111/j.1759-3441.2009.00022.x.

Tanton, R, Vidyattama, Y, Nepal, B & McNamara, J. (2011). Small area estimation using a reweighting algorithm. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. doi: 10.1111/j.1467-985X.2011.00690.x.

Tanton, R, Williamson, P & Harding, A. (2007). Comparing two methods of reweighting a survey file to small area data: generalised regression and combinatorial optimisation. *First General Conference of the International Microsimulation Association, Vienna, 20 - 22 August 2007.*

Van Imhoff, E. & Post, W. (1998). Microsimulation methods for population projection. *Population. English selection*, 10(1), pp. 97-138. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12157954.

Vencatasawmy, Holm, Rephann, Esko, Swan, Öhman, Åström, Alfredsson, Holme & Siikavaara (1999). Building a spatial microsimulation model. *Paper presented at the 11th European Colloquium on Quantitative and Theoretical Geography in Durham, England, on September 3-7, 1999.*

Vidyattama, Y, Cassells, R., Harding, A & McNamara, J. (2011). Rich or poor in retirement? A small area analysis of Australian superannuation savings in 2006 using spatial microsimulation. *Regional Studies*, forthcoming

Vidyattama, Y & Tanton, R. (2010). Projecting small area statistics with Australian spatial microsimulation model (SpatialMSM). *Australian Journal of Regional Studies*, 16(1), pp. 99-126.

Voas, D & 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.

Voas, D & Williamson, P. (2001). Evaluating Goodness-of-Fit Measures for Synthetic Microdata. *Geographical and Environmental Modelling*, 5(2), pp. 177-200. doi: 10.1080/13615930120086078.

Vu, Q. N. & Tanton, R. (2010). The distributional impact of the Australian Government's Household Stimulus Package. *Australian Journal of Regional Studies*, 16(1), pp. 127-145.

Williamson, P, Mitchell, G. & McDonald, A. (2002). Domestic Water Demand Forecasting: A Static Microsimulation Approach. *Water and Environment Journal*, 16(4), pp. 243-248. Retrieved from http://dx.doi.org/10.1111/j.1747-6593.2002.tb00410.x .

Williamson, P. (1996). Community Care policies for the Elderly, 1981 and 1991: a microsimulation approach. In G. P. Clarke (Ed.), *Microsimulation for Urban and Regional Policy Analysis,* pp. 64 - 87. Pion Ltd.

Williamson, P. (2007). *CO Instruction Manual.* Population Microdata Unit, Dept. of Geography, University of Liverpool. Retrieved from < http://pcwww.liv.ac.uk/~william/microdata/workingpapers/CO Instruction Manual 070615.pdf >.

Williamson, P, Birkin, M & Rees, P. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*, 30(5), pp. 785-816.

Williamson, P, Clarke, G & McDonald, A. (1996). Estimating small area demands for water with the use of microsimulation. In Graham Clarke (Ed.), *Microsimulation for Urban and Regional Policy Analysis,* pp. 117 - 148. Pion Ltd.