SPATIAL MICROSIMULATION USING SYNTHETIC SMALL-AREA ESTIMATES OF INCOME, TAX AND SOCIAL SECURITY BENEFITS

Shih-Foong Chin
Senior Research Fellow, National Centre for Social and Economic Modelling (NATSEM), University of Canberra ACT 2601

Ann Harding
Professor of Applied Economics and Social Policy and Director, NATSEM, University of Canberra ACT 2601

Rachel Lloyd
Principal Research Fellow, NATSEM, University of Canberra ACT 2601

Justine Mc Namara
Senior Research Fellow, NATSEM, University of Canberra ACT 2601

Ben Phillips
Senior Research Fellow, NATSEM, University of Canberra ACT 2601

Quoc Ngu Vu
Senior Research Officer, NATSEM, University of Canberra ACT 2601

ABSTRACT: A spatial microsimulation technique using the 'reweighting' approach has been developed at NATSEM to produce synthetic estimates for small areas. This is a novel way to create small-area socio-economic data that are otherwise unavailable from other sources. This paper reports on the application of the technique to produce estimates of income, tax, and social security benefits for the Statistical Local Areas (SLAs) in three States and one Territory in Australia. The spatial technique utilises the *1998-99 Household Expenditure Survey* (HES) data and the 2001 Census data, both from the Australian Bureau of Statistics. The process involves the 'reweighting' of the household unit record weights at the national level obtained from the HES to produce household unit record weights for each SLA. These SLA household weights are then applied to the selected output variables generated by NATSEM's STINMOD Model to produce the required small-area synthetic estimates for SLAs. This paper explains how spatial microsimulation is undertaken to 'regionalise' STINMOD. It presents the results of the synthetic small-area estimates and measures the reliability of these estimates against Census counts and against data from independent sources. This paper also provides an example to illustrate how the spatial technique may be used to assess the local impact of government policy changes, by simulating the spatial effects of the tax cuts announced in the recent Australian Government Budget.

1. INTRODUCTION

This paper reports on the application of a spatial microsimulation technique developed at NATSEM for generating synthetic, small-area, socio-economic data to support policy decisions in the four participating states and territory of NSW, Victoria, Queensland and the ACT, under a joint-funding involving these jurisdictions and the Australian Research Council.

The paper is divided into two main parts. The first part (comprising Sections 2, 3 and 4) explains how spatial microsimulation could be used to produce small-area estimates, what the resultant estimates are, and how the reliability of these results may be measured. The second part (comprising Section 5) demonstrates how the spatial technique may be used to simulate the local impact of changes in government policy, by using as an example, the recent tax cuts announced in the May 2005 Australian Government Budget for 2005-06.

NATSEM research findings are generally based on estimated characteristics of the population. Such estimates are usually derived from the application of microsimulation modelling techniques to microdata based on sample surveys. These estimates may be different from the actual characteristics of the population because of sampling and non-sampling errors in the microdata and because of the assumptions underlying the modelling techniques. The microdata do not contain any information that enables identification of the individuals or families to which they refer.

2. SPATIAL MICROSIMULATION

2.1 Overview

Microsimulation is the use of microdata – i.e. data at the level of individual persons or households – to model real life events such as changes in government policy and to simulate how these events may impact on individuals. Spatial microsimulation simply adds to the simulation a spatial dimension, by creating and using synthetic microdata for small areas. Because such spatial microdata are usually unavailable, they need to be synthesized.

This paper reports on the creation of small-area estimates of income, tax, and selected social security benefits. The overall process is shown in Figure 1. It involves two major steps. First, a spatial microsimulation technique called 'reweighting' is used to create household weights for small areas (output #1 in Figure 1). Second, these small-area household weights are then applied to the income, tax and selected social security benefits for households obtained from NATSEM's static microsimulation model – STINMOD (output #2)[1] The end results are the synthetic small-area estimates of income, tax and selected social security benefits (output #3). The various components of the spatial

[1] NATSEM's static model – STINMOD, simulates the impact of major federal government cash transfers, income tax and the Medicare levy on individuals and families in Australia. The model is used by the Australian Treasury, the Department of Family and Community Services, and other government agencies, in policy formulation.

microsimulation depicted in Figure 1 will be explained in the ensuing discussion.

In this study, the geographical units used for creating small-area data are the Statistical Local Areas (SLA). The SLA is a standard geographical unit maintained by the ABS for the purpose of disseminating spatial statistics. In the 2001 Census year, there were 1,353 SLAs covering all of Australia. For New South Wales (NSW), Victoria and most parts of Queensland, an SLA maps to a Local Government Area, which is a gazetted legal entity. For the Australian Capital Territory (ACT) and for the Brisbane area in Queensland, an SLA maps to a suburb (ABS, 2001).

ABS data shows that as at June 2003, the average estimated residential population for the SLAs is 33,942 for NSW, 23,987 for Victoria, 8,098 for Queensland and 3,690 for the ACT. These populations are very heterogenous, as indicated by the respective standard deviations of 45,151 (NSW), 24,690 (Victoria), 8,502 (Queensland) and 3,235 (ACT).

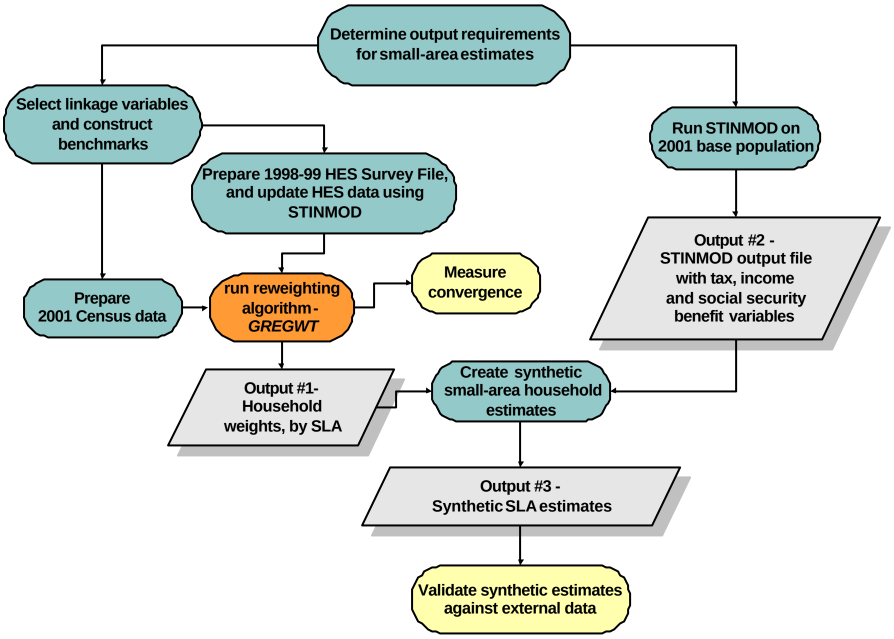


Figure 1. Creation of Small-Area Synthetic Household Estimates of Tax, Income and Social Security Benefits for 2001

2.2 Creation of Small-Area Household Weights Using Reweighting

There are several techniques for creating small-area weights: data matching or fusion, iterative proportional fitting (Tranmer et al, 2005), synthetic reconstruction and combinatorial optimisation (Williamson et al, 2001), and reweighting (Ballas, et al, 2003). This project uses the reweighting approach. As the name suggests, reweighting 'reweights' (or converts) a set of unit record weights (e.g. household weights) at a larger spatial unit (e.g. Australia) to sets of unit record weights at a smaller spatial unit (e.g. SLA). One set of weights is created for one SLA.

Linking survey and Census data

Reweighting is typically achieved by linking a national survey file and a Census file. The idea behind the linking is that while a national survey file contains rich microdata information of interest (e.g. employment status, household income etc.), it has very poor spatial information since the data are reported at a highly aggregated (national) level. In contrast, the Census data are spatially rich because they are often available at much finer levels of geography such as the SLA. By combining the two sets of data, and by benchmarking the rich information from a national survey to the small-area characteristics in the Census, the former can be given a small-area, spatial dimension.

The reweighting process reported here combines the 1998-99 Household Expenditure Survey (HES) Confidentialised Unit Record File (ABS, 2002) with the 2001 Census Expanded Community Profile (XCP) datasets (ABS, 2003). Both were obtained from the ABS.

Linkage variables and their selection

The reweighting process begins with the selection of the appropriate variables from the HES survey file. These variables need to satisfy two conditions. One, they need to adequately represent the key characteristics associated with the output estimates we are trying to synthesize — in this case income, tax and social security benefits. For this purpose, regression analysis was first carried out to identify and select those variables in the HES survey file that were relatively highly correlated with the dependent variables of income and social security benefits. Two, the selected HES variables also need to be present in the Census XCP datasets because the reweighting process requires linking the HES data with the Census data. The selected variables, which are common to both the HES and the Census, are known as linkage variables. The linkage variables used in this report are: Age, Sex, Occupation, Labour Force Status and Education Level, of the person; and Household Type[2], Household Size, Tenure type, Amount of Mortgage Paid and Amount of Rent Paid, of the household.

Benchmarks for reweighting

Benchmarks are the constraints against which the HES data are reweighted. In other words, they are the Census counts which the HES estimates produced by reweighting are trying to meet. A benchmark can be a one-dimensional linkage variable (e.g. Age), or a combination of two or more linkage variables (e.g. Age by Sex by Labour Force Status). Each benchmark contains a number of discrete classes such as '$100-$199' weekly rental, or 'male' and 'female'.

---

2 The ABS defines a household as a group of related or unrelated people who usually live in the same dwelling and make common provision for food and other living arrangements; or a lone person who makes his or her own provision without combining with any other person.

The selection of benchmarks needs to balance two opposing requirements. On the one hand, we need to maximise the number of benchmarks and their classes because this increases the level of information in our estimates (e.g. “males over 65” is more informative than “total number of males”). On the other hand, we need to be mindful that an increase in benchmarks and benchmark classes reduces the ability of the reweighting process to converge – good convergence means achieving a small or nil difference between the estimate and the Census count.

After due deliberation, twelve benchmarks and 116 benchmark classes were selected for reweighting (Table 1). The benchmarks were predominantly single-dimensional (i.e. non-cross tabulated). All benchmarks except one were ‘fully-specified’ (meaning they cover the entire population). Finally, the benchmarks came in three different levels (i.e. persons, family and household level).

Table 1. Benchmarks and Benchmark Classes for Reweighting

Benchmark	Description	Level	Fully specified	Benchmark classes (no.)
1	Age by sex by labour force status	Person	Yes	40
2	Occupation	Person	Yes	10
3	Individual weekly income	Person	Yes	12
4	Non-school qualification	Person	Yes	6
5	Education status	Person	Yes	4
6	Family type	Family	No[1]	3
7	Household type	Household	Yes	2
8	Weekly household income	Household	Yes	11
9	Dwelling tenure type	Household	Yes	6
10	Monthly mortgage	Household	Yes	10
11	Weekly rental	Household	Yes	5
12	Number of persons per household	Household	Yes	7
		Total number of benchmark classes		*116*

Note: [1] The benchmark classes excluded are: “Non-private Dwelling”, “Other Family” and “Non-Family Member”. **Source:** NATSEM

Reweighting algorithm

The reweighting process uses a generalised regression routine written in the SAS programming language and developed by the ABS called GREGWT (ABS, 2000, and Bell, 2000). This routine carries out iterative calculations to derive an ‘optimal’ set of household weights for each SLA which, when applied to the HES unit record file, will produce estimates that best fit the Census counts for that SLA. In other words, optimisation is reached when the difference (or ‘residual’) between the estimates and the Census counts is zero or approaching zero. At this point, convergence is considered achieved.

Preparation of HES and Census data

For the GREGWT routine to work properly, some pre-processing of the HES data and the Census data needs to be done. Preparation of the HES data includes the following processes:

- updating the 1998-99 HES population to the 2001 world to render it comparable to the 2001 Census population (this is accomplished in STINMOD);
- ensuring the definition of the HES population is consistent with the 2001 Census population. For example, for the linkage variable "Individual weekly income", the HES population covers "all persons" whereas the population in the Census tables covers only "all persons employed". In this case, a subset of the HES population is taken in order to match the Census population;
- mapping the entire HES population to more than one Census table. For example, for the linkage variable "Labour force status", the HES population needs to be mapped to two separate Census XCP tables in order to ensure the entire HES population is covered.

Likewise, some preparation of the Census data is also required to address the issue of data 'perturbance' resulting from the confidentialisation process applied by the ABS to alter the true value of a data cell that has less than four counts. Because of this 'perturbance', the totals and sub-totals of different Census XCP tables with the same population base may not be consistent and this will cause the reweighting process to not converge.

This issue is addressed by 'balancing' the Census data. It involves the redistribution of non-response counts and 'balancing' the marginal totals of the Census tables. The former process involves the redistribution of non-response counts (i.e. 'not applicable' and 'not stated') to 'known' classes for the variable. The classes needing redistribution will have their counts redistributed to 'known' classes in accordance with the relative frequency of the known classes. The latter involves the determination of the 'hierarchy' of a suite of Census XCP tables used for reweighting so that subordinate totals can be adjusted to meet the equivalent 'totals' higher up in the table hierarchy.

Outputs from Reweighting

The reweighting process produces two key outputs: a dataset containing unit record household weights for every SLA in each of the four States or Territory; and a dataset containing the convergence measures. The former dataset is used for creating the small-area synthetic estimates for SLAs. The latter dataset is used for separating convergent SLAs from the non-convergent ones. Convergence is discussed in Section 4.2.

2.3 Creation of Synthetic Small-Area Estimates

Synthetic estimates for each SLA are created by applying the household weights produced by the reweighting process to a STINMOD output file containing the desired output variables by household. This process involves the following steps:

Step 1. NATSEM's STINMOD Model is run against the 2001 base population to produce an output file containing selected output variables on income, tax, and social security benefits, by household.
Step 2. The household-level output file from Step 1 is merged with the file containing the household weights by SLA to produce, for each SLA, a dataset containing the output variables and the weights, by household.
Step 3. The output variables in the SLA files from Step 2 are aggregated to SLA totals to create the synthetic estimates by SLA, in the following manner:

- For each of the categorical output variables (e.g. "employed", "unemployed"), the household weights of each category are summed to produce the SLA total (e.g. summing all household weights in the SLA for the "employed"); and
- For each of the numerical output variables (e.g. "taxable income"), the value of the output variable is first multiplied by the household weight and the weighted value is then summed to produce SLA totals, or averaged to produce SLA means.

Finally, the separate SLA datasets for each State or Territory are merged into one single dataset an example of which is shown in Table 2.

3. RESULTS – SYNTHETIC SMALL-AREA ESTIMATES

Four datasets containing the synthetic estimates for SLAs were produced for NSW, Victoria, Queensland and the ACT. Each dataset contains about one hundred output variables organised into the following groups:

- SLA code and SLA name;
- Number of persons and income units;
- The convergence measures: absolute-sum-of-residuals and average-absolute-sum-of-residuals-per-household (explained in Section 4.2);
- Estimated number of households and Census counts of households;
- Number of persons employed and unemployed, and their proportion as per cent of labour force;
- Number of wage and salary earners and their income per week;
- Number of taxpayers, their taxable income and tax paid per week;
- Number of persons receiving social security benefits from the Department of Family and Community Services (FaCS) and the dollar amount received per week. The selected benefits are: Age Pension, Disability Support Pension, Wife Pension, Widow B Pension, Carer's Pension, Parenting Payment Unpartnered Allowance, Parenting Payment Partnered Allowance, Newstart Allowance, Mature Age Allowance, Partner Allowance, Sickness Allowance, Special Allowance, Youth Allowance, AUSTUDY Allowance, Family Tax Benefit Part A and B, and Rent Assistance;
- Number of persons receiving pensions from the Department of Veterans' Affairs (DVA) and the dollar amount received per week. The selected pensions are: Disability Pension, Service Pension and War Widow Pension; and

- SLA averages and SLA rankings of selected output variables.

An example of this output for NSW is shown in Table 2, which only shows a handful of the 100 or so synthetic estimates for the top twenty SLAs in terms of convergence.

Table 2. Sample Output Dataset from Reweighting Showing Selected Synthetic Estimates for Twenty SLAs in NSW

SLA ID	SLA NAME	Convergence	No. of households	No. of aged pensioners	Total income $	Average weekly tax $	SLA ranking on average weekly tax	Per cent unemployed (%)
110055902	Newcastle (C) – Remainder	0.002	53019	14515	41537970	162	143	11
120057551	Tweed (A) – Pt A	0.003	18739	8363	11448151	123	31	12
140158100	Weddin (A)	0.003	1446	457	919785	123	30	6
135107900	Walgett (A)	0.003	2893	528	1858425	133	73	13
145103600	Gunning (A)	0.003	811	137	827766	184	159	4
130100400	Barraba (A)	0.003	885	329	488597	104	2	10
140103300	Greater Lithgow (C)	0.003	7083	1752	5502684	166	146	10
125055000	Maclean (A)	0.003	6438	2681	3634505	115	9	14
115106952	Shoalhaven (C Pt B	0.003	21410	9124	13316784	128	48	12
125050600	Bellingen (A)	0.003	4618	1463	2659907	117	16	14
125055700	Nambucca (A)	0.003	6785	2614	3571941	113	7	19
120074851	Lismore (C)– Pt A	0.003	11040	2755	7423950	131	63	13
140106750	Rylstone (A)	0.003	1463	464	954506	150	121	10
135055400	Mudgee (A)	0.003	6385	1684	4820790	147	113	8
135058150	Wellington (A)	0.004	3077	993	1880157	123	34	10
105555350	Mosman (A)	0.004	10632	1638	18211969	376	198	3
155152500	Deniliquin (A)	0.004	3027	899	2301452	136	81	6
125103754	Hastings (A) – Pt B	0.004	10013	3941	6089700	116	14	12
135151200	Brewarrina (A)	0.004	648	71	495925	155	131	11
160101250	Broken Hill (C)	0.004	8011	2476	5097015	147	114	13

Note: Convergence is measures in terms of average-absolute sum of residuals per household and is explained in Section 4.2. **Source:** NATSEM modelling.

Note that the estimates of social security benefits and veteran pension payments are not readily available at the SLA level. Therefore the use of spatial microsimulation reported here represents a novel way for generating estimates that are otherwise unavailable from other sources.

The quality of the synthetic estimates depends largely on the careful choice of the benchmarks used for reweighting. Experience shows that good estimates are produced for constrained variables (i.e. the linkage variables explicitly chosen for reweighting), and for constrained variables that have converged (i.e. those with zero residuals). For a constrained variable, better estimates are produced for those benchmark classes that are 'fully specified' (i.e. covering the entire population) than those that are allowed to 'float' (i.e. covering part of the population). Finally, for the unconstrained variables (i.e. variables not included in the reweighting but for which synthetic estimates are produced), the reliability of the estimates will depend on how well the unconstrained variables (e.g. "taxpayer number") correlate with the constrained variables (e.g. "income" and "labour force status").

4. CONVERGENCE AND VALIDATION

4.1 Overview

How reliable is the reweighting process and the synthetic estimates it produces? These are evaluated in terms of convergence and validation respectively.

Convergence measures the reliability of reweighting. It applies only to the constrained variables (i.e. linkage variables). Convergence measures how well an SLA estimate of a constrained variable matches the Census count for the same SLA. The difference between the two is known as a residual. A zero residual means the estimate coincides exactly with the Census count.

Validation, on the other hand, measures the reliability of the synthetic estimates. It measures how well the synthetic estimate for an SLA compares with the best available data from external sources for the same SLA (for example, comparing the estimated number of Aged pension recipients with the official data). Not surprisingly, such external data are often hard to get and the lack of small-area data is precisely the *raison d'etre* for synthesizing them.

4.2 Convergence

Convergence measured at the SLA level

Convergence measure

A convergence measure called the **absolute-sum-of-residuals** has been devised to define what a 'good' convergence is. This is the sum of all the differences between the synthetic estimates and the Census counts (i.e. 'residuals') in absolute terms. Absolute-sum-of-residuals was calculated for each of the 116 benchmark classes in every SLA. These residuals were then summed to produce a residual measure for each SLA. If an SLA converges across all benchmark classes, a zero absolute-sum-of-residuals would be achieved.

However, the absolute-sum-of-residuals value of an SLA is influenced by the number of households in the SLA. For example, an absolute-sum-of-residuals of '100' in an SLA with 1,000 households would represent a worse convergence than an absolute-sum-of-residuals of '500' in an SLA with 10,000 households. Thus, a relative convergence measure called the **average-absolute-sum-of-**

residuals-per-household (AV_ABSSUM_HH) was devised. This is calculated by dividing the absolute-sum-of-residuals of an SLA by the number of households in the SLA. Note that despite its name, this is not an 'average' in the true sense but rather a normalisation of the absolute-sum-of-residuals figure by the household number.

AV_ABSSUM_HH is a simple but effective measure for separating non-convergent SLAs from the convergent ones. The former are those SLAs whose household characteristics have less in common with the characteristics of a 'typical' household in the HES Survey and the latter are those SLAs whose household characteristics have more in common with a typical household in the Survey. The household weights for the non-convergent SLAs produced by the GREGWT reweighting algorithm are often very large numbers. These numbers are highly spurious and are discarded. In other words, for the non-convergent SLAs, no synthetic estimates are produced.

Convergence results

In the current work, an AV_HH_ABSSUM of '0.1' or less is used to arbitrarily define a convergent SLA. Conceptually, an AV_HH_ABSSUM of '0.1' or less means for every household in the SLA, the estimate is out by '0.1' or less. Table 3 shows the convergent results. Over 90 per cent of the SLAs in NSW, Victoria and Queensland have 'converged' using the convergent measure of AV_ABSSUM_HH of '0.1' or less.

Table 3 Convergent and Non-Convergent SLAs

AV_HH_ABSSUM £ 0.1	**NSW**	**VIC**	**QLD**	**ACT**
Convergent SLAs (% of All SLAs)	189 (96%)	191 (96%)	421 (93%)	84 (79%)
Non-convergent SLAs	9	8	32	23
AV_HH_ABSSUM £ 1	**NSW**	**VIC**	**QLD**	**ACT**
Convergent SLAs (% of All SLAs)	194 (98%)	194 (97%)	442 (98%)	89 (83%)
Non-convergent SLAs	4	5	11	18
All SLAs	198	199	453	107

Note: Excludes offshore and migratory SLAs. **Source:** NATSEM, ABS.

It should be noted that non-convergent SLAs are usually 'atypical', in the sense that they either have:

- Few households or persons (e.g. commercial centres, industrial areas, sparsely populated rural areas and off-shore islands);
- A high proportion of people living in non-private dwellings (e.g. institutions, prisons, military base); or
- A population with 'unusual' characteristics compared to the HES Survey, such as having large student populations, having many households in caravans, etc.

'Atypical' SLAs have poor convergence results. When these SLAs were identified by the State and Territory partners using their local knowledge and were excluded from the list of non-convergent SLAs, the percentage of convergent SLAs improved. This is especially the case in the ACT where the removal of 15 'atypical' SLAs improves the percentage of convergent SLAs from 79 to 91 per cent. 'Atypical' SLAs in the ACT include the industrial areas of Hume, Mitchell and Fyshwick, and the Defence headquarters at Russell. The estimated residential populations of these SLAs in 2003 were respectively 18, 3, 89, and 0, according to ABS data.

For comparison, using the less stringent cut-off of AV_HH_ABSSUM of '1' or less would increase the number of convergent SLAs to 194 in both NSW and Victoria, 442 in Queensland, and 89 in the ACT.

Convergence measured at the level of the benchmark class

Convergence can also be examined at the more detailed level of the individual benchmark class. Here, the convergence of each of the 116 benchmark classes in an SLA is measured by first calculating the absolute relative residual in percentage (i.e. *benchmark class convergence* = a*bsolute (Census count – estimate)/Census count*100*). Next, the SLA average of this measure in a State or Territory is calculated and is used for reporting. We call this final measure the *Mean benchmark class residual of SLA.*. Table 4 shows the convergence results of two selected benchmarks: Labour Force Status and Individual Weekly Income. The former benchmark has four classes and the latter has ten.

Table 4 shows that for the benchmark Individual Weekly Income, perfect convergence (i.e. zero residuals) has been achieved in all the ten benchmark classes in all States and the Territory. However, for the benchmark Labour Force Status, two benchmarks classes have converged perfectly and two have not – namely "persons not in labour force" and "unemployed". Of these two, the latter benchmark class has converged better than the former, by showing less residuals.

Not shown in Table 4 are the convergence results of the remaining ten benchmarks and their 102 benchmark classes. The only one of these remaining benchmarks that did not converge perfectly is Occupation. However, the residuals for each of the ten classes of this benchmark are all less than about 0.5 percent.

Convergence can be investigated even further by examining the residual of a particular benchmark class within a particular SLA. By doing so, an SLA that did not achieve a good overall convergence can be interrogated to determine what benchmark class has caused it to not converge. This analysis has indeed been used to fine-tune the reweighting process.

Summing up the discussion on convergence, it can be concluded that very good results have been achieved either at the SLA level using the more stringent measure of *AV_HH_ABSSUM* of less than or equal to '0.1', or at the benchmark class level using the *Mean benchmark class convergence of SLA* measure.

Table 4 Mean Benchmark Class Residuals (SLA Average) for Two Selected Benchmarks

Benchmark	Benchmark class	NSW	VIC	QLD	ACT
Labour Force Status	Full-time employed	0.00	0.00	0.00	0.00
	Part-time employed	0.00	0.00	0.00	0.00
	Persons not in labour force	**1.16**	**0.97**	**1.50**	**1.79**
	Unemployed	**0.17**	**0.10**	**0.24**	**0.28**
Individual Weekly Income	<$200	0.00	0.00	0.00	0.00
	$200-$299	0.00	0.00	0.00	0.00
	$300-$399	0.00	0.00	0.00	0.00
	$400-$499	0.00	0.00	0.00	0.00
	$500-$599	0.00	0.00	0.00	0.00
	$600-$699	0.00	0.00	0.00	0.00
	$700-$799	0.00	0.00	0.00	0.00
	$800-$999	0.00	0.00	0.00	0.00
	$1000-$1499	0.00	0.00	0.00	0.00
	$1500 or more	0.00	0.00	0.00	0.00

Note: The residual are calculated by the formula: (absolute(Census count – estimate))/Census count*100. **Source:** NATSEM.

4.3 Validation

Validation measures

Validation is the comparison of the synthetic estimates with external data. Only convergent SLAs (i.e. those with an AV_ABSSUM_HH value of '0.1' or less) were included for validation because, as earlier explained, the weights of non-convergent SLAs are unreliable and no estimates were produced for them. In addition, SLAs which the States and Territory had identified as very atypical and/or having extremely low populations were removed from the validation analysis.

Where possible, the estimates are validated at the SLA level, if the external data are available at this level. Otherwise, the estimates are aggregated to a higher spatial unit (e.g. State) for validation, if the external data are only available at this level of geography. Only SLA level validation results are presented here.

The external data used for validation are mainly administrative data obtained from various government agencies including the ABS, the Australian Taxation Office (ATO), the Department of Family and Community Service (FaCS), the Department of Veterans' Affairs (DVA), and the Department of Employment and Workplace Relations (DEWR).

External data are collected for 2001 (or for as close to 2001 as possible) in order to match the synthetic estimates for 2001. It is important to note that precise matches between external data and the synthetic estimates cannot always be expected for various reasons including differences in data definitions, different time frames and different geography. Also, some of the validation data are themselves estimates. When reviewing the validation results these differences will need to be considered.

Validation results

Table 5 shows how well the estimates for a selected set of variables (for which external data are available at an SLA level) compare with the external data. Negative values in the table show that the estimates are lower than the external data. Given that some external data are of better quality than others, and that a perfect match between estimates and external data is often unachievable due to the factors noted above, a match of between plus or minus 15 percent is considered reasonable. Using this yardstick, estimates of Wage and Salary, Taxation, and Labour Force data compare better than estimates of FaCS payment recipients and DVA pension data. Variations are also observed among the different States and Territory, and among the different items within the same group of estimates.

In interpreting Table 5, readers are reminded that: (1) the synthetic estimates included in the validation are for 2001; (2) estimates were only produced for convergent SLAs; (3) the convergence measure used was AV_HH_ABSSUM of less than or equal to '0.1'; and (4) some SLAs had been excluded on the advice of the research partners because they were 'atypical', even if they had converged.

Validation of estimates of Wage and Salary, and Taxation

These estimates compared very well with the external data of wages and salary (from the ABS), and with the tax data from the ATO. These external data are available for the same reference year and for the same geography as the estimates. Figures 2 and 3 provide two examples to show the close correlation between the estimates and the external data. One dot on the diagram represents one SLA. The fact that the ATO data are for individuals who submitted a tax return, whereas the estimate is based on all individuals who earn a taxable income, may explain the discrepancies between the tax estimates and the ATO data.

Validation of estimates of Centrelink payment recipients

Synthetic estimates of the number of recipients of selected FaCS payments compare less well with the administrative data produced by the ABS using data from Centrelink – Australia's welfare payment agency. The ACT estimates compare better than the other States, and Age Pension estimates compare better than the other pensions across all States and Territory. Figure 4 provides an example to show a reasonably good comparison for ACT Age pension estimates.

The synthetic estimates are consistently lower than the validation data, particularly for Newstart Allowance and Youth Allowance (Table 5). This may be due to differences in data definition. Another possibility is that the HES

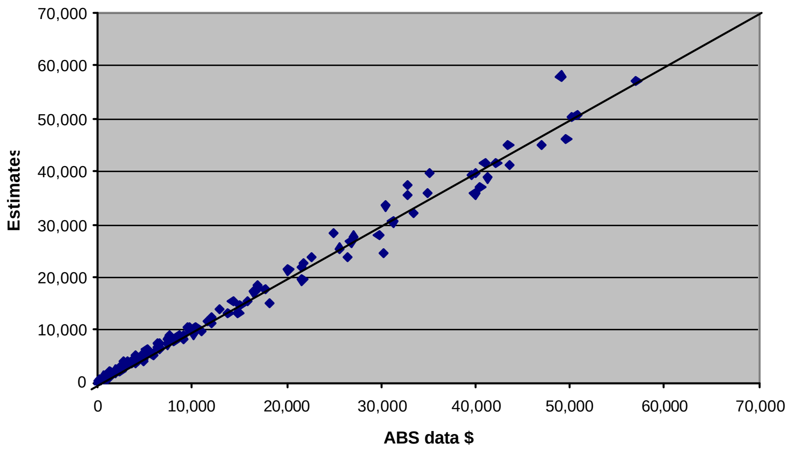
Survey file used in the reweighting does not include people living in non-private dwellings (such as residents of nursing homes and aged care homes) but many of them do receive an Age Pension. Therefore, at least in the case of this payment, the estimates would be expected to underestimate the actual counts.

Table 5. Percentage Difference Between Synthetic Estimates and External Data

Synthetic estimates	NSW	VIC	QLD	ACT
Wage and Salary				
Number of wage and salary earners	-4.6%	-4.2%	-4.8%	0.2%
Total income per week	-4.3%	-3.7%	-0.8%	3.3%
Total income per week from wage and salary	1.0%	2.0%	7.1%	3.6%
Average income per person per week	2.3%	3.1%	5.2%	3.3%
Average income from wage and salary per person per week	11.1%	12.6%	12.2%	8.0%
Taxation				
Taxable individuals	-5.4%	-6.3%	-5.0%	-3.7%
Total taxable income per week	1.0%	2.0%	7.1%	3.6%
Average taxable income per person per week	11.1%	12.6%	12.2%	8.0%
Total tax paid per week	-11.1%	-9.4%	-3.3%	1.0%
Average tax paid per person per week	-3.7%	0.2%	1.6%	4.9%
FaCS payments (number of recipients)				
Age Pension	-8.8%	-11.6%	-11.1%	-8.0%
Disability Support Pension	-13.4%	-16.5%	-13.7%	-12.0%
Newstart Allowance	-14.9%	-20.7%	-21.6%	-5.3%
Parenting Payment Single	-13.3%	-13.1%	-14.9%	8.2%
Youth Allowance	-19.2%	-26.7%	-22.3%	-19.5%
DVA payments				
DVA Disability Pensioners	-9.0%	11.3%	-29.9%	-42.5%
DVA Service Pensioners	-9.5%	2.6%	-27.7%	-8.0%
DVA Disability Pension expenditures	-34.4%	-15.0%	-50.1%	-37.6%
DVA Service Pension expenditures	-22.2%	-11.2%	-36.4%	-4.7%
Labour force				
In labour force	-7.2%	-7.9%	-9.0%	-3.1%
Unemployment	12.2%	-1.3%	-14.5%	-11.2%

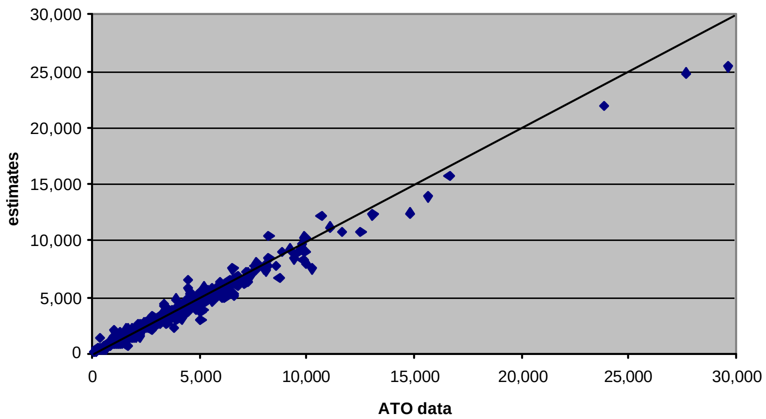
Note: Negative values mean the synthetic estimates are lower than the external data.
Source: NATSEM modelling and external data from ATO TaxStats 2000-01; ABS Regional Wage and Salary Earner Statistics, Australia, Catalogue No. 5673.0; DEWR Small Area Labour Market data, June 2001; ABS Centrelink Income Support Customers - Selected Main Payment Type by SLA, June 2002; and DVA Pensioners by SLA - December 2002.

Other factors that could have contributed to the 'mismatches' include: one, the external data are for June 2002 whereas the estimates are for 2001; and two, some SLAs in the external data have missing values, generally due to low cell counts. Finally, the ABS used a concordance procedure to convert Centrelink postcode-level data to SLA data, which could possibly lead to some geographical differences between the synthetic and the external data.



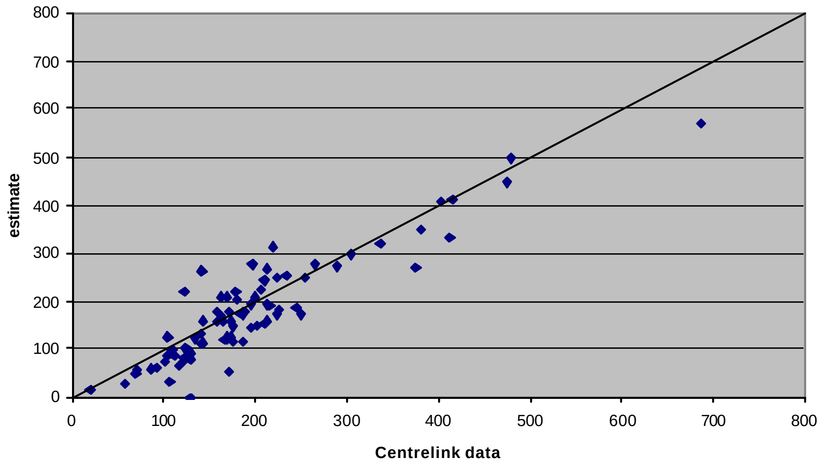
Source: NATSEM Synthetic Estimates and ABS Regional Wage and Salary Earner Statistics, Australia, Catalogue No. 5673.0

Figure 2. Wage and Salary Income, NSW SLAs



Source: NATSEM synthetic estimates and ATO Tax Statistics 2000-0

Figure 3. Number of Taxpayers Paying Positive Tax, Queensland SLAs



Source: NATSEM Synthetic Estimates and unpublished ABS data on Centrelink Income Support Customers – Selected Main Payment Type by SLA, June 2002

Figure 4. Number of Age Pension Recipients, ACT SLAs

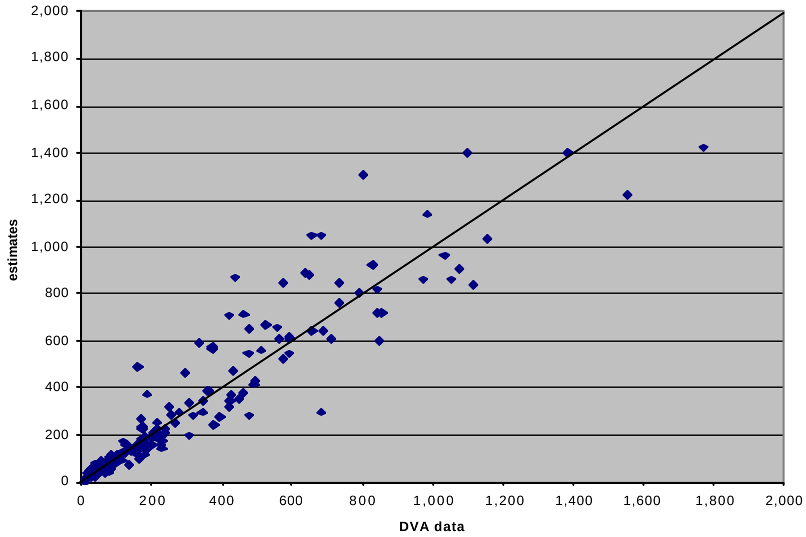
Validation of estimates of DVA pension payments

Estimates of DVA pension recipients and payments received have poor validation results. Here, the 2001 estimates of recipient numbers are validated against DVA data for December 2002; and 2001 payment estimates are validated against DVA's own broad estimates for 2002. These shortcomings in data matching may also explain the considerable variations of the validation results seen among the States and Territory (Table 5). However, it should be noted that relatively few Australians receive DVA pensions, so a large percentage difference may still represent a small number. Figure 5 provides an example to show the considerable divergence between the synthetic estimates and DVA's own estimated data.

Validation of estimates of labour force data

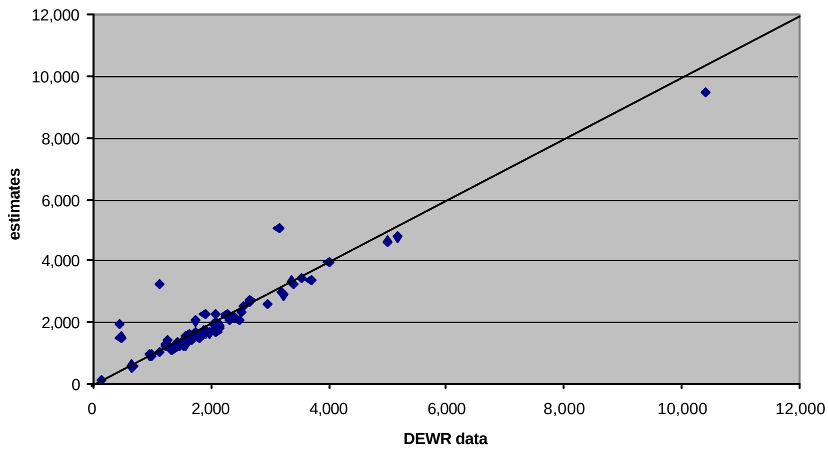
Labour force data are validated against DEWR small-area labour market data for June 2001. Although the reference period for comparison is the same as that for the synthetic estimates, the DEWR data are based on the 1996 SLAs instead of the 2001 SLAs. While it was possible in most cases to adjust for these changes, it was necessary to exclude a small number of the non-matching SLAs. Despite the difference in geography, the validation is reasonably good (Figure 6).

Summing up the discussion on validation, it can be concluded that some synthetic estimates compare better with the external administrative data than others. Also, some comparisons are more valid than others, depending on the quality and comparability of the external data themselves. We would suggest that variations within plus or minus 10 percent are good, and within 15 percent acceptable.



Source: NATSEM synthetic estimates and unpublished DVA Pensioners by SLA data for December 2002

Figure 5. DVA Service Pension Recipients, Victoria SLAs



Source: NATSEM synthetic estimates and DEWR by Small-Area Labour Market Data, June 2001

Figure 6. Number of Persons in Labour Force, ACT SLAs

By this measure, good estimates have been produced for Wage and Salary data, Taxation data and Labor Force data, and acceptable estimates have been produced for many FaCS payments. The reliability of the estimates of the DVA payments data is inconclusive given the very small numbers of recipients in many SLAs, and the availability of only very broad estimates of payment amounts for use in the validation. Overall, we can conclude that the synthetic estimates produced by reweighting are reasonably reliable.

5. SPATIAL IMPACTS OF TAX CUTS

5.1 Background

In the 2005-06 Budget released in May this year, the Australian Government announced measures to further reduce personal income tax, on top of the reduction announced in the 2004-05 Budget. The key changes include reducing the 17 percent tax rate to 15 percent, and raising the tax threshold for the 42 percent tax rate from $58,000 to $63,000, and for the 47 percent tax rate from $70,000 to $95,000 (Table 6). These changes have taken effect from 1 July 2005.

Table 6. Tax Schedules for 2004-05 and 2005-06

2004-05 (before 1 July 2005)		2005-06 (before 1 July 2005)		2005-06 (after 1 July 2005)	
Tax threshold	Tax rate	Tax threshold	Tax rate	Tax threshold	Tax rate
$6,000	0.17	$6,000	0.17	$6,000	0.15
$21,600	0.30	$21,600	0.30	$21,600	0.30
$58,000	0.42	$63,000	0.42	$63,000	0.42
$70,000	0.47	$80,000	0.47	$95,000	0.47

Source: Australian Treasury

In this report, we aim to demonstrate how the spatial microsimulation technique may be used to assess the spatial impact of the announced tax cuts. In particular, we will demonstrate how the 2001 household weights for SLA produced by reweighting may be used to create tax cut estimates for SLAs in 2005-06, and how these estimates may then be used to simulate where the 'winning' and 'losing' regions are.

5.2 Method

Tax estimates for SLAs in 2005-06 were separately generated for the 2004-05 tax regime and the 2005-06 tax regime effective after 1 July 2005. Figure 7 below depicts the overall process, which involves three key steps.

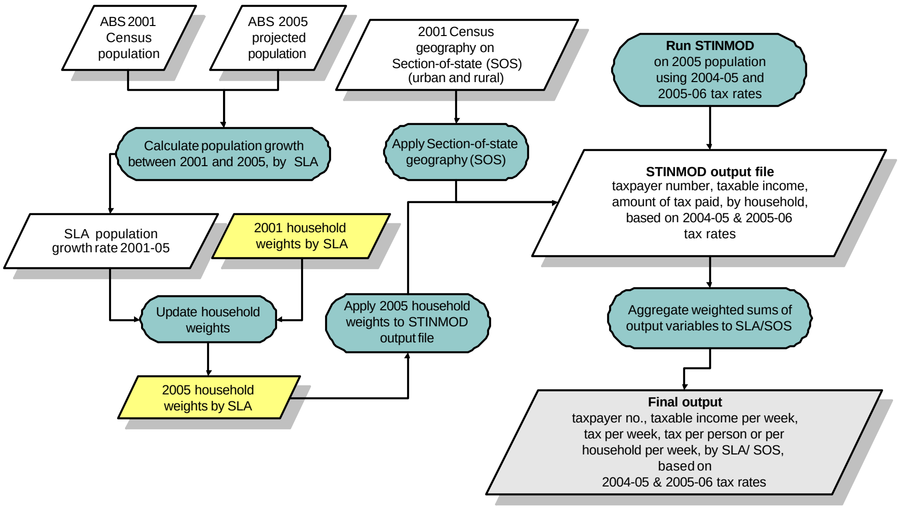


Figure 7. Creation of Small Area Household Tax Estimates for 2005-06 Based on the Tax Rates of 2004-05 and 2005-06

Step 1. Update synthetic household weights for SLAs from 2001 to 2005

The 2001 household weights for each SLA generated by the reweighting process were updated to 2005. This was done by applying the ABS estimates of population growth for each SLA between 2001 and 2005 to the 2001 household weights for each SLA. The assumption here is that population growth is a reasonable proxy for projecting the growth in household numbers.

Step 2. Generate household tax estimates for 2005-06 using STINMOD

The version of the NATSEM STINMOD Model based on the *ABS Household Expenditure Survey 1998-99 (HES)* was run to generate two STINMOD output files containing tax estimates for households in 2005-06. The two files are respectively based on the tax regimes of 2004-05 and 2005-06 (post 1-July 2005) and the tax estimates are:

- Number of taxpayers paying positive tax (and number of households)
- Weekly taxable income
- Weekly amount of tax paid.

Step 3. Create SLA synthetic tax estimates

The 2005 household weights for SLA (from Step 1) were applied to the 2005 STINMOD output file (from Step 2) to produce the SLA total of the weighted sums of the tax variables. Two sets of results were produced – one based on the 2004-05 tax rates and one on the 2005-06 rates. From these results, the following tax cut estimates were derived for each SLA:

- Absolute and per cent tax cut per week per person paying positive tax;
- Absolute and per cent tax cut per week per household.

It should be noted that estimates are only produced for convergent SLAs (for the reasons explained in Section 4.2). The convergent measure used in this case is average-absolute-sum-of-residuals-per-household (AV_HH_ABSSUM) of less than or equal to '1'. A less 'stringent' measure is adopted here so that less SLAs are excluded from the synthetic estimation.

Output Geography

Tax-cut estimates are reported for the convergent SLAs in NSW, Victoria, Queensland and the ACT. Non-convergent SLAs have no estimated data and will show up as blank regions when mapped.

Tax-cut estimates are also reported in broad 'urban' and 'rural' regions. For this purpose, the ABS Census geographical structure – Section of State (SOS), which is part of the 2001 Australian Standard Geographical Classification (ASGC) – was used to generate the following five regions:

- **Capital Cities** – These are the 'Major Urban' areas in the SOS structure with a population of 100,000 to 1 million or more. They include the Statistical Divisions: Sydney, Melbourne, Brisbane and Canberra;
- **Major Urban Areas** – These are the 'Major Urban' areas in the SOS structure with a population of 100,000 to 249,999. They fall outside the above Statistical Divisions;
- **Regional Towns** – These are the 'Other Urban' areas in the SOS structure with a population of 1,000 to 99,999;
- **Rural Towns** – These are the 'Bounded Locality' areas in the SOS structure with a population range of 200 to 999; and
- **Rural Areas** – These are the 'Rural Balance' areas in the SOS structure with population less than 200.

Note that the 2001 Census geography has been used to output 2005-06 tax cut estimates. Although the boundaries of most SLAs remain stable, some changes have occurred between 2001 and 2005. Tracking the spatial changes would require the use of population-based concordance files, which, at the time of writing this report, were not yet available from the ABS. Hence no adjustment could be made and the relatively few changes in SLAs are not considered significant enough to alter the overall results. The SOS, on the other hand, is a Census year spatial structure and will not change until the next population Census in 2006.

5.3 Results

Distribution of tax cuts by taxable income

The magnitude of the 2005-06 tax cuts is determined by the level of taxable income. Before we examine the spatial distribution of tax cuts, it will be useful to gain an insight into how tax cuts are distributed across the taxable income spectrum for a hypothetical single taxpayer (Figure 8).

As the figure shows, in dollar terms, people on higher taxable incomes will benefit more from the tax cuts than those on lower taxable incomes. For example, people on $60,000 annual taxable income will receive a $10.60 per week tax cut and this will increase to $41.60 per week for people on a taxable income of $100,000 or more. On the other hand, people with a taxable income of between $20,001 and $50,000 will only receive a tax cut of $5.40 to $6 a week. This group of taxpayers made up about 54 per cent of all taxpayers in 2002-03 (Table 7).

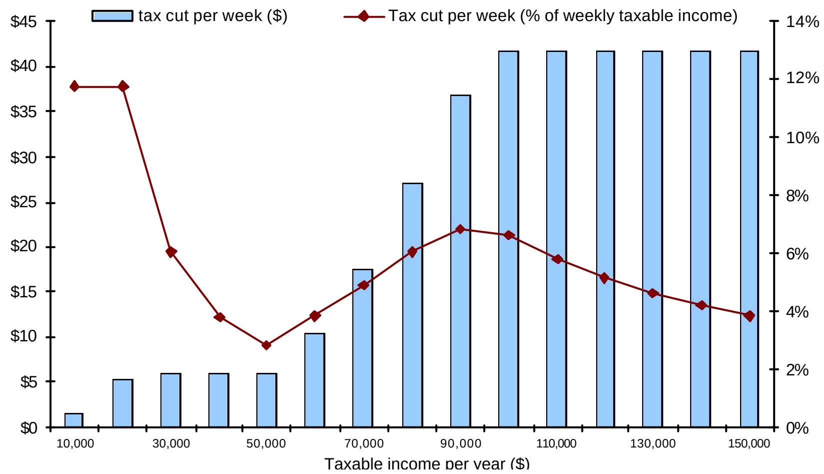


Figure 8. Distribution of the 2005-06 Tax Cuts by Taxable Income for a Single Taxpayer

In *percentage terms* however, people on very low taxable incomes of $20,000 or less will benefit most (11.8 per cent tax cut) – Table 7 shows this group made up 22 per cent of all taxpayers in 2002-03. The next income groups to benefit most from the tax cuts are people on $30,000, and those on $80,000 to $100,000, taxable incomes. These groups will receive a tax cut of about 6 per cent. People receiving a less than 4 per cent tax cut will be those with a taxable income of $40,000 to $60,000, and those on $150,000 or more.

Table 7. Taxpayer Number by Taxable Income, 2002-03

Taxable income pa	Taxpayer number	Per cent of total	Taxable income pa	Taxpayer number	Per cent of total
$10,000 or less	367,503	4%	$60,001-$80,000	678,249	8%
$10,001-$20,000	1,535,506	18%	$80,001-$100,000	240,947	3%
$20,001-$30,000	1,835,095	21%	$100,001-$200,000	240,249	3%
$30,001-$40,000	1,637,874	19%	$200,001or more	64,785	1%
$40,001-$50,000	1,237,339	14%			
$50,001-$60,000	796,699	9%	Total	8,634,246	100%

Note: Taxpayers with net tax payable >$0. **Source:** Australian Taxation Office FY2002-03 (latest published data).

Synthetic estimates of tax cut by urban and rural regions

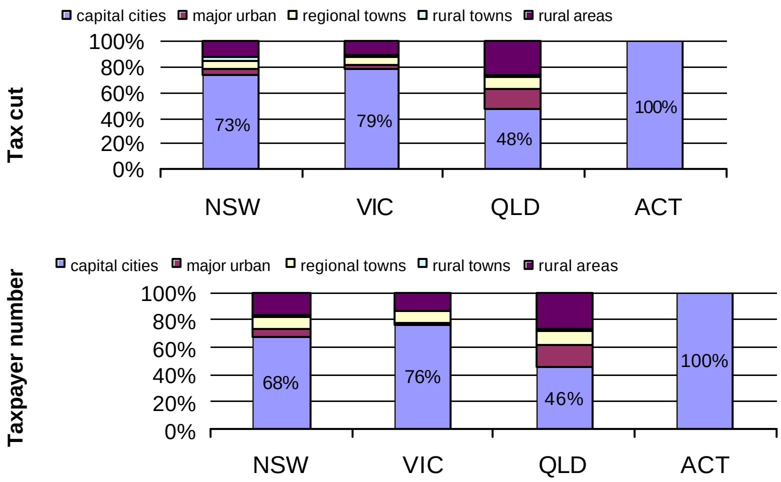
Table 8 shows the State/Territory totals of tax cut per week for all convergent SLAs in the five urban and rural regions. Clearly, the urban regions will benefit more from the tax cuts than the rural regions. This is hardly surprising as we have previously seen in Figure 8 that those on higher taxable incomes will receive higher dollar tax cuts, and these income groups are more common in the urban areas. In addition of course, more of the population live in the urban areas.

Table 8. Synthetic Estimates of Total Tax Cut ($000), by Urban and Rural Regions, 2005-06

State/Territory	Capital city	Major urban	Regional town	Rural town	Rural area	All regions
NSW	21,290	1,379	2,015	716	3,715	29,115
VIC	16,276	407	1,433	129	2,325	20,570
QLD	6,469	2,037	1,402	186	3,515	13,609
ACT	1,850	0	0	0	1	1,851
Total	45,885	3,823	4,850	1,031	9,556	65,145

Note: Convergent SLAs only. **Source:** NATSEM modelling.

In addition, Figure 9 also shows that the distribution of tax cuts among the urban and rural regions in each jurisdiction closely mirrors the distribution of the taxpayer population in these regions. For example, NSW capital city has 68 per cent of taxpayers and 73 percent of the weekly tax cut in that state.



Note: Convergent SLAs only. **Source:** NATSEM modelling

Figure 9. Synthetic Estimates of Tax Cut per Week, and Taxpayer Number, as Percentage of State/Territory Totals, by Urban and Rural Regions, 2005-06

Synthetic estimates of dollar tax cuts by urban and rural regions

Table 9 and 10 respectively show that for all regions in all the four States and Territory, the overall average dollar tax cut is $8.14 a week per taxpayer paying positive tax, and $11.50 a week per household (regardless of whether it contains a taxpaying individual).

Table 9. Synthetic Estimates of Average Dollar Tax Cut per Week per Taxpayer, by Urban and Rural Regions, 2005-06

State/Territory	Capital city	Major urban	Regional town	Rural town	Rural area	All regions
NSW	$9.69 (1.19)	$6.98 (0.86)	$6.80 (0.84)	$6.81 (0.84)	$6.38 (0.78)	$8.61 (1.06)
VIC	$8.60 (1.06)	$7.20 (0.88)	$6.38 (0.78)	$6.40 (0.79)	$6.27 (0.77)	$8.02 (0.99)
QLD	$7.81 (0.96)	$6.76 (0.83)	$6.84 (0.84)	$5.68 (0.70)	$6.89 (0.85)	$7.25 (0.89)
ACT	$10.14 (1.25)	na	na	na	$8.93 (1.10)	$10.14 (1.25)
All four States/Territory	$8.99 (1.11)	$6.89 (0.85)	$6.68 (0.82)	$6.52 (0.80)	$6.53 (0.80)	$8.14 (1.00)

Notes: Individual taxpayers with net tax payable greater than $0. The tax cut figures are for convergent SLAs only. They are also expressed (in bracket) as multiples of the all-states/territory-all-region tax cut figure of $8.14. **Source:** NATSEM modelling.

Table 10. Synthetic Estimates of Average Dollar Tax Cut per Week per Household, by Urban and Rural Regions, 2005-06

State/ Territory	Capital city	Major urban	Regional town	Rural town	Rural area	All regions
NSW	$14.59 (1.27)	$9.13 (0.79)	$8.76 (0.76)	$8.99 (0.78)	$8.03 (0.70)	$12.22 (1.06)
VIC	$12.64 (1.10)	$9.71 (0.84)	$8.26 (0.72)	$8.40 (0.73)	$8.53 (0.74)	$11.48 (1.00)
QLD	$11.01 (0.96)	$9.01 (0.78)	$8.78 (0.76)	$7.10 (0.62)	$9.39 (0.82)	$9.91 (0.86)
ACT	$16.03 (1.39)	na	na	na	$11.67 (1.01)	$16.03 (1.39)
All four States/Territory	$13.30 (1.16)	$9.12 (0.79)	$8.61 (0.75)	$8.51 (0.74)	$8.61 (0.75)	$11.50 (1.00)

Notes: All households, regardless of whether they contain taxpaying individuals. The tax cut figures are for convergent SLAs only. They are also expressed (in bracket) as multiples of the all-states/territory-all-region tax cut figure of $11.50. Source: NATSEM modeling.

Both Tables also show that the capital cities will receive higher average tax cuts than all the other regions. In fact, in the capital cities, tax cut per taxpayer

($8.99) will be 1.11 times that of the overall figure of $8.14, and tax cut per household ($13.30) will be 1.16 times that of the overall figure of $11.50.

Comparing all jurisdictions, the ACT, being relatively affluent and with almost all its SLAs in the capital city region, will unsurprising be the biggest 'winner'. The tax cut per taxpayer it will receive ($10.14) is 1.25 times that of the all-regions-all-States figure of $8.14. Similarly, the tax cut per household it will receive ($16.03) is 1.39 times that of the overall figure of $11.50.

Surprisingly in Queensland, the Rural Areas will receive a slightly higher tax cut per taxpayer than the major urban areas (Table 9), and a higher tax cut per household than the major urban areas (Table 10). This result is driven by a group of SLAs with a relatively high taxable income, in the Mackay-Rockhampton area in the north and in the Mt Isa area in the west.

Synthetic estimates of percentage tax cuts by urban and rural regions

Tables 11 and 12 respectively show the weekly tax cut a taxpayer and a household will receive as a percentage of the weekly amount of tax paid based on the 'old' tax rates of 2004-05. Table 11 shows an overall tax cut of 4.16% per taxpayer and Table 12 shows an overall tax cut of 4.92% per household.

Table 11. Synthetic Estimates of Average Percentage Tax Cut per Week per Taxpayer, by Urban and Rural Regions, 2005-06

State/Territory	Capital city	Major urban	Regional town	Rural town	Rural area	All regions
NSW	4.19	3.98	4.05	4.06	4.10	4.12
VIC	4.24	4.08	4.14	4.04	4.06	4.19
QLD	4.25	4.15	4.25	4.01	4.19	4.21
ACT	4.20	na	na	na	4.85	4.20
All four	4.21	4.08	4.13	4.04	4.12	4.16

Notes: Individual taxpayers with net tax payable greater than $0. Percentage tax cut figures are for convergent SLAs only. **Source:** NATSEM modelling.

Table 12. Synthetic Estimates of Average Percentage Tax Cut per week per Household, by Urban and Rural Regions, 2005-06

State/Territory	Capital city	Major urban	Regional town	Rural town	Rural area	All regions
NSW	4.77	5.02	5.10	5.14	5.30	4.88
VIC	4.90	5.04	5.20	5.26	5.23	4.96
QLD	4.89	5.07	5.07	5.37	5.14	5.00
ACT	4.70	na	na	na	4.97	4.70
All four	4.83	5.04	5.12	5.20	5.22	4.92

Notes: All households, regardless of whether they contain taxpaying individuals. Percentage tax cut figures are for convergent SLAs only. **Source:** NATSEM modelling.

Both Tables also show that the percentage tax cuts in all regions are very similar. Unlike the dollar tax cuts, the capital cities no longer have a clear lead over the other regions in terms of tax cut 'benefit'. On the contrary, Table 12 shows that the average percentage fall in income tax paid (for household) is slightly lower in the capital cities than in the other four regional categories.

Similarly, the percentage cuts in income tax across all States and the Territory are also very similar. The clear lead of the ACT, evident when the 'benefit' was measured as 'tax cut dollars', has now disappeared.

The reason why the percentage tax cut is so similar in all regions and in all States and the Territory becomes self-evident when we once again refer to Figure 8 and Table 7. These show that the 43 percent of taxpayers with the lowest taxable incomes of $30,000 or less receive tax cuts of about 6 to 12 percent, and the other 57 percent receive tax cuts of between 3 and 6 percent. The net effect is the smoothing out of all results and this effect is reflected in the spatial distribution of the SLAs when mapped.

Synthetic estimates of dollar and percentage tax cut by SLAs

A series of maps have been produced to show the spatial distribution of tax cut per household (or per taxpayer) in both dollar and percentage terms, for the convergent SLAs (see Appendix A of Chin et al, 2005, pp. 183-192. These maps are also available on NATSEM's website www.natsem.canberra.edu.au). The key observations from these maps are as follows:

- In dollar terms, there are more SLAs receiving higher tax cuts in the capital cities than in the other regions and this is true for all States and Territory. This pattern is true for both average tax cut per household and average tax cut per taxpayer, although only the former was mapped.
- In dollar terms and within the capital cities, SLAs receiving the highest tax cuts are generally found in 'wealthy' SLAs such as Mosman and Woollahra in Sydney; Port Phillip - West and Bayside - Brighton in Melbourne; Figtree Pocket and Brookfield - Mount Cootha in Brisbane; and Bruce and Campbell in Canberra. Once again, these observations for household tax cuts equally apply to the tax cuts per taxpayer.
- However, a handful of SLAs that are outside the capital cities also receive a relatively high amount of tax cut. In Queensland, these include Broadsound and Peak Downs (mining and sugar) in the Mackay region in the north, and Mount Isa (mining) in the west. In NSW, these include Singleton and Yarrowlumla in the Queanbeyan region next to the ACT.
- In percentage terms, SLAs in the capital cities will generally receive lower tax cuts than their counterparts in the other regions. This is true for all State and Territory and the pattern is generally opposite to that for tax cuts in dollar terms.
- Tax cut estimates have also been mapped in both dollar and percentage terms concurrently. In this case, the estimated tax cuts per taxpayer were mapped to the four groups of SLAs shown in Table 13. The distribution of these four groups of SLAs closely matches the distribution of personal taxable incomes. Mapping results show that the

majority of the SLAs have a personal taxable income ranging ≤$10,000 to $60,000 (i.e. in the 'LH' and 'LL' classes). This observation agrees with the 2002-03 ATO data, which show that these income groups made up approximately 85 per cent of all taxpayers (Table 13).

- Mapping results also show that with a few exceptions, SLAs located outside the capital cities are relatively less 'affluent' (i.e. in the 'LH' and 'LL' classes). The few exceptions being those SLAs with a personal taxable income of between $70,000 and $140,000 (i.e. in the 'HH' class), in northern and western Queensland and in NSW, as we noted before. On the other hand, the more 'affluent' SLAs (i.e. in the 'HH' and 'HL' classes) are mostly found in the capital cities.
- Overall, one might conclude that the group of SLAs (i.e. those in the 'LL' class) with a personal taxable income of $40,000 to $60,000 – which made up 23 per cent of all taxpayers in 2002-03 – are the clear 'losers', as they would receive the least 'benefit' from the tax cuts, in both dollar and percentage terms.

Table 13. Characteristics of Four Classes of SLAs (Maps 11-15,Chin et al, 2005)

SLA class	Weekly tax cut[1] ($)	Tax cut (as % of wkly taxable income)[1]	Personal taxable income[2]	Taxpayer no. as % of total[2]
HH	≥$9.50	≥4.25%	$70,000 - $140,000	-
HL	≥$9.50	<4.25%	$60,000 & >$140,000	-
LH	<$9.50	≥4.25%	≤$10,000 - $40,000	62%
LL	<$9.50	<4.25%	$40,001 - $60,000	23%

Notes: [1]ArcMap Jenk's natural break. [2] Approximation using Figure 8 and Table 7 (2002-03 data). **Source:** NATSEM modelling.

Evaluation of the 2005-06 synthetic tax estimates

Direct comparison of the 2005-06 synthetic tax estimates with independent data is not possible at the small-area level because the latter data are unavailable. However, estimates of tax cuts may be compared against Australian Treasury data for the whole of Australia, and the other tax estimates may be compared against ATO data for 2002-03 and earlier years.

Evaluating tax cut estimates

Two rounds of tax cuts were announced for 2005-06: round one in the 2004-05 Budget and round two in the 2005-06 Budget. Round one would have reduced the 2005-06 tax revenues by $1.875 billion, by moving from the 2004-05 tax rates to the 2005-06 tax rates announced before 1 July 2005. Round two further reduced the tax revenues by another $3.1 billion, by moving from the 'old' 2005-06 tax rates of pre-1 July 2005 to the 'revised' 2005-06 tax rates of post-1 July 2005 (*Australian Government Budget Paper No.2*, 2004-05, 2005-06). These three tax schedules are previously shown in Table 6. Thus, changing from the 2004-05 tax rates to the 2005-06 rates of post-1 July 2005 will cost a total of $4.975 billion in tax revenues in 2005-06 according to the forecasts.

This above tax cut figure from the budget is very close to the $4.905 figure for Australia generated by NATSEM's 'national' STINMOD/04B Model (Table 14), with an absolute difference of only $75 million. Thus, STINMOD/04B figures for the individual States/Territory were generated for comparison with the synthetic estimates produced by reweighting.

Table 14. Synthetic Estimates of Tax Cut by State/Territory, 2005-06

State/Territory	NATSEM spatial synthetic estimate[1] ($m)	NATSEM STINMOD/04B estimate ($m)	Treasury budget forecast ($m)	Difference between NATSEM and STINMOD/04B estimate ($m)	(% of STINMOD)
ACT	*98*	*na*[2]	-	-	-
NSW	1,546	1,743	-	-197	-11%
VIC	1,079	1,258	-	-179	-14%
QLD	710	828	-	-118	-14%
NSW/VIC/QLD	3,335	3,829	-	-494	-13%
Australia		4,905	4,975	-	-

Notes: [1] The estimates have been adjusted upward to include non-convergent SLAs. [2] The ACT cannot be separately identified due to confidentiality measures undertaken by the ABS. **Source:** NATSEM modelling and *Australian Government Budget Paper No.2*, 2004-05 and 2005-06

STINMOD/04B is based on the ABS 2000-01 Survey of Income and Housing Costs (SIHC). It can provide national and State level estimates of the impact of changes in tax and cash transfer policy. However, it cannot provide estimates of the impact of such policies at an SLA level – which is, of course, why we are trying to develop the spatial version described in this paper. A user-friendly version of the 'national' STINMOD Model is publicly available.

Table 14 shows that although the tax cut estimates for NSW, Victoria and Queensland do not match the STINMOD/04B figures exactly, being 11 to 14 percent lower, they are in the same magnitude as the STINOMD figures and are distributed in similar proportions across NSW, Victoria and Queensland. Overall, the spatial synthetic estimates appear credible.

Regarding the difference between STINMOD/04B and the spatial synthetic estimates, it should be noted that STINMOD/04B is constructed on the ABS SIHC Survey, whereas the spatial synthetic estimates are based on reweighting the *ABS 1998-99 Household Expenditure Survey (HES)* to the 2001 Census Extended Community Profile data. Our initial investigation reveals that the number of households in the ABS SIHC is about 7 to 11 percent higher than the household numbers in the 2001 Census. This is a key reason why the aggregated STINMOD/04B tax cut figures are higher than the spatial synthetic estimates.

5.4 Evaluating the Synthetic Estimates of Taxpayers, Taxable Income and Tax Paid

Available tax data for the three most recent financial years were obtained from the ATO. The compounded annual growth rate between the three years was calculated and was used to project the ATO numbers forward to 2005-06. The synthetic estimates were then compared against the projected ATO figures. It is very important to bear in mind that the projected ATO figures should not be treated as the 'true' values because the projection does not reflect the two rounds of tax cuts that have occurred in the last two financial years. Thus, the projected figures are most likely to be over-stated.

Two more points need to be considered when evaluating the synthetic estimates against projected ATO data. One, the synthetic estimates cover all taxpaying individuals whereas the ATO data only includes persons who have lodged a tax return. Two, the synthetic estimates only include convergent SLAs so adjustments are made to account for the non-convergent SLAs, by using the ATO 2002-03 figures (which are for ALL SLAs) to work out what the values of these non-convergent SLAs are, as a percentage of the SLA totals.

Before we examine the 2005-06 synthetic tax estimates, it will be useful to recall how the synthetic estimates of 2001-02 compared with actual (not projected) ATO data for the same year (Table 15). This Table is included here as a reference for our ensuing discussion. In particular, we would repeatedly refer to the 'difference' columns in Table 15 when we discuss the size and direction of the differences between the synthetic estimates and the projected ATO data.

Table 15. Synthetic Tax Estimates and ATO Data by State/Territory, 2001-02

State/ Territory	Taxpayer number ('000)			Wkly taxable income ($m)			Weekly tax paid ($m)[1]		
	NAT-SEM[2]	ATO 2001-02	Differ-ence[3]	NAT-SEM[2]	ATO 2001-02	Differ-ence[3]	NAT-SEM[2]	ATO 2001-02	Differ-ence[3]
ACT	156	166	-6.4%	$139	$141	-1.4%	$34	$36	-5.9%
NSW	2,672	2,790	-4.4%	$2,268	$2,253	0.7%	$508	$575	-13.2%
VIC	1,985	2,129	-7.3%	$1,599	$1,614	-0.9%	$344	$393	-14.2%
QLD	1,443	1,552	-7.5%	$1,093	$1,074	1.7%	$223	$246	-10.3%
All	6,256	6,637	-6.1%	$5,099	$5,082	0.3%	$1,109	$1,250	-12.7%

Notes: [1] Medicare levy (1.5% of taxable income in most cases) are included in the synthetic estimates but excluded from the ATO figures. [2] The estimates have been adjusted upward to include non-convergent SLAs. [3] Negative value indicates that the synthetic estimates are lower than the ATO data. **Source:** NATSEM modelling and ATO TaxStats.

It should also be noted that the results in Table 15 are not directly comparable to the tax results in Table 5 for at least the following reasons: (1) Table 5 has excluded SLAs that did not converge, whereas Table 15 has included ALL SLAs by adjusting the synthetic estimates upward; (2) Table 5 has also excluded a number of 'atypical' SLAs identified by the research partners whereas Table 15 has not; and (3) Table 5 is based on comparison with 2000-01 ATO data whereas Table 15 is based on comparison with 2001-02 ATO data. Despite these differences, Table 5 and Table 15 are broadly consistent in the size and the direction of the variations between the synthetic estimates and the ATO data.

Synthetic estimates of taxpayer numbers

Table 16 shows that the estimated number of taxable individuals paying a positive tax is lower than the projected ATO figures by about 5 percent across all States and Territory. The range of variation is between about 1 to 12 percent. These observations are broadly consistent with the corresponding results in Table 15, in both magnitude and direction of the variations.

Table 16. Synthetic estimates of taxpayer numbers by State/Territory, 2005-06

State/ Territory	NATSEM estimates based on 2005-06 tax rate of post-1July 2005[1] ('000) (A)	ATO 2000-01 ('000)	ATO 2001-02 ('000)	ATO 2002-03 ('000)	ATO 2-yr compounded growth pa	NATSEM-projected ATO 2005-06 ('000) (B)	% Difference[2] (A-B)/A
ACT	165	164	166	171	2.1%	181	-9.7%
NSW	2,877	2,813	2,790	2,852	0.7%	2,910	-1.1%
VIC	2,162	2,114	2,129	2,182	1.6%	2,289	-5.9%
QLD	1,572	1,523	1,552	1,611	2.8%	1,754	-11.6%
All	6,776	6,614	6,637	6,816	1.5%	7,134	-5.4%

Notes: [1]The estimates have been adjusted upward to include the non-convergent SLAs.
[2] Negative value indicates that the synthetic estimates are lower than the ATO data.
Source: NATSEM modelling and ATO TaxStats.

Synthetic estimates of weekly taxable income

Table 17 shows that across all States and Territory, the difference between the synthetic estimate and the projected ATO data is about 0.4 percent. This result is consistent with Table 15 (0.3 percent). However, greater variations can be seen in the individual States and Territory – with a narrower range of variations for the 2001-02 comparisons (Table 15) than the 2005-06 comparisons (Table 17). This discrepancy reflects the 'inflated' nature of the projected ATO figures, which have not accounted for the tax cuts that have taken place in the last two financial years.

Synthetic estimates of weekly tax paid

Table 18 compares the synthetic estimates of tax paid per week (which include the Medicare levy) for all taxpaying individuals paying a positive tax, with the projected ATO data for net tax paid per week (which exclude Medicare levy but include tax rebates) for all individuals who have lodged a tax return. Since the comparison is not like-to-like, greater discrepancies are expected and are indeed shown in Table 18.

Table 18 also shows that for all States and Territory, the projected figures are higher than the synthetic estimates. This is consistent with the 2001-02 comparisons shown in Table 15. However, this time, the differences for the 2005-06 data (about 16 to 31 percent) are a lot higher than the differences for the

2001-02 figures (about 6 to 14 percent). Once again, this would suggest that the projected ATO figures for 2005-06 are over-stated as they fail to reflect the two rounds of recent tax cuts.

Table 17. Synthetic Estimates of Weekly Taxable Income by State/Territory, 2005-06

State/ Territory	NATSEM estimates based on 2005-06 tax rate of post-1July 2005[1] ($m) (A)	ATO 2000-01 ($m)	ATO 2001-02 ($m)	ATO 2002-03 ($m)	ATO 2-yr compounded growth pa	NATSEM-projected ATO 2005-06 ($m) (B)	% Difference[2] (A-B)/A
ACT	$164	$136	$141	$151	5.4%	$176	-7.3%
NSW	$2,738	$2,252	$2,253	$2,394	3.1%	$2,623	4.2%
VIC	$1,948	$1,571	$1,614	$1,712	4.4%	$1,949	-0.0%
QLD	$1,339	$1,024	$1,074	$1,166	6.7%	$1,417	-5.8%
All	$6,189	$4,983	$5,082	$5,423	4.3%	$6,165	0.4%

Notes: [1] The estimates have been adjusted upward to include the non-convergent SLAs. [2] Negative value indicates that the synthetic estimates are lower than the ATO data. **Source:** NATSEM modelling and ATO TaxStats.

Table 18. Synthetic Estimates of Tax Paid per Annum by State and Territory, 2005-06

State/ Territory	NATSEM estimates based on 2005-06 tax rate of post-1July 2005[1] ($m) (A)	ATO 2000-01 ($m)	ATO 2001-02 ($m)	ATO 2002-03 ($m)	ATO 2-yr compounded growth pa	NATSEM-projected ATO 2005-06 ($m) (B)	% Difference[2] (A-B)/A
ACT	$2,087	$1,788	$1,873	$2,052	7.1%	$2,521	-20.8%
NSW	$31,715	$29,956	$29,914	$32,469	4.1%	$36,640	-15.5%
VIC	$21,750	$19,814	$20,437	$22,205	5.9%	$26,343	-21.1%
QLD	$14,205	$12,026	$12,785	$14,297	9.0%	$18,533	-30.5%
All	$69,757	$63,584	$65,009	$71,023	5.7%	$84,037	-20.5%

Notes: [1.] Medicare levy has been included in the NATSEM estimates but excluded from the ATO figures. NATSEM estimates have been adjusted upward to include the non-convergent SLAs. [2] Negative value indicates that the synthetic estimates are lower than the ATO data. **Source:** NATSEM modelling and ATO TaxStats.

In summary, it is obvious that there are limitations to the extent to which the synthetic estimates can be meaningfully evaluated. Given all the data limitations and all the caveats mentioned, the synthetic tax estimates would appear to be reasonably reliable, both in terms of the magnitude of the numbers, and their proportions in the States and Territory. Ultimately, the acid test is to compare the synthetic estimates with the actual, small-area ATO data when they finally become available.

6. CONCLUSION

A spatial microsimlation technique using the reweighting approach has been successfully applied to NATSEM's STINMOD Model to produce estimates of income, tax and social security benefits for SLAs. This 'regionalisation' of STINMOD has extended the usefulness of the Model because, for the first time, the output data from the Model can now be given a new, spatial dimension.

The reweighting process produced a set of small-area household weights for 2001. These weights were used to produce the 2001-02 synthetic estimates of income, tax and social security benefits for SLAs.

The reliability of the reweighting process is measured by 'convergence', which measures how well the estimate of an SLA matches the Census count for the same SLA, for the linkage (or constrained) variables. Good convergence has been achieved for over ninety per cent of the SLAs in the three States and the ACT. Convergence is measured both at the SLA level and at the level of the individual benchmark classes.

The reliability of the small-area synthetic estimates has been measured against external administrative data. The estimates of wages and salary, taxation, and labour force variables have a fairly close match with the external data. The matches are not as close for the estimates of recipients of FaCS and DVA payments, although some of these discrepancies may be due to comparability problems between the administrative data and the estimates, and the fact that some of the administrative data itself involves estimation. Overall however, we can conclude that reweighting produces reasonable and useable small-area synthetic estimates.

To demonstrate the application of small-area estimates for spatial microsimulation, 2005-06 tax estimates for SLAs were synthesized using the 2005 household weights for each SLA, which were themselves updated from the 2001 synthetic household weights for SLA. Using these 2005-06 synthetic tax estimates for SLAs, the spatial effects of the tax cuts announced in the recent Australian Government Budget was simulated.

This study shows that the estimated average weekly tax cut in NSW, Victoria, Queensland and the ACT is about $8.14 a week per taxpaying individuals, or about $11.50 per household. The capital cities will receive more tax cut than the other regions in dollar terms, but in percentage terms they will receive slightly less. This pattern is repeated by the SLAs when their spatial distribution for tax cuts is mapped. Finally, mapping results show that the distribution of the tax cut 'benefit' among the SLAs depends on their taxable incomes. SLAs with a relatively high taxable income will receive a higher dollar tax cut but less in percentage terms, and these SLAs tend to be located in the capital cities.

Evaluating the tax estimates against external data from the Australian Treasury and the ATO proved problematic. Direct comparison is impossible because of the lack of directly comparable data. Comparisons are made with projected ATO figures and they show that for taxpayer number and taxable income, the variations between the estimates and the projections are within a narrow band of about plus or minus ten per cent. The discrepancies for 'tax paid' are greater. A plausible explanation for this is that our projections of ATO tax revenues for 2005-06 are too high because they do not reflect the two rounds of tax cuts that have now taken place. Given the data limitations and all the caveats, the synthetic estimates of tax cuts are considered sufficiently credible to be used for policy analysis, especially in the absence of such small-area data from any other sources.

ACKNOWLEDGMENTS

This Regional Dimensions Linkage Project is being funded by NATSEM, the Australian Research Council (Project No LP0349152), the NSW Premier's Department, the Queensland Department of Premier and Cabinet, the Queensland Treasury, the ACT Chief Minister's Department, the Victorian Department of Sustainability and Environment, and the Australian Bureau of Statistics. NATSEM would like to thank the research partners for their enthusiasm and their funding support to the project.

The authors would also like to thank the following public institutions for their assistance in providing administrative data for use in the validation of the synthetic estimates: the Australian Bureau of Statistics, the Department of Employment and Workplace Relations, and the Department of Veterans Affairs.

Finally, the authors would like to acknowledge the on-going technical support provided by Stephen Leicester of NATSEM and from our chief investigator – Dr Paul Williamson from the University of Liverpool (UK). Many past NATSEM staff had made invaluable contributions to the Regional Dimensions Project. We would particularly like to thank Anthony King, Tony Melhuish, Susan Day, Marcus Blake, Elizabeth Taylor and Anthea Bill, among others.

REFERENCES

Australian Bureau of Statistics (2003) *2001 Census Expanded Community Profile, Australia, Confidentialised Unit Record File (CURF),* ABS Catalogue no. 2005.0.

Australian Bureau of Statistics (2002) *Household Expenditure Survey, Australia, Confidentialised Unit Record File (CURF), Technical Paper, Second Edition (including Fiscal Incidence Study)*, ABS Catalogue no. 6544.0.30.001.

Australian Bureau of Statistics (2001) *Statistical Geograhpy Volume 1, Australian Standard Geographical Classification (ASGC) 2001*, ABS Catalogue no. 1216.0.

Australian Bureau of Statistics (2000) *GREGWT and Table Macro – Users Guide*.

Ballas, D., Clarke, G.P. and Turton, I. (2003) 'A Spatial Microsimulation Model for Social Policy Evaluation.' In Boots, B. and Thomas, R. (eds.) *Modelling Geographical Systems*. Kluwer: Amsterdam, pp. 143-168.

Bell, P. (2000) *Weighting and Standard Error Estimation for ABS Household Surveys, paper prepared for ABS Methodology Advisory Committee*, July 2000, Australian Bureau of Statistics.

Chin, S-F., Harding, A. Lloyd, R., McNamara, J., Phillips, B. and Vu, Q.N. (2005) Spatial Microsimulation Using Synthetic Small-Area Estimates of Income, Tax and Social Security Benefits. In P. Dalziel (Ed.) *Proceedings of the 29th Annual Conference of the Australia and New Zealand Regional Science Association International, Manukau City 27-30 September 2005*. CD-ROM published by AERU, Lincoln University, pp. 156-193.

Day, S., Bill, A., Farbotko, C., Harding, A., Leicester, S., King, A., Melhuish, A. and Taylor, E. (2003) *ARC Regional Dimensions Research Project: Work in Progress.* Presentation at the 27th Australia & New Zealand Regional Science Association International Annual Conference, Fremantle, Western Australia 28th September - 1st October 2003.

Taylor, E., Harding, A., Lloyd, R. and Blake, M. (2004) *Housing Unaffordability at the Statistical Local Area Level: New Estimates Using Spatial Microsimulation.* Presentation at the 28th Australia & New Zealand Regional Science Association International Annual Conference, Wollongong, NSW, Australia, September 2004.

Tranmer, M., Pickles, A., Fieldhouse, E.,. Elliot, M., Dale, A., Brown, M., Martin, D., Steel, D. and Gardiner, C. (2005) The Case for Small Area Microdata. *Journal of the Royal Statistical Society: Series A*, 168(1), p. 29.

Williamson, P. (2001) 'A comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata.' Working Paper 2001/2, Population Microdata Unit, Department of Geography, University of Liverpool.

Williamson, P. (2001) *Spatial Microsimulation: Adding Geography to Microdata.* Seminar presentation to SAGE/Dept. of Social Security, London School of Economics, October 2001.