Comparing Two Methods of Reweighting a Survey File to Small Area Data

Robert Tanton

National Centre for Social and Economic Modelling (NATSEM)
Institute for Governance and Policy Analysis, University of Canberra
ACT 2601
AUSTRALIA
robert.tanton@canberra.edu.au

Paul Williamson

Department of Geography and Planning
University of Liverpool
L69 7ZT United Kingdom
P.Williamson@liv.ac.uk

Ann Harding

National Centre for Social and Economic Modelling (NATSEM)
Institute for Governance and Policy Analysis, University of Canberra
ACT 2601
AUSTRALIA
ann.harding@canberra.edu.au

ABSTRACT: One method of calculating small area estimates using survey data involves deriving new weights for each respondent in the survey. These new weights are derived so that the survey data sums to some known totals for a small area (from either a Census or administrative data). There are different methods for calculating these weights, and this paper analyses the results from two different methods - a generalised regression method and combinatorial optimisation. The weights derived from each method are compared, and advantages and disadvantages of each method are assessed. Estimates of housing stress at a Statistical Local Area in Australia from each method are then calculated, and these estimates are then validated against a third reliable source, Australian Census data from 2001.

1.   INTRODUCTION

Spatial microsimulation is a microsimulation method that can be used for small area estimation (Anderson, 2007; Clarke *et al.*, 1997), projections (Williamson *et al.*, 2002; Harding *et al.*, 2011) and analysing small area policy change (Hynes *et al.*, 2009; Tanton *et al.*, 2009). Recent reviews of the field include Hermes and Poulsen (2012) and Tanton and Edwards (2013a). There are a number of methods that can be used for spatial microsimulation, and they have developed over a number of years. A brief review and history of the methods is described in a paper in this special issue (Tanton, 2014). There are also many applications of spatial microsimulation, and these are outlined in another paper in this special issue (O'Donoghue *et al.*, 2014).

As far as the authors are aware, there have been two recent papers that have compared different spatial microsimulation methods (Williamson, 2013; Harland *et al.*, 2012). This paper adds to this literature by comparing the generalised regression technique and the combinatorial optimisation technique. Both methods are described in Tanton and Edwards (2013b). The first uses a Generalised Regression (GREG) approach. GREG is a well-established approach to reweighting survey data (Lehtonen and Veijanen, 2009). The particular algorithm used in this paper iteratively attempts to minimise a truncated exponential distance function using a SAS program called GREGWT, developed by the Australian Bureau of Statistics from INSEE's reweighting macro CALMAR (Bell, 2000). GREGWT is used by the ABS to benchmark their survey data to known State totals. The NATSEM adaptation of this approach to small-area estimation is documented in Tanton *et al.* (2011; 2013) and elaborated below in Sections 2 and 3 of this paper. The code for GREGWT is available on request from the ABS.

The second method of survey reweighting considered in this paper uses an iterative 'combinatorial optimisation' algorithm. An initial set (combination) of households are drawn from a survey at random (with replacement), following which a succession of random changes in the households selected are made, with a view to optimising the fit of the household combination to the specified small area benchmarks. To reduce the risk of getting trapped in a sub-optimal solution a simulated annealing approach to the acceptance of household changes is adopted. This algorithm is implemented using the program CO, developed by Dr Paul Williamson at the University of Liverpool (Williamson, 2007, 2013; Williamson, *et al.*, 1998) and now available as part of an R library (Kavroudakis, 2013).

Unless earlier exit criteria are met (convergence measures for GREGWT; estimate fit thresholds for CO), both algorithms continue until a maximum number of user-specified iterations has been

exceeded.

This paper compares the two algorithms in terms of their advantages and disadvantages for small-area estimation, including the number of areas which fail to satisfy minimum fit criteria, and the resulting weights from each method.

Section two of this paper outlines the data and benchmarks used for both methods. The data and benchmarks used for each method are exactly the same. Section three summarises the differences between the two methods, in terms of the methods and assumptions. Section four compares the results from each method, looking at the total difference and differences for each benchmark. Section 5 provides further analysis of the weights - including looking at the distribution of the weights from each procedure - while Section 6 provides conclusions and directions for further work.

2. BACKGROUND

There has been considerable work in Australia and Britain on generating small area estimates using survey data. The attraction of small area models is that they allow a survey designed for generating reliable estimates for a large area to be used to derive reliable estimates for a small area, without increasing the sample size, which is an expensive process.

In 2006, the Australian Bureau of Statistics (ABS) produced a small area estimation practice manual (ABS, 2006), which outlines some of the available techniques, and includes a section on diagnostics and quality measurement. The manual covers simple small area methods, like broad Area Ratio Estimator and Calibration estimators; and then covers regression methods, including random effects regression models. While this manual is theoretical, the ABS has also produced a number of small area estimates using a variety of techniques. Similar reviews have been compiled by, amongst others, Pfeffermann (2002), a consortium of European Statistical Agencies (EURAREA, 2005) and Marshall (2010).

In 2005, the ABS produced small area estimates of disability (ABS, 2005), which looked at three different methods of estimating disability for small areas; a Poisson regression model; a Bernoulli model; and ratio estimation. The report found that the Bernoulli model and the ratio estimator gave the best results, with the Poisson model performing poorly, possibly due to overdispersion.

The ABS also used a number of methods to generate small area estimates of crime. While some of this was unpublished, a method using a regression estimator was published (Tanton, *et al.*,

2001). The results for small area estimates of crime from a number of other methods, including a ratio method based on an article by Purcell and Lincare (Purcell and Linacre, 1976) and a Structure Preserving Regression Estimator (SPREE) used for estimating labour force by the ABS, were unpublished, due to the fact that the results were difficult to validate.

This has been the biggest difficulty with the modelled small area estimates derived by the ABS - there is no estimate of the reliability of the results, for example, standard errors or confidence intervals. Recent work by the ABS has focussed on modelling labour force status, and assessing the quality of the estimates using relative root mean squared errors (RRMSEs) (ABS, 2011).

Outside of the ABS, the Australian Commonwealth Department of Employment and Workplace Relations uses a SPREE approach to estimate small area labour force statistics (Commonwealth Department of Employment and Workplace Relations, 2007). The ABS is now re-examining the estimation of the labour force using small area estimation techniques (ABS, 2012).

The National Centre for Social and Economic Modelling has also produced small area estimates of poverty and housing stress, using the methods outlined in this paper (see Harding, *et al.*, 2006; Tanton *et al.*, 2009; Tanton, 2011; Phillips, *et al.*, 2006).

In the UK the Office for National Statistics has conducted a review of alternative methods for updating small area population estimates between censuses, in lieu of a population register, concluding by favouring a ratio change approach (Bates, 2006). An allied initiative has seen the development of small area estimates using a multilevel regression-based synthetic estimator fitted using area-level covariates. The result has been the release of a series of 'experimental' small-area statistics covering topics such as household overcrowding and social capital (Heady, *et al.*, 2003), leading to the development of a regular series of official small area income estimates (Bond and Campos, 2010).

2.1.   Data and benchmarks

Both the methods of spatial microsimulation being compared here require two sets of data. One set is the survey which is being reweighted; and the second is the set of benchmarks that the survey is being reweighted to. The benchmarks must be reliable for the small area being estimated.

In this case, we have used data from the 1998-99 Household Expenditure Survey from the Australian Bureau of Statistics (ABS). This survey is a survey conducted by the ABS collecting

data on income and expenditure from a sample of households in Australia. The file used for this work is a Confidentialised Unit Record File (CURF), which is a file of every person in the sample with all the information collected from them, but with no identifying information. We then manipulate this file by adding a record for each child (the CURF only has the total number of children in the household) and for each person living in a non-private dwelling (which are not on the CURF but are on the data being benchmarked to). A description of the process to prepare a survey CURF for spatial microsimulation is described in Cassells *et al.* (2013).

The benchmarks for the reweighting process come from the 2001 Australian Census. The Census provides reliable data for Statistical Local Areas (SLAs), which range in population from 12 to 181,327 people, and the Expanded Community Profile tables provide the cross-tabulations that we require. Unfortunately the raw Census data include 'Not Stated' counts. These are counts of people or households that did not respond to certain questions on the Census. This can be partial non-response (eg, they said they were employed, but did not say whether this was full time or part time); or full non-response (e.g., they didn't answer the employment question). The 'Not Stated' values are distributed across all valid responses, using an integer pro rata method in which any unit remainder is allocated to the category with the highest value. Further information on how the survey and Census data are adjusted can be found in NATSEM's technical papers, available from the NATSEM website (Cassells *et al.*, 2010; Chin, *et al.*, 2006).

From the survey and Census data, we have chosen a set of benchmark variables that are available on both datasets, aggregating variable categories in one or other set of data until the categories are exactly the same. The final set of benchmarks used for this project are shown in Table 1.

Table 1   Benchmarks used for creating small-area weights

Census table	Type	Dimension[1]	Fully specified[2]	Benchmarks (no.)
Age by sex by labour force status	Person	Multi	Yes	32
Residents in different types of non-private dwelling	Person	Single	Yes	8
Household Type	Household	Single	No	1
Household size - number usual residents	Household	Single	Yes	7
Dwelling tenure by weekly household rent	Household	Multi	No	7
Dwelling tenure by household type	Household	Multi	Yes	15
Dwelling tenure by weekly household income	Household	Multi	No	16
Monthly household mortgage by weekly household income	Household	Multi	Yes	12
Weekly household rental by weekly household income	Household	Multi	Yes	20
Dwelling structure by household family composition	Household	Multi	No	12
Total number of benchmark tabulations				10
Total number of benchmarks				130

1. Multi-dimensional means cross-tabulations of variables
2. Not fully specified means that one or more of the cells in a benchmark tabulation were not used for weight production. For example, for the benchmark table of 'Household Type', the count of 'Private households' was extracted for use as a benchmark, whilst the count of 'Non-private dwellings' was excluded from the reweighting process

3.   DIFFERENCES BETWEEN THE METHODS

There are a number of theoretical differences between GREGWT and CO that need to be outlined.

3.1.   Algorithms

The algorithms used for each method are quite different. The GREGWT algorithm is essentially a constrained distance minimisation function. The method uses a truncated linear regression model to get an initial weight (constrained so the weights cannot be below 0 or above a preset limit set by the user). Because these weights are truncated, the boundary conditions may not be met by the initial weights, so an iterative approach is used to match the boundary conditions after a number of iterations. The iterations continue until convergence is reached (so the difference between the estimated benchmark and the actual benchmark for the area from the Census data is within a set limit), or a set number of iterations is made, at which time the iteration stops. The process needs a start weight, and this is set to the original ABS survey weight for the survey record divided by the population of the SLA. In many cases, there is no iteration as the initial regression estimate provides weights that are within the tolerance set.

The method implemented in GREGWT is from Singh and Mohl (1996) and its application in spatial microsimulation is described in Tanton *et al.* (2011). Full information on the GREGWT macro can be found in the user manual for GREGWT (Bell, 2000).

At the end of the reweighting process, every household in the survey dataset will have a weight for each Census SLA for which benchmark counts were supplied. In a small number of cases these weights will have been generated even though convergence was not achieved, due to the algorithm halting after exceeding a user-specified number of iterations (30 for the work reported in this paper). In these non-convergent cases the weights from the terminal iteration typically, but not always, include a small number of exceptionally high household weights, leading to very poor fits to one or more benchmark counts. We have chosen not to discard all non-convergent GREGWT outputs, as a few technically non-convergent estimates actually give rise to sets of weights that fit all benchmarks reasonably well. Instead, for the purposes of this paper, we identify as 'non-convergent' any SLA for which the sum of the absolute value of all errors across all benchmarks is greater than the number of households in the SLA; so where:

$$\frac{\sum_{Benchmark} ABS(Actual - Estimated)}{N_{hh}} > 1$$

For an account of the rates of non-convergence see Section 5.

The Combinatorial Optimisation algorithm, as currently implemented, may be viewed as an integer reweighting algorithm. For each household in an SLA (as recorded in the Census benchmark counts), CO randomly selects one household (with replacement) from the survey dataset. This is equivalent to setting all survey household weights to 0, then incrementing household weights at random by a count of 1 until the sum of weighted households matches the equivalent benchmark count. In each subsequent iteration the weight of one survey household is randomly increased by one, whilst the weight of another (non-zero weighted) household is randomly decreased by one. This is equivalent to randomly swapping households in and out of the set (combination) of households currently selected to represent the SLA. If the change in weights leads to improved fit, the change is retained; a few adverse changes in weights are also accepted, with a probability that diminishes in proportion to (i) size of the adverse impact and (ii) number of iterations, in order to avoid getting stuck in a local sub-optima; otherwise the change is rejected and the weights are reverted to their previous values. CO will continue to iterate until either a minimum fit threshold is achieved, or until a maximum number of user-specified iterations has been exceeded (5 million for the results reported in this paper). Full details of the algorithm and its links to simulated annealing are published in an article by Williamson *et al.* (Williamson, *et al.*, 1998). The user manual for the publicly available fortran-based version of the CO program has been published by Paul Williamson (Williamson, 2007), whilst more recently a version of CO has been made available via the R package 'sms' (Kavroudakis, 2013).

When deciding whether or not to accept a change in household weights, CO can evaluate one of two measures of fit. The first is the Overall Total Absolute Error. This is a conventional measure of fit that seeks to minimise the absolute difference between benchmark counts and their weighted survey equivalents, and is given by

$$OTAE = \sum_{ij} \left| o_{ij} - e_{ij} \right|$$

where e_{ij} is the expected (census) count for cell j in benchmark table i and o_{ij} is its estimated (weighted survey) equivalent

The second measure is the Overall Relative Sum of Squared modified Z-Scores (ORSSZm2), defined as

$$ORSSZm^2 = \sum_i \frac{1}{c_i} \left[\sum_j \frac{\left(o_{ij} - e_{ij} \right)^2}{e_{ij} \left(1 - \dfrac{e_{ij}}{\sum_j e_{ij}} \right)} \right]$$

where c_i = the χ^2 critical value for benchmark table i, with p=0.05 and $d.f.$ = number of cells in table. The derivation of this second measure is explained in full in an article by Voas and Williamson (Voas and Williamson, 2001). The underlying principle behind this second measure is the use of a modified Z-score for each benchmark count that takes into account not only the proportional difference between observed and expected counts (as with a normal Z-score), but also the absolute difference between estimated and observed benchmark table totals. The resulting sum of squared modified Z-scores for each benchmark table is divided by the relevant table-specific χ^2 critical value to standardise for the number of benchmark counts in each benchmark table. These table-specific relative scores are then summed to yield the overall measure, the main focus of which is upon proportional rather than absolute fit.

Whereas the results presented in this paper (Sections 4 and 5) mainly exclude 'non-convergent' (i.e. very poorly fitting) GREGWT estimates, they include all CO estimates, whether or not these estimates satisfied the minimum fit thresholds specified for triggering early termination of a CO

run on the grounds that non-convergence, *per se*, is not an issue for CO.

3.2. Weights

Because of the methods used, GREGWT produces floating point weights - whereas CO produces integer weights. There is no real advantage to either type of weight. In fact, the CO routine could implement floating point weights by adding partial 'units' of individuals, rather than whole records.

3.3. Efficiency

Both the CO and GREGWT routine are computationally intensive. In testing the different algorithms on exactly the same computer (dual processor dual core 2 Ghz processor, 2 GB Memory), the CO routine calculated weights for the 107 SLAs in the Australian Capital Territory in about ½ hour; where the GREGWT routine took 2 ½ hours. This was possibly due to the way the algorithms are coded (GREGWT is in a SAS macro; whereas CO uses compiled FORTRAN code), but may also reflect algorithmic efficiency.

3.4. Summary of differences

A summary of the differences in each method is shown in Table 2.

Table 2   Comparison of methods in summary

	GREGWT	CO
Approach	National household weights from a national survey dataset are reweighted to household weights for SLAs by constraining to small-area census counts	Selection of a combination of households from a national survey microdata set that best fit small-area census counts
Weights	In fractional numbers	In integer numbers
Preparation of census data	Needs to address re-allocation of 'not-stated' and 'not applicable' counts	Needs to address re-allocation of 'not-stated' and 'not applicable' counts
Conflicting benchmark counts due to statistical disclosure measures	Causes non-convergence because no set of weights can be found that simultaneously satisfies all benchmarks	Seeks to minimise the difference between the final weights and the target benchmarks which typically results in weights that match the average of any discrepant benchmarks
Optimisation strategy	Algorithm reaches an optimised solution when residual (i.e. difference between an synthetic estimate and the benchmark count) approaches zero	Minimise absolute or proportional error
'Convergent' & 'non-convergent' SLAs	In some cases no convergent solution may be found; Average Household Absolute Sum of Residuals is >1 provides a proxy indicator for this non-convergence.	No convergence issues, although final 'optimal' estimates may still fail to fit all user-supplied benchmarks

Source: NATSEM (GREGWT) and Williamson (CO)

4.   RESULTS FROM EACH METHOD

This section shows summary results from the estimates produced by GREGWT and CO for the 307 SLAs in the ACT (the Australian Capital Territory) and NSW (New South Wales). NSW contains a range of urban and rural SLAs fairly representative of Australia as a whole, whilst the ACT contains some SLAs which are highly atypical in an Australian context, containing high concentrations of professionals and students. Results from GREGWT exlclude the 14% of SLAs classified, as per Section 3.1, as 'non-convergent'. These non-convergent SLAs are predominantly less populous SLAs drawn from the industrial and remote areas of the ACT and NSW. Only 4.5% of the ACT population were in non-converging areas, and 4.6% of the NSW population.

The first set of results we consider here show how well each method has hit the benchmarks specified. The second set of results, more interestingly, show the usefulness of each method for predicting values not present in the benchmarks.

Predicting the benchmarks (constrained variables) is of limited interest, as by definition we already know their values. On the other hand, the difference between the estimated (weighted survey) and observed (census) values does at least provide some initial indicator of estimate quality. More usefully, the weighted data also yield estimates for unbenchmarked values. Two kinds of value-added estimates may be identified. The first involves the unbenchmarked interactions between benchmarked variables (margin constrained estimates). The second involves the unbenchmarked interactions between unbenchmarked variables (unconstrained estimates). The greater the degree of correlation between the benchmark constraints and the unbenchmarked estimates, the greater the quality of the estimate is likely to be. In constrast, values estimated using unbenchmarked variables that have no correlation with the benchmarked variables will necessarily be highly unreliable.

In this paper we evaluate the efficacy of our small area estimates in predicting housing stress. Housing stress is directly correlated with three of our constrained variables: income, rent paid and mortgage paid. A household is defined here as being in housing stress when they spend more than 30 per cent of their gross income on rent or a housing loan (a definition that can be matched from the ABS Census data so it can be validated against reliable small area data).

The average SLA-level fit to the benchmarks listed in Table 1 is summarised in Tables 4 and 5 for two States in Australia, New South Wales and the Australian Capital Territory, using a range of summary statistical measures described in Table 3. There are 107 areas estimated in the Australian

Capital Territory and 199 in New South Wales. The SLAs in the ACT are atypical for Australia, both in terms of socio-demographic composition and in terms of size (ACT SLAs contain considerably fewer people than the average Australian SLA). The SLAs in NSW may be regarded as more 'typical' Australian SLAs.

Table 3   Summary measures of goodness of fit

Measure	Description
Overall Total Absolute Error (OTAE)	Absolute Sum of Residuals summed across all benchmark counts
Overall Total Absolute Error per household (OTAE/HH)	Absolute sum of residuals per household across all benchmark counts
Overall Total Absolute Proportional Error (OTAPE)	Absolute difference between benchmark counts when expressed as fraction of the table total
Overall relative sum of Z-square scores (ORSumZ2)	For each benchmark table, the Z-score of each benchmark count squared, and summed for the table; then divide by chi-square critical value for table (--> RSumZ2), then sum across all tables (-->ORSumZ2). For a given table, RSumZ2>1 shows it is not fitting.

The results are shown in Table 4 and Table 5, averaged across the SLAs for which GREGWT produced 'convergent' estimates.

Both variants of CO produced a better proportional fit to the estimation benchmarks than GREGWT (lower ORSumZ2 and OTAPE) but performed more variably when it came to matching GREGWT's absolute fit to the estimation benchmarks. Of the two CO variants, CO (Min Proportion) unsurprisingly produced by far the lowest proportional error for both States; more surprisingly it also produced the lowest absolute error in NSW. This may reflect the fact that the SLAs in NSW are more populous, making the link between absolute and proportional values more tenuous. Overall the evidence presented in Tables 4 and 5 suggests a performance advantage favouring selection of CO (Min Proportion) over CO (Min Absolute), if such a choice has to be made.

Table 4   Results for constrained variables, Australian Capital Territory
(GREGWT 'convergent' SLAs only)

Measure	GREGWT	CO (Min Proportion)	CO (Min Absolute)
OTAE	139.6	133.4	92.2
OTAE/HH	0.1	0.1	0.1
OTAPE	0.4	0.2	0.2
ORSumZ2	48.4	0.5	27.8

Note: Lower numbers signify greater accuracy

Table 5   Results for constrained variables, New South Wales
(GREGWT 'convergent' SLAs only)

Measure	GREGWT	CO (Min Proportion)	CO (Min Absolute)
OTAE	602.9	483.1	979.3
OTAE/HH	0.1	0.1	0.1
OTAPE	0.2	0.1	0.2
ORSumZ2	60.5	1.9	29.2

Note: Lower numbers signify greater accuracy

As mentioned above, the main use for these reweighting techniques is to get estimates for variables that were not on the Census. We should be able to get reasonable estimates for variables that are correlated with the benchmarked variables (margin-constrained variables). For this paper, we calculated estimates of housing stress, which are correlated with some of the benchmarked variables (Income and Housing costs). Estimates of housing stress supplied by the Australian Bureau of Statistics from Census data provide an independent source against which to compare our own estimates.

A comparison of our various modelled estimates with those supplied by the ABS is presented in Table 6 for the Australian Capital Territory and New South Wales. It can be seen that all the models produced reasonable estimates of proportions, with the GREGWT estimates coming closest to the actual ABS per cent figures.

Table 6   Results for Housing Stress, Australian Capital Territory and New South Wales
(GREGWT 'convergent' SLAs only)

State	ABS	GREGWT	CO (Min Proportion)	CO (Min Absolute)
Number Unaffordable				
Australian Capital Territory	5,526	6,147	5,924	5,821
New South Wales	169,823	194,394	191,720	189,269
Total	175,349	200,541	197,644	195,090
% Unaffordable				
Australian Capital Territory	5.9	5.9	5.7	5.6
New South Wales	9.1	9.2	9.1	8.9
Combined	9.0	9.0	8.9	8.8

In terms of the number of people in housing stress, Table 6 appears to suggest that neither GREGWT nor CO estimate absolute numbers all that well. However, in considering Table 6 it should be borne in mind that the population base for the independent ABS housing stress estimates includes only households providing full returns on income and housing costs via their Census form. In contrast, the GREGWT and CO estimates have been weighted to fit benchmarks in which non-response households have been included via pro rating (see section 3). Given that pro rating should more or less preserve the proportional distribution of households by income and housing cost, and given that the CO and GREGWT estimates closely replicate

ABS estimates of the proportion of households in housing stress, it could in fact be argued that it is actually the ABS absolute estimates that are out of line. In any case, it is certainly true that GREGWT and CO outputs do need to be recognised as modelled estimates. As such, estimated proportions are more likely to be accurate than estimated levels - and so results are best presented as proportions or grouped into quantiles.

Another way of looking at these results is to look at the correlation between the different methods for each SLA in the area. In many recent papers on spatial microsimulation, this has been done using the Standard Error around Identity (SEI) (Ballas *et al.*, 2007; Edwards and Tanton, 2013). The advantage of the SEI is that it provides a precise measure of the accuracy of the modelled estimates around reliable small area estimates.

In this work, we have used a correlation coefficient rather than the SEI. The reason for this was twofold. The first was that the SEI assumes a slope of 1. In all three of the estimate comparisons made here (Figures 1 - 3), the slope is less than 1. This means that the CO and GREGWT estimates are typically higher than the ABS estimates for low levels of housing stress; and vice versa for high levels of housing stress. The second was that the SEI assumes an intercept of 0. In all three graphs, the intercept is not 0, suggesting a consistent bias in our results that the SEI would treat as error.

This means there is some obvious bias in our estimates, and in these cases the coefficient of determination (R^2) may be a better estimate of the accuracy of the estimates, because of the enforced 0 intercept and slope of 1 for the SEI. Recent work by Vidyattama *et al.* (2013) used both the R^2 and the SEI to show up a slight bias in their estimates due to the different datasets used in the spatial microsimulation process.

The graphs below (Figures 1 – 3) show the correlations between the ABS estimate and the different estimation methods we have used for all convergent SLAs in the Australian Capital Territory and New South Wales. It can be seen that the correlations are all very high (0.86 – 0.89). The highest correlation (and therefore lowest error) is from using the CO-Min Proportion model. Clearly the ranking of the three model estimates depends upon the precise measure used. But overall all three approaches appear to have done a good job of estimating the unknown three-way interaction between income and tenure-specific housing cost, from which the final estimate of housing stress is derived.

Figure 1: SLA level Housing stress estimates (%)
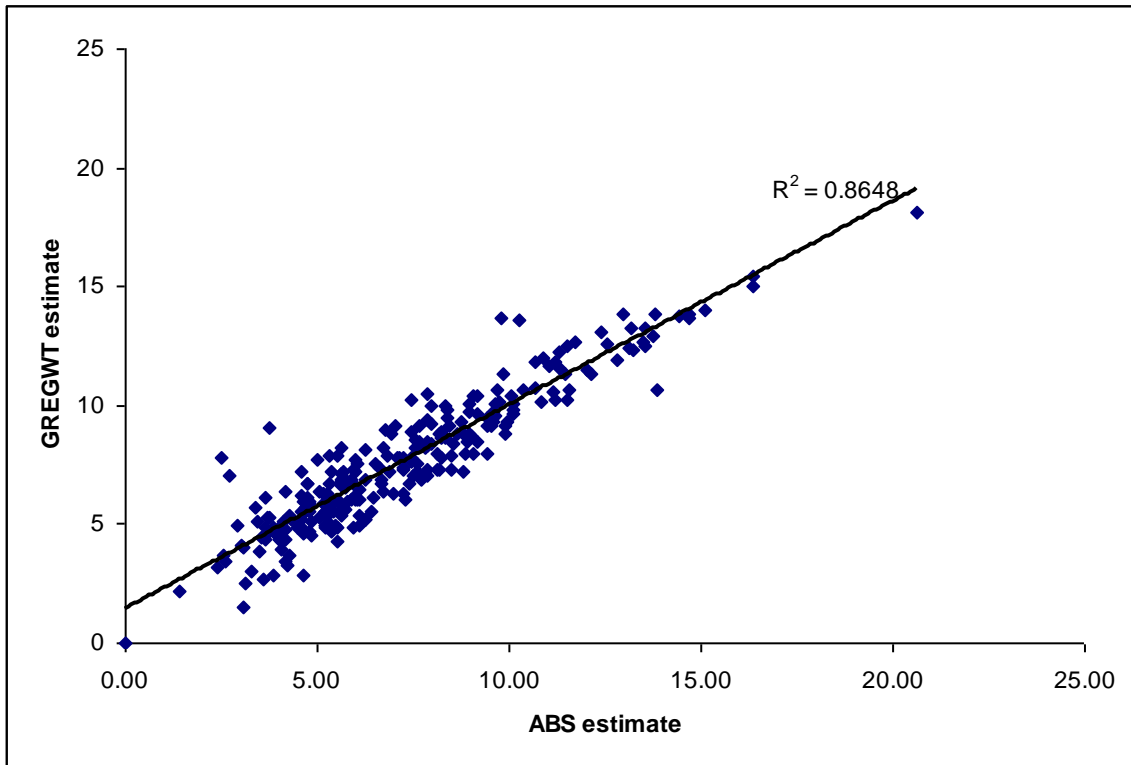*GREGWT (GREGWT convergent SLAs)*

Figure 2: SLA level Housing stress estimates (%)
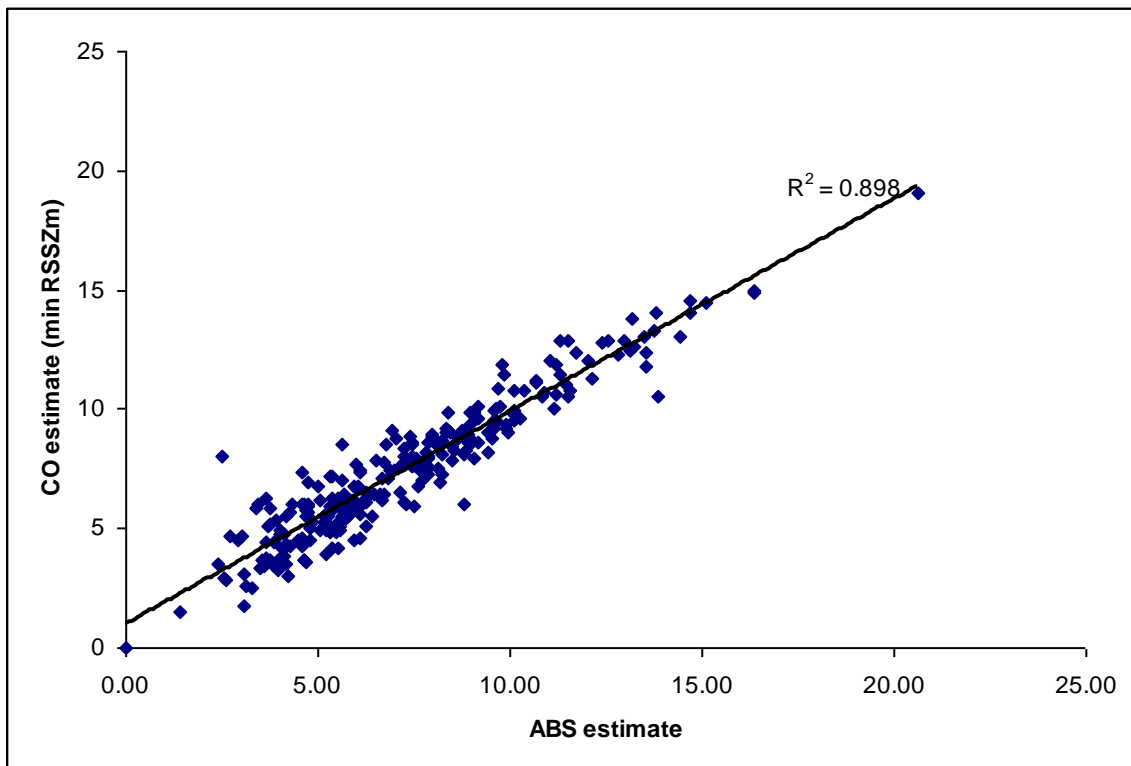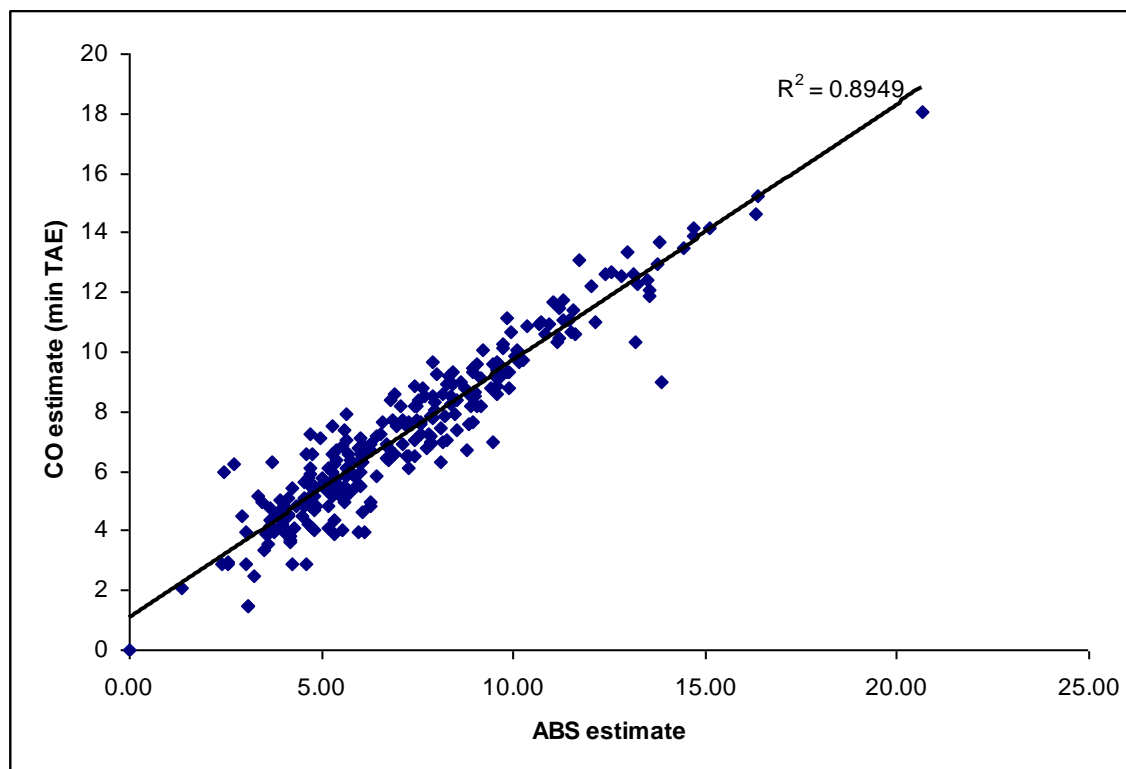*CO-Min Proportion (GREGWT convergent SLAs)*

Figure 3: SLA level Housing stress estimates (%): CO-Min Absolute
            (GREGWT convergent SLAs)

5.   FURTHER ANALYSIS OF THE WEIGHTS DERIVED BY CO AND GREGWT

While Section 4 has shown that the results from estimating different variables using these two methods are similar, the weights derived by each method are quite different. The CO routine derives integer weights - whereas the GREGWT routine doesn't. We also expect more zero weights from the CO routine than we get using GREGWT. This section looks at the distribution of these weights.

The number of weights derived by each routine is massive. There are 6,892 households on our survey file that we derive weights for. In the Australian Capital Territory, there are 107 SLAs. So in the dataset of weights for the Australian Capital Territory, there are a total of about 740,000 weights calculated. For New South Wales, with 199 areas, there are about 1.4 million weights calculated.

Table 7 shows the size distribution of the weights from CO and GREGWT. The CO routine only uses integer weights, so if there are fewer households in an SLA than in a survey, there will inevitably be survey households with a weight of 0. In contrast, GREGWT shares out the weights in small fractions across a large number of households. It can be seen, therefore, that the

CO routine produces many more 0 weights than GREGWT.

Table 7   Size of weights from CO and GREGWT
(Table legend)

Method	Australian Capital Territory			New South Wales		
	0	>0 – 1[a]	>1	0	>0 – 1[a]	>1
	%	%	%	%	%	%
CO	95	3	1	79	8	13
GREGWT	53	47	1	36	48	16

a.   For >0 - 1, the CO value is 1.

This also means that the CO routine is relying on fewer households to calculate the values of housing stress in the previous example. The GREGWT routine will use more households, with lower weights. The distribution of the weights from GREGWT for the Australian Capital Territory and New South Wales is shown in Figure 4 and Figure 5. Both these frequency distributions have class boundaries increasing by 0.01 until the value of one is reached, and the final category is weights greater than one. It can be seen that most of the weights in the Australian Capital Territory are below one (99 per cent of them, according to Table 7); in New South Wales, about 15 per cent are above one, but of these, half of them are under 2.

Frequency distributions have not been shown for the CO routine, as 95 per cent of the weights are 0 for each State, so the frequency distributions are dominated by this.

The other interesting statistic to look at with the weights is the maximum and average weights. For the GREGWT routine, non-convergence can lead to weights that are ridiculously large (in the order of 100,000 or more). Those areas where non-convergence occurs are usually less populous areas in the Northern Territory and West Australia, or industrial areas in cities where few people live. This is the case in the ACT and NSW.

One solution to the convergence issue is to reduce the number of benchmarks, or aggregate the benchmarks differently. Reducing the number of benchmarks is now standard in the SpatialMSM model to derive estimates for more areas, but this also affects the accuracy of the estimates, as with fewer benchmarks the estimates become less accurate against the benchmarked Census data. To enable consistency across all our areas estimated using SpatialMSM, we have not reduced the number of benchmarks or aggregated them differently in this analysis.

Excluding SLAs for which GREGWT produced non-convergent estimates, the maximum and average weights produced by GREGWT and CO are shown in Table 8. It can be seen that the CO routine produces a higher maximum for the Australian Capital Territory - but GREGWT

produces a higher maximum for New South Wales.

The average values shown in Table 8 are calculated without zeroes, as there are so many in each procedure that they dominate an average. The values in Table 8 show that the CO routine calculates higher weights on average than the GREGWT routine for both New South Wales and the Australian Capital Territory. This is to be expected, given the number of zero weights calculated by the CO routine. To get to the same population in an area, the positive weights must be higher in the CO routine than the GREGWT routine, which has fewer zero weights.
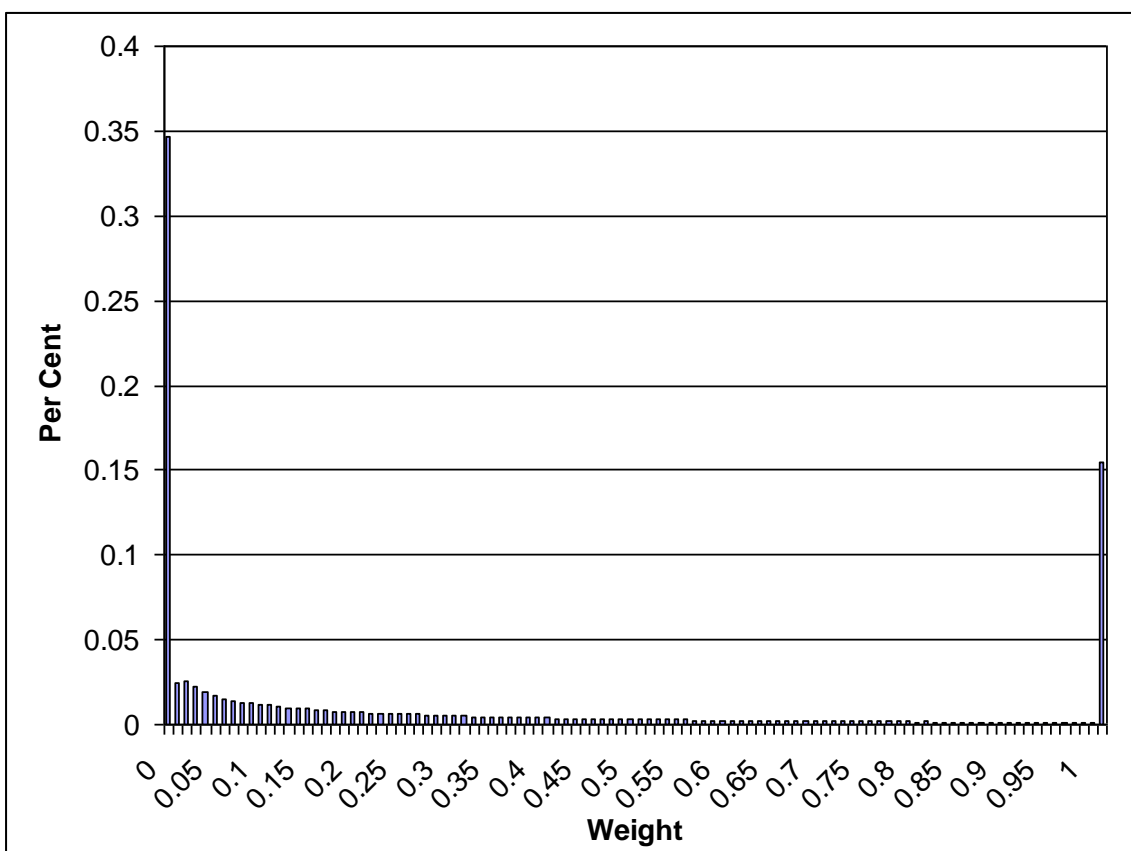
Figure 4: Distribution of GREGWT weights for NSW

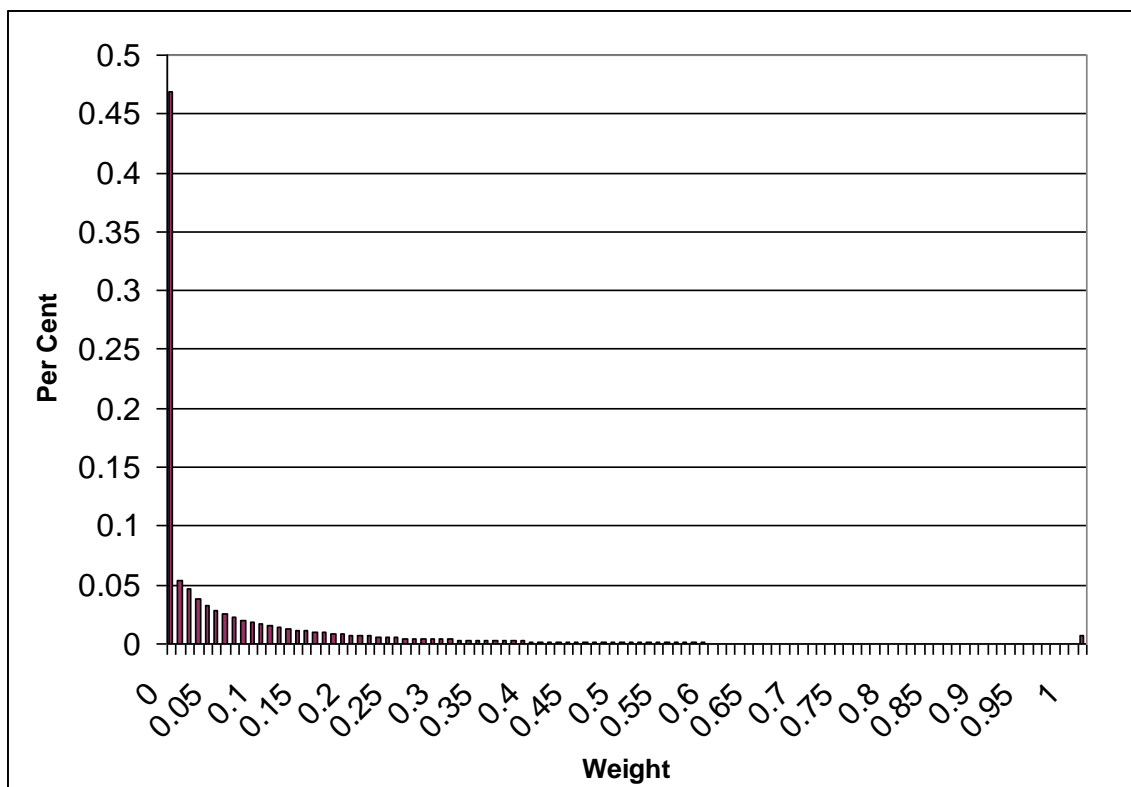**Figure 5: Distribution of GREGWT weights for the Australian Capital Territory**

Table 8   Weights for GREGWT and CO
(GREGWT convergent SLAs only)

Method	Maximum		Average non-zero value	
	New South Wales	Australian Capital Territory	New South Wales Average	Australian Capital Territory Average
CO (Min Proportion)	443	24	3.49	1.45
GREGWT	647	18	1.11	0.15

This section has shown that even though the weights from the CO and GREGWT algorithm give similar results when calculating variables like housing stress, they are actually very different. The CO routine tends to include fewer households, but give them higher weights — while the GREGWT routine will select more households to represent an SLA, but will give them smaller weights.

6.   CONCLUSIONS

The GREGWT algorithm has been shown to be capable of producing good results, but 14% of the total number of SLAs did not converge. These areas are industrial areas and remote areas in the ACT and NSW, and are usually less populous areas. Only 4.5% of the ACT population were in non-converging areas, and 4.6% of the NSW population.

The GREGWT algorithm also takes a long time to run compared to CO, particularly when estimating a large number of areas. It is unknown whether this is due to the programming language being used for each algorithm (CO uses a compiled FORTRAN code, whereas GREGWT is a macro running in SAS) or whether this reflects the relative efficiencies of the underlying algorithms.

Head-to-head, when results are compared for those SLAs for which GREGWT converged, the fit to benchmarks and estimates of housing stress produced by GREGWT and CO (Min Proportion) are broadly comparable. On balance, however, the CO Min Proportion model is perhaps slightly to be favoured. It has a better proportional fit to benchmarks (Table 4), the lowest error when estimating each SLA's value (Figure 2), and produces reasonable estimates when the SLA values were aggregated to State (see Table 6). In addition, for those SLAs for which GREGWT produces no usable estimate, CO (Min Proportion) appears to continue to produce estimates of reasonable quality. CO (Min Absolute) competes slightly less well head to head with GREGWT, and in almost all circumstances produces estimates of at least marginally lesser quality than those offered by CO (Min Proportion).

ACKNOWLEDGEMENT

REFERENCES

ABS (2005) Small area estimates of Disability in Australia, Canberra: ABS, Pub. # 1351.0.55.006

ABS (2006) A guide to small area estimation - Version 1.0, http://www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/3a60738d0abd f98cca2571ab00242664/$FILE/May%2006.pdf, Last accessed 19 May 2014

ABS (2012) Wage and salary earner statistics for small areas, Time series, 2009 - 10, Cat # 5673.0.55.003, ABS: Canberra

ABS (2011) Small Area Estimation with Simulated Samples from the Population Census. Cat # 1352.0.55.106, ABS: Canberra.

Anderson, B (2007). "Creating Small-Area Income Estimates: Spatial Microsimulation Modelling." UK Department of Communities, Accessed from http://www.communities.gov.uk/documents/communities/pdf/325286.pdf, 21 Feb 2014

Ballas, Clarke, Dorling, and Rossiter (2007) "Using SimBritain to Model the Geographical Impact of National Government Policies." *Geographical Analysis 39(1):* 44–77.

Bates, A (2006) 'Methodology used for producing ONS's small area population estimates', *Population Trends*, 125, 30 – 36.

Bell, P (2000) GREGWT and TABLE macros - Users guide, Canberra: ABS, Unpublished.

Bond, S and Campos, C (2010) 'Understanding income at small area level', *Regional Trends 42: 2010 edition*, London: Office for National Statistics, 80-94.

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 Paper 36, NATSEM

Cassells, R., Miranti, R. and Harding, A (2013). Building a static spatial microsimulation model: Data preparation. In R. Tanton & K. Edwards (Eds.), *Spatial Microsimulation: A Reference Guide for Users*, pp. 87–103. Springer: Netherlands.

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

Clarke, G., Kashti, A. McDonald, A. and Williamson, P. (1997) "Estimating Small Area Demand for Water: A New Methodology." *Water and Environment Journal* 11 (3): 186–192.

Commonwealth Department of Employment and Workplace Relations (2007) Small Area Labour Markets, Commonwealth Government.

Edwards and Tanton (2013). "Validation of Spatial Microsimulation Models." In Tanton and Edwards (eds), *Spatial Microsimulation: A Reference Guide for Users*, 249 – 258. Springer.

EURAREA (2005) 'Enhancing Small Area Estimation Techniques to meet European Needs', Project Reference Volume, The EURAREA Consortium.

Harding, A, Lloyd, R, Bill, A and King, A (2006) Assessing poverty and inequality at a detailed

regional level: New advances in spatial microsimulation, Helsinki: United Nations University Press.

Harding, A, Vidyattama, Y and Tanton, R (2011). "Demographic Change and the Needs-Based Planning of Government Services: Projecting Small Area Populations Using Spatial Microsimulation." *Journal of Population Research 28* (2-3): 203–224.

Harland, K., Heppenstall, A., Smith, D., & Birkin, M. (2012). Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques. *Journal of Artificial Societies and Social Simulation*, 15(1), 1–24.

Heady, P, Clarke, P, Brown, G, Ellis, K, Heasman, D, Hennell, S, Longhurst, J and Mitchell, B (2003) Model-Based Small Area Estimation Series No. 2 - *Small Area Estimation Project Report*, Office of National Statistics.

Hermes, K and Poulsen, M (2012). "A Review of Current Methods to Generate Synthetic Spatial Microdata Using Reweighting and Future Directions." Computers, Environment and Urban Systems 36 (4): 281–290.

Hynes, S, O'Donoghue, C, Morrissey, K and Clarke, G (2009). "A Spatial Micro-Simulation Analysis of Methane Emissions from Irish Agriculture." *Ecological Complexity*, 6 (2): 135–146.

Kavroudakis, D (2013) 'sms: Spatial Microsimulation', R Package, http://cran.r-project.org/web/packages/sms/index.html

Lehtonen, R and Veijanen, A (2009) 'Design-based methods of estimation for domains and small areas' in: C. R. Rao and D. Pfeffermann (eds.), *Handbook of statistics*, vol. 29(B). Sample surveys: theory, methods and inference. Elsevier.

Marshall, A (2010) 'Small area estimation using ESDS government surveys – an introductory guide', ESDS Government, Economic and Social Data Service, University of Essex.

O'Donoghue, C, Morrissey, K. and Lennon, J (2014). 'Spatial Microsimulation Modelling: a Review of Applications and Methodological Choices', *International Journal of Microsimulation*, 7(1) 26-75.

Pfefferman, D (2002) 'Small area estimation – new developments and directions', *International Statistical Review*, 70, 125-143.

Phillips, B, Chin, SF and Harding, A (2006) 'Housing Stress Today: Estimates for Statistical Local Areas in 2005', Paper given at Australian Consortium for Social and Political Research Incorporated Conference, Sydney, 10-13 December 2006

Purcell, N and Linacre, S (1976) 'Techniques for the Estimation of Small Area Characteristics', Paper given at Third Australian Statistical Conference, 18 - 20 August 1976

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

Tanton, R. (2011). Spatial microsimulation as a method for estimating different poverty rates in Australia. *Population, Space and Place*, 17(3), 222 – 235.

Tanton, R. (2014), A Review of Spatial Microsimulation Methods, *International Journal of Microsimulation*, 7(1) 4-25.

Tanton, R., and Edwards, K. L. (2013a). Spatial Microsimulation: A Reference Guide for Users, Understanding Population Trends and Processes, Vol. 6, Springer: Netherlands

Tanton, R and Edwards, K. L. (2013b). "Introduction to Spatial Microsimulation: History, Methods and Applications." In R. Tanton & K. Edwards (Eds.), *Spatial Microsimulation: A Reference Guide for Users*, 3–8. Springer: Netherlands.

Tanton, R., Harding, A., and McNamara, J. (2013). Spatial Microsimulation Using a Generalised Regression Model. In R. Tanton & K. Edwards (Eds.), *Spatial Microsimulation: A Reference Guide for Users*, 87–103. Springer: Netherlands.

Tanton, R., McNamara, J., Harding, A., and Morrison, T. (2009). Small Area Poverty Estimates for Australia's Eastern Seaboard in 2006. In A. Harding & A. Zaidi (Eds.), *New Frontiers in Microsimulation Modelling* 79 – 97. Vienna, Austria: Ashgate.

Tanton, R, Vidyattama, Y, McNamara, J, Vu, Q and 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): 102–120.

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)*, 174(4),

931–951.

Tanton, R, Jones, R and Lubulwa, G (2001) 'Analyses of the 1998 Australian National Crime and Safety Survey', Paper given at The Character, Impact and Prevention of Crime in Regional Australia, Townsville, 2 - 3 August 2001.

Vidyattama, Y, McNamara, J, Miranti, R, Tanton, R and Harding, A (2013). "The Challenges of Combining Two Databases in Small-Area Estimation: An Example Using Spatial Microsimulation of Child Poverty." *Environment and Planning* A 45 (2): 344–361.

Voas, D and Williamson, P (2001) 'Evaluating goodness-of-fit measures for synthetic microdata', *Geographical and Environmental Modelling*, 5(2), 177 – 200.

Williamson, P (2007) CO Instruction Manual, Working Paper 2007/1, Population Micordata Unit, Dept. of Geography, University of Liverpool.

Williamson, P. (2013). An Evaluation of Two Synthetic Small-Area Microdata Simulation Methodologies: Synthetic Reconstruction and Combinatorial Optimisation. In R. Tanton & K. Edwards (Eds.), *Spatial Microsimulation: A Reference Guide for Users*, pp. 19–47. Springer: Netherlands.

Williamson, P., Birkin, M. and 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), 785-816.

Williamson, P, Mitchell, G. and McDonald, A (2002). "Domestic Water Demand Forecasting: A Static Microsimulation Approach." *Water and Environment Journal* 16 (4): 243–248.