1352.0.55.065

Australian Bureau of Statistics

Research Paper

Methodology for Producing Synthetic Microdata for Income in Non-Survey Years

www.abs.gov.au

New Issue

Research Paper

Methodology for Producing Synthetic Microdata for Income in Non-Survey Years

Anil Kumar

Analytical Services Branch

Methodology Advisory Committee

26 November 2004, Canberra

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) THURS 23 MAR 2006

ABS Catalogue no. 1352.0.55.065
ISBN 0 642 48169 5

© Commonwealth of Australia 2005

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from the Commonwealth. Requests and inquiries concerning reproduction and rights in this publication should be addressed to The Manager, Intermediary Management, Australian Bureau of Statistics, Locked Bag 10, Belconnen ACT 2616, by telephone (02) 6252 6998, fax (02) 6252 7102, or email <intermediary.management@abs.gov.au>.

Views expressed in this paper are those of the author(s), and do not necessarily represent those of the Australian Bureau of Statistics. Where quoted, they should be attributed clearly to the author(s).

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Mr Anil Kumar, Analytical Services Branch on Canberra (02) 6252 5344 or email <anil.kumar@abs.gov.au>.

METHODOLOGY FOR PRODUCING SYNTHETIC MICRODATA FOR INCOME IN NON-SURVEY YEARS

Anil Kumar
Analytical Services Branch

EXECUTIVE SUMMARY

Beginning with the 2003–04 reference year, the Survey of Income and Housing (SIH) became a biennial rather than an annual survey. The ABS was asked to explore whether income micro data could still be made available for non-survey years. In response the ABS investigated methodological possibilities for producing synthetic micro data for non-survey years.

This paper, presented to the Methodology Advisory Committee in November 2004, outlines the early stages of this investigation. It describes the static ageing technique and discusses how it could be used to produce forecasts of income micro data for non-survey years. Preliminary results from the exploratory study undertaken up to that point are also briefly discussed.

Results from the subsequent exploratory study indicated that the proposed methodology produced reasonably good estimates at aggregated levels of income. However, it was less suitable for producing acceptable estimates of income at the unit record level as would be required for research purposes. Synthetic income data produced through the proposed methodology is unlikely to capture the impact of policy changes on income distribution as accurately or adequately as would be the case if an actual survey was carried out. Consequently we have decided not to take forward the research into production.

QUESTIONS FOR MAC MEMBERS

Priorities are 6, 1, 3 and 7.

1. Have we identified the appropriate methodology to address the task at hand? (Section 2)
2. Have we correctly identified the capabilities of the chosen technique? (Section 2)
3. Are there other approaches/techniques that we should consider or be aware of for producing microdata in non-survey years?
4. Are there other datasets we should be aware of that could be used for demographic and economic ageing? (Table 4.1)
5. Are there other methods which we could use to derive uprating factors for income, especially from self-employment and investment? (Section 4.2)
6. What criteria should we use to decide the feasibility or otherwise of the proposed methodology at the exploratory stage? (Section 5)
7. If implemented, how should we validate/evaluate whether the synthetic micro dataset obtained is reasonably accurate? (Section 6)

CONTENTS

The role of the Methodology Advisory Committee (MAC) is to review and direct research into the collection, estimation, dissemination and analytical methodologies associated with ABS statistics. Papers presented to the MAC are often in the early stages of development, and therefore do not represent the considered views of the Australian Bureau of Statistics or the members of the Committee. Readers interested in the subsequent development of a research topic are encouraged to contact either the author or the Australian Bureau of Statistics.

METHODOLOGY FOR PRODUCING SYNTHETIC MICRODATA FOR INCOME IN NON-SURVEY YEARS

Anil Kumar
Analytical Services

1. INTRODUCTION

The ABS Survey of Income and Housing (SIH) provides information on sources and amounts of income as well as selected housing costs for persons resident in private dwellings throughout Australia. It produces estimates of current weekly income and estimates of annual income for the previous financial year. Data from the survey are used to compile information on individual and household income levels and its distribution.

Data from the survey are important for researchers and policy makers. Estimates of individual, family or household incomes are particularly important for informed and evidence-based policy making. Researchers from academia are particularly interested in SIH unit record income data to model relationships or distributions following changes in policies. Government policy departments, social and economic researchers and academics are interested in annual income data in order to study changes in income levels and its distribution over time. Income data are also used for taxation policy, the planning of social security income support programs and for labour market analysis.

The Survey of Income and Housing (SIH) – formerly the Survey of Income and Housing Costs (SIHC) – was conducted by the ABS on an annual basis from 1994–95 to 1997–98. In 1998–99, the Household Expenditure Survey (HES) was conducted in place of the SIH. Subsequent surveys took place in 1999–2000, 2000–01 and 2002–03, with no survey being conducted for 2001–02. A combined SIH/HES was conducted in 2003–04. After this, the ABS plans to conduct SIH every two years instead of annually. This decision by the ABS has prompted questions from policy departments and academics whether income microdata can still be made available for non-survey years.

In response, the Analytical Services Branch of the ABS is currently involved in a project investigating possibilities for producing synthetic microdata for non-survey years. This project is divided into two phases. Phase 1 is a 'proof-of-concept' phase exploring and testing possible methods to produce synthetic microdata for income using past data. If Phase 1 proves feasible, with a sound methodology being identified and peer-reviewed, then Phase 2 of the project will proceed to implement the chosen methodology in non-survey years to produce synthetic microdata for income.

This paper discusses a methodology for producing synthetic income microdata. It provides a brief description of the proposed methodology, discusses the main processes/steps and benchmark data required to apply this method, and identifies some possible ways to validate the results produced using this method. It also presents some preliminary results from an exploratory study underway and identifies areas for further research, prior to moving into Phase 2.

2. PROPOSED METHODOLOGY

While no generally agreed or readily available methods exist for forecasting income at the microdata level, the ageing techniques of microsimulation models (MSMs) offer the potential to update base year data to future years. MSMs can be divided into two main types: static and dynamic. In principle, the main difference between a static and a dynamic microsimulation model is the ageing procedure (Merz, 1994). Ageing is the process of updating a database to represent current conditions, or of projecting a database for one or more years to represent expected future conditions before simulating and analysing the impact of particular policy changes.

Static microsimulation models typically use a combination of reweighting of microdata and indexation of monetary amounts to age the original cross-section microdata to the required point in time. Dynamic models allow for ageing of the original microdata on the basis of survival or transition probabilities of different real life events (such as marriage, birth, dying, unemployment) occurring. Associated incomes, at each stage, are updated based on current status and circumstances, and past history.

In static ageing, the relationships among the variables and the structure of the sample itself (e.g. age, income type distributions) is not modified and the number of individual units before and after ageing remains same. In dynamic ageing, the demographic characteristics of the individual units are altered and any side effects on the behaviour of further units may be directly accounted for (e.g. if a child is born, this might immediately affect the mother's labour force participation in a simulation period). Since under dynamic ageing the transition probabilities generate new individual units and therefore new populations each year, the number of units after ageing will not necessarily be the same as the number of units before ageing, as is the case under static ageing.

The choice of whether to use a static or dynamic microsimulation model depends, in principle, on the task or question to be addressed and also on the quality and suitability of available data[1] (Mitton, *et al.*, 2000). Static microsimulation models are generally used to measure the instantaneous or the 'first round' effects of policy changes, before individuals have had time to adjust their behaviour to these changes. A static microsimulation model is relatively well suited for short to medium term (one to three years) forecasts on the assumption that the characteristics of the population under examination do not change rapidly. Since time-consuming simulation of demographic processes, with interactions among different individual units (such as marriages) do not need to be estimated, static models are less expensive and the data and modelling/computational requirements of this method are not onerous. Static

1 In practice it also depends on the institutional context and the speed with which an answer to the question is necessary.

models, however, may be less suitable if the demographic and economic structure of the population changes rather quickly even in the short run as such rapid changes may not be properly handled by static ageing (Devine and Wertheimer, 1981).

Dynamic microsimulation models are more suited for long-term projections and where the demographic structure and the levels of economic variables essentially change. In addition, dynamic models also allow for incorporation of behavioural responses to government policy (Devine and Wertheimer, 1981). Dynamic microsimulation models, however, are not easy to implement. They require more data and more processing than static microsimulation models. Such models also require relationships between variables to be established, and interactions between individual units to be considered and at least attempted to be captured.

Our aim is to produce forecasts of income at the unit record level for one year ahead. Given that most population characteristics and economic variables are likely to change only slowly over such short periods, the static ageing technique is deemed more appropriate for our purposes. Furthermore, this technique is easier to implement as the data, modelling and time required to generate new micro datasets appear less onerous on an ongoing basis.

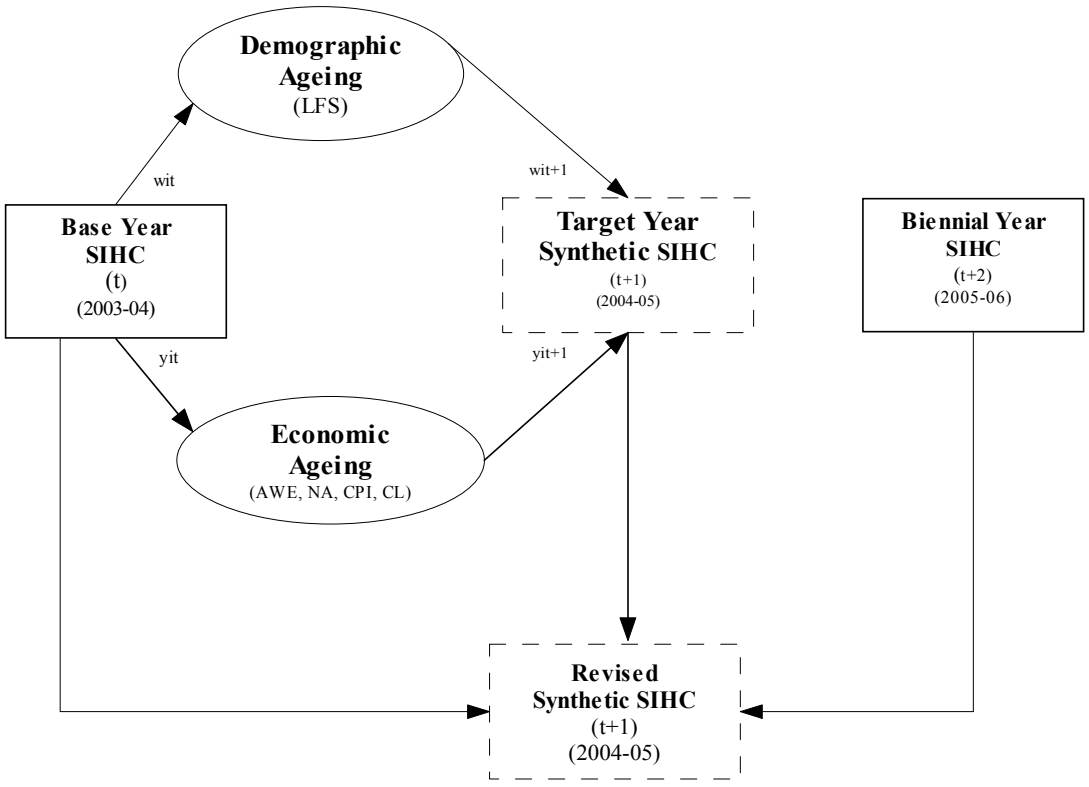
3. PROCEDURE FOR STATIC AGEING

As stated earlier, static ageing consists of two components – demographic ageing or reweighting, and economic ageing or uprating. Demographic ageing involves calculating new sampling weights for each unit record by adjusting the original unit weights in line with movements over time in the characteristics of the populations that the record represents. The reweighting process in essence tries to capture the change in demographic, family, education, and labour force characteristics (e.g. the increase in unemployment) since the last survey period (Lambert, *et al.*, 1994).

Economic ageing involves uprating (either inflating or deflating) components of income of each unit record to reflect changes in income levels that have occurred since the last survey was undertaken. Each income component can be aged separately by applying appropriate uprating factors that more closely reflect movements in that component of income in the economy over time.

Figure 3.1 below presents an overview of the framework for static ageing to be applied in our case. The microdata containing the associated population weights (w_i) and income (y_i) for each unit record (i) from the last survey period (t) e.g. 2003–04 (the base year) will be extrapolated to create a new synthetic microdata file for the non-survey period ($t+1$) e.g. 2004–05 (the target year) by applying the appropriate population reweighting and income uprating factors.

3.1 Framework for Static Ageing



A revised synthetic microdata file could also be produced using interpolation between the previous and following years' survey data when the next survey data becomes available. This refinement could exploit the 'recall' data on the previous year's income.[2]

2 For example when the 2005–06 SIH data (including recall data on 2004–05 income) become available, we could interpolate the data between 2003–04 and 2005–06 to produce a revised synthetic unit record file for 2004–05.

4. BENCHMARK DATA FOR STATIC AGEING

Benchmark data are required to undertake both demographic and economic ageing. Table 4.1 below presents possible data sources that could be used for ageing the demographic and economic variables, each of which is discussed in more detail in this section.

4.1 Possible data sources for ageing demographic and economic variables

Variables	*Data source for updating*
Demographic	
Persons (State by Sex by LF Status by Age)	Labour Force Survey (LFS), Estimated Resident Population (population benchmark by household composition, number of children, etc.)
Economic	
Income	
Wages and Salaries	Average Weekly Earnings (AWE), Employee Earnings, Benefits and Trade Union Membership (EEBTUM) (Sex by F/T/P/T, FT/PT by quintiles)
Own Unincorporated Business Income	National Accounts, LFS (by Farm/Non-Farm)
Investment Income	
– Interest	National Accounts
– Dividends	National Accounts
– Rent	National Accounts
– Other investments	National Accounts/CPI
Govt Cash payments	
– Pensions	Official rates or Imputation (Centrelink/FaCS)
– Allowances	Official rates or Imputation (Centrelink/FaCS)
– Family Tax Benefits	Official rates or Imputation (Centrelink/FaCS)
– Others	Official rates or Imputation (Centrelink/FaCS)
Private cash transfers	
– Superannuation	CPI
– Workers compensation	CPI
– Child support/maintenance	CPI
– Other sources	CPI
Residual Income	CPI

4.1 Demographic ageing

Benchmark data for demographic ageing can be obtained from labour force surveys or estimated resident population. The ABS Labour Force Survey (LFS) provides estimates on a monthly basis of labour force status, unemployment and participation rates by a range of socioeconomic indicators such as state of residence, sex, age, occupation, family status and education status. This allows us to use benchmarks for ageing at a finer level of disaggregation than is possible from estimated resident population benchmarks.

Demographic ageing uses the benchmark data to adjust the original sampling weights. Since not all SIH unit records find a corresponding match in the LFS data (and vice versa), the SIH microdata are normally divided into subgroups or cells based on a combination of socioeconomic groupings. The initial SIH sub-groupings could consist of State by Sex by Labour Force Status by Age cross-classification. Further modifications to this initial matrix could be undertaken to maximise matching of unit records from the two datasets.

Once the sub-groupings or cells for the SIH are produced, a ratio reweighting approach is used to assign new sampling weights to the initial SIH base file using benchmark data from the LFS corresponding to these cells. The reweighting factor for a given cell (and consequently each individual unit in that cell) is calculated by dividing the number of people in the corresponding cell from the LFS in the target year by the number of people in the corresponding cell in the base year. The new weights can be expressed as follows.

$$\text{New weight for person } i \text{ in cell } j \ (w_{ijt+1}) = \text{Old weight for person } i \text{ in cell } j \ (w_{ijt}) \times \frac{\text{No. of LFS persons in cell } j \ (LFS_{jt+1})}{\text{No. of LFS persons in cell } j \ (LFS_{jt})}$$

To illustrate, if there were 200 persons with a particular set of characteristics in the base year SIH (say, males aged between 45 and 54 years employed full-time and living in NSW), and the LFS showed the number of persons with the same characteristics rose from 300 in the base year to 600 in the target year, then the sampling weight of each of the 200 original records would be multiplied by 2.

4.2 Economic ageing

Benchmark data for uprating income (both current weekly income and previous financial year income) are generally derived from a variety of sources. The formula for calculating uprating factors for income can be expressed as follows.

$$\text{Uprating factor} = \frac{\text{Benchmark Income for Target year}}{\text{Benchmark Income for Base year}}$$

As highlighted in table 4.1, the sources for income benchmark variables are as follows:

i. Income from wages and salaries are generally uprated using average weekly earnings data. Changes in earnings by sex, full-time/part-time status or by income quintile could be used to derive the uprating factors.

ii. Uprating factors for income from own unincorporated business (i.e. self-employment) and investment (interests, dividends, rent etc.) could be derived from data from the National Accounts.[3]

iii. Income from social security payments could be uprated using changes in the official rates of payments or imputed on the basis of eligibility criteria and income tests (Bremner, *et al.*, 2002).

iv. Income from private cash transfers (e.g. superannuation, workers compensation, child support etc.) could be uprated using the CPI.

It may also be possible that instead of uprating income by applying average growth factors in macroeconomic variables, alternative forecasts of major components of income (e.g. wages and salaries, business income and property income) could be derived using regression techniques.[4]

3 The National Accounts provides data at the aggregate level on household income from farm and non-farm unincorporated enterprises, interest , dividends, rent, etc. which could be used to derive uprating factors for each of these components.

4 For example wages and salaries could be estimated using a regression equation where wages and salaries are regressed on age, sex, education and industry using pooled cross-section data from past SIH data.

5. EXPLORATORY STUDY

An exploratory study is being undertaken to test the usefulness and accuracy of the static ageing technique in producing synthetic income microdata using past SIH data. Presently consecutive year SIH data exist for 1994–95 to 1997–98 and 1999–2000 to 2000–01.[5] Using data for these years, synthetic income microdata for one year ahead will be generated and compared with the actual microdata to assess the accuracy of the method. For example using 1996–97 as base year, income for 1997–98 could be forecast and compared with the actual survey data for 1997–98. Similarly, using 1999–00 as base year, income for 2000–01 could be forecast and compared with the actual survey data for 2000–01.

To assess how good the microdata forecasts are, we will need to specify some tolerance levels. We will consider the synthetic estimates derived from the microdata forecasts to be reasonable if they are within 5% of the actual estimates at aggregated and disaggregated level. We will also examine where the synthetic estimates are positioned in the spectrum between the actuals of the base year (aged by CPI, say) and the actuals of the year under investigation. Furthermore we will examine the effectiveness of the model in estimating the data in the main tables of the SIH publication *Household Income and Income Distribution* (6523.0) such as Gini coefficients, distribution of income by quintiles, etc. We will do this assessment at various levels of aggregation. We also need to determine the interaction between sampling error and our assessment criteria.

5 For 1998–99 HES was conducted in place of SIH while no SIH was conducted in 2001–02. The 2002–03 SIH data is not yet released.

6. MODEL VALIDATION DURING PHASE 2

A major component of any modelling exercise should be an evaluation or validation of the model results to see how they measure against reality. One of the major uses of the SIH data is to calculate various measures of income and its distribution, and a synthetic SIH file will need to do this accurately. If implemented, no actual data in the non-survey years will be available to assess the accuracy of the methodology or forecasts, so this will not be an easy task. Hence one of the major challenges for this project would be how we define and assess the 'success' of it during Phase 2 (i.e. when in production).

During Phase 1, comparing forecasts of income using past data with the actual survey microdata would provide us with some indication of how much confidence we can place on the estimates derived from the adopted method. Furthermore, comparing the synthetic microdata with the results from interpolation after the next SIH data are available, will also give us some indication of the reliability and accuracy of the forecasts. However, these options will not be available to us when we start producing estimates for the non-survey years. As such, alternatives means of model validation during Phase 2 will be necessary.

Some possible methods of validation are briefly discussed below. These methods should not be seen as mutually exclusive but complementary to each other.

6.1 Internal validation

A 'debugging' exercise could be undertaken to verify the accuracy of the computer codes, particularly for income from social security payments, to see whether they accurately reflect program rules and therefore produce correct outcomes. This approach could be complemented by the standard programming practice of code walk-throughs. The verification of results by FaCS is another way of ensuring that the model produces valid program outcomes (Bremner, *et al.*, 2002).

6.2 External validation

External validation involves benchmarking the results from the method against "the truth" i.e. against data from administrative records or other sources that are considered to represent a standard for comparison. Compositional changes could be examined to see whether the results are consistent with actual changes. In addition to checking the aggregates (numbers, outlays) the program profiles (e.g. age, family status) could also be undertaken. The validation process should examine the distribution of the unit records and the marginal totals in addition to aggregates.

6.3 Sensitivity analysis

This technique could be used to examine the effect on the results of alternative choices about the key assumptions of the ageing method. For example we could look at the sensitivity of the income projections if we vary unemployment rates, participation rates and wage levels. The extent of variation in the results helps gauge the technique's susceptibility to bias in the results based on different scenarios. We could also look at the sensitivity of the income forecasts if we use alternative benchmark data, or alternative subgroupings for population reweighting.

6.4 Variance estimation

The jack-knife technique can be used to estimate the variance or relative standard errors (RSEs) of the synthetic microdata for various socioeconomic and income variables. The jack-knife variance of an estimator is the average squared distance between estimates derived from repeated sub-sampling (with replacement) of the original sample and the estimate derived from the (whole) original sample. The estimates from the whole sample can be considered as a proxy for the 'true' estimate. Higher variances or RSEs for a particular variable imply less precise estimates so caution should be exercised in using these results. A comparison of the RSEs from the synthetic micro dataset against RSEs from actual micro datasets, even from previous years, would help gauge whether extra variation has been introduced by the ageing methodology.

6.5 Quality statements

Finally, some quality statements could also be made in terms of what the synthetic data is good for and where/what it can be used for and where it cannot be used.

7. PRELIMINARY RESULTS FROM EXPLORATORY STUDY

In this section, we present some preliminary results using past SIH data to test the accuracy of the static ageing technique. Using SIH 1999–00 as the base year, we produce synthetic forecasts of the population weights and income, for each unit record for 2000–01. We then obtain aggregates or estimates from the synthetic microdata, and compare those estimates with actual estimates from SIH 2000–01. At this stage, results from the exploratory study are reported for two variables of interest – population size and estimates of current weekly income.

7.1 Synthetic estimates of Population

For population or demographic ageing, we first divided the SIH 1999–00 observations (and their corresponding population weights) into subgroups (cells) using the following cross-classification: state (8) by sex (2) by labour force status (4)[6] by age group (7).[7] For each of these cells[8] we then obtained corresponding reweighting factors using benchmark data from the LFS.[9] The number of persons for each cell in the LFS matrix for 1999–00 (the base year) and 2000–01 (the target year) have been derived by averaging the labour force numbers over the twelve months.[10] Dividing the 2000–01 labour force numbers for each cell by the corresponding labour force numbers in 1999–00 gave us the population reweighting factors for each cell. For those cells in the SIH 1999–00 file for which no matches were found in the LFS data, the above process was repeated by collapsing the cells to broader sub-groupings in reverse order till a match was found.[11]

Table 7.1 below presents a comparison of actual and synthetic total population and its breakdown by selected characteristics for 2000–01. The last column of table 7.1 gives the percentage difference between the synthetic and the actual estimates. As can be seen, the static ageing technique predicts the total SIH 2000–01 population quite well. The predicted population is just 0.3% above the actual population recorded in SIH. It

6 The four labour force status groups are: Employed full-time; Employed part-time; Unemployed; and Not in the Labour Force.

7 The seven age groups are: 15–24; 25–34; 35–44; 45–54; 55–64; 65–69; 70+.

8 Note that while potentially there are 448 cells, only 373 cells in the SIH 1999–00 dataset had observations.

9 Note that since the SIH covers individuals from private dwellings only, we excluded individuals living in non-private dwellings from the LFS file before calculating the reweighting factors for each cell.

10 Given that the SIH is conducted over twelve months (as part of the monthly LFS) we can either use the average over the twelve months (July–June) or the mid-month (December) of the financial year to derive the labour force numbers for each cell. Both methods were used and it was found the method based on averages gave a better estimate of total population for 2000–01 compared to the method based on using the mid-month labour force numbers. Total population was forecast to be just 0.3% above the actual population in 2000–01 based on the first method compared to 0.5% based on the second method.

11 There were seven cells in the SIH 1999–00 dataset for which no matches were found in the original LFS sub-grouping. The reweighting factors for these cells were obtained using the state by sex by labour force status cross-classification.

gives reasonably good estimates of population by sex and age group. In terms of labour force status, while it provides good estimates of those not in the labour force and those employed part-time, it over-predicts those employed part-time by 3.7% and those unemployed by 3.1%. In terms of population numbers by State/Territory, with the exception of the Northern Territory, the method produces reasonably good estimates for the remaining jurisdictions. The method also gives reasonably good estimates of the total number of government benefit recipients, with the predicted total being just 0.3% higher than the actual total.

7.1 Actual and synthetic estimates of Population – 2000–01

	Actual estimates 2000–01 (based on SIH 2000–01) (No.)	*Synthetic estimates 2000–01 (based on Static ageing) (No.)*	*Percent difference (%)*
Total population	14,962,960	15,000,404	0.3
Sex			
Males	7,405,725	7,427,943	0.3
Females	7,557,235	7,572,461	0.2
Labour force status			
Employed full-time	6,679,760	6,611,483	–1.0
Employed part-time	2,348,985	2,435,183	3.7
Unemployed	621,892	641,307	3.1
Not in the labour force	5,312,322	5,312,432	0.0
Age Group			
15–24	2,631,255	2,635,880	0.2
25–34	2,885,223	2,863,965	–0.7
35–44	2,919,568	2,935,366	0.5
45–54	2,596,286	2,603,945	0.3
55–64	1,756,342	1,754,481	–0.1
65–69	635,311	637,857	0.4
70+	1,538,975	1,568,911	1.9
State			
NSW	5,064,384	5,077,647	0.3
Vic.	3,767,319	3,768,117	0.0
Qld	2,769,593	2,766,598	–0.1
WA	1,465,353	1,471,404	0.4
SA	1,182,162	1,192,276	0.9
Tas.	363,772	364,298	0.1
NT	109,755	114,940	4.7
ACT	240,622	245,125	1.9
Total Govt Benefit Recipients	4,267,374	4,280,930	0.3

However, past experience has shown that income estimates are sensitive to the benchmarking regime used. The demographic ageing process needs to incorporate the appropriate benchmarking regime as much as possible. We need to further benchmark the aged weights so that they reflect the wider SIH benchmarking regime, including equal person weights within a household. This work will be done in the next stages of the exploratory study.

7.2 Synthetic estimates of Income

The SIH identifies five major sources of income: wages and salaries; self-employed (unincorporated business) income; investment income; government benefits; and private cash transfers. Each of these income components has been uprated using data from a number of different sources. The uprating factors used for each income type are presented in the Appendix.

For wages and salaries the uprating factors are based on the changes in each of the quintiles of average weekly earnings between 1999 and 2000.[12] Adjustment factors have been calculated separately for full-time and part-time workers. The data source used is the August ABS Employee Earnings, Benefits and Trade Union Membership Survey (EEBTUM).[13]

Income from self-employment (own-business) has been uprated using National Accounts data and unpublished LFS data.[14] Adjustment factors have been calculated separately for farm and non-farm sectors. An estimate of the average farm self-employment income was calculated by dividing the total farm income of unincorporated enterprises (from the National Accounts) by the number of self-employed people in the farm sector recorded in the LFS. An estimate of the average non-farm income of self-employed people is derived in the similar way.

Income from investment has been uprated using data from the National Accounts and the CPI. The National Accounts data provide the aggregate level of income from interest, dividends and rents which allows us to derive uprating factors (with necessary adjustments) for each of these components. The CPI has been used to uprate investment income from other sources.

However, given that the current weekly incomes from self-employment and investment in the SIH are derived from the previous financial year's income from these sources[15], the rating factors for these two income components are based on changes in income between 1998–99 and 1999–00 rather than 1999–00 and 2000 in order to be consistent with the SIH methodology.

Uprating factors for income from government benefits are derived using changes in general rates of payment by major benefit types. In SIH 1999–00, 19 different types of

12 Changes in quintile earnings are better able to capture the dispersion in the distribution of income from wages and salaries than changes at a more aggregated level.

13 We used the EEBTUM Survey which is run in August every year rather than the quarterly AWE Survey because only the former provides a breakdown of weekly earnings by income range which enables us to calculate earnings by relevant percentiles.

14 The National Accounts provides data on farm and non-farm income for unincorporated enterprises while the LFS provides data on the number of self-employed by farm/non-farm type.

15 In SIH income from self-employment and all components of investment income (interest, dividends, rent, etc.) relate to questions that ask about income for the previous financial year and these amounts are converted to weekly amounts by dividing by either 52.14 or the number of weeks in business.

government benefits were identified. These benefit types were grouped into four major categories: Pensions; Allowances; Family Payments; and Others. For the first three benefit categories changes in the standard rate of payment for Age Pension, Newstart Allowance and Family Payment respectively were used to uprate income from these sources. The CPI was used to uprate the remaining government benefits classified as 'Others'.

Income from private transfers (superannuation, workers compensation, child support etc.) was uprated using the CPI.

Table 7.2 presents a comparison of the actual and synthetic estimates of mean and total current weekly income by major income source. As can be seen, the synthetic estimates of total income from wages and salaries, government benefits and private transfers fall well within the 5% range of acceptance. The methodology, however, does poorly in terms of estimating income from self-employment and investment. Despite this overall total income appears to be reasonably well estimated with the synthetic total being just 1% above the actual total, reflecting the fact that unincorporated income and investment income are small components of total income.

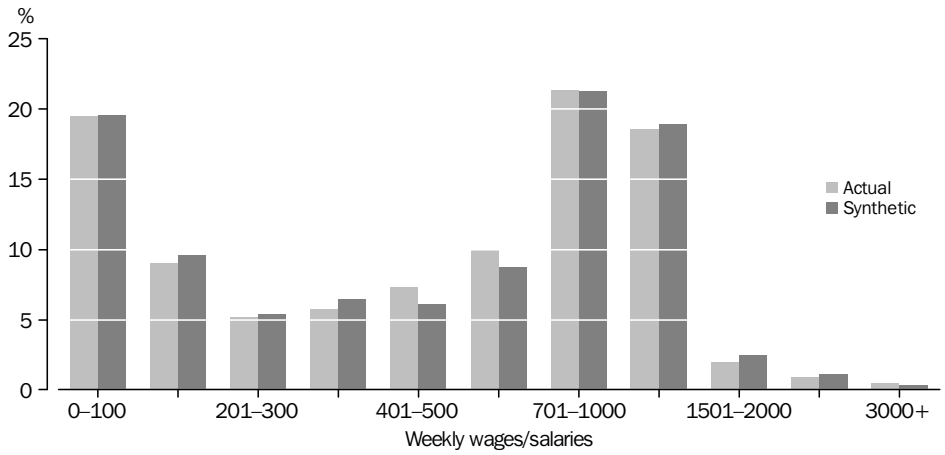
7.2 Actual and synthetic estimates of current weekly income, by major source

	Actual			*Synthetic*		*Percent Difference*	
	Mean ($)*	*Total ($'000)*	*RSE (%)*	*Mean* ($)*	*Total ($'000)*	*Mean (%)*	*Total (%)*
Income source							
Wages/Salaries	569.86	5,145,126	1.2	567.91	5,137,656	–0.3	–0.1
Unincorporated income	35.56	532,109	7.5	38.68	580,260	8.8	9.0
Investment income	20.93	313,235	9.4	24.25	363,774	15.8	16.1
Government benefits	157.81	891,661	0.1	153.41	873,942	–2.8	–2.0
Private transfers	14.99	224,326	4.1	14.71	220,685	–1.9	–1.6
Total income	474.94	7,106,457	1.0	478.41	7,176,317	0.7	1.0

* The mean for wages and salaries has been calculated over those who identify themselves as being either employed full-time or part-time including those who reported zero income from this source. Mean government benefits are calculated over those who receive positive income from this source. For all other income mean is calculated over all observations.

The method quite accurately produces estimates of current weekly total income from wages and salaries. Synthetic estimate of total weekly wages and salaries is just 0.1% below the actual. In terms of the distribution of the current weekly income by income range, as can be seen from figure 7.3, the shares produced by the synthetic estimates appear to be broadly in line with the actual shares for most income ranges except for a few (e.g. $401–500 and $501–700).

7.3 Distribution of actual and synthetic estimates of current weekly wages and salaries



The method significantly overestimates income from self-employment and investment by 9% and 16% respectively. It appears that the income uprating factors derived from the National Accounts overstate the growth in such incomes over the period. It may be noted that the coverage of income in the National Accounts is generally much broader than the SIH, and not all income is reported or collected under SIH. In light of these there may be a need to have a closer look at the data used, including exploring alternative data sources or some other techniques (such as trend smoothing using regression) to derive the uprating factors for these two income components.

Synthetic estimates of income from government benefits is underestimated by 2%. While this estimate appears reasonable, there is scope for further refinement of the estimates. It is possible that the broad benefit categories and the corresponding rates we have applied do not capture all the government benefits received during the period and a more disaggregated approach may be necessary. It may be noted that the standard benefit rates used for estimating income from pensions, allowances and family payments do not take into account additional benefits received through rent assistance, pharmaceutical benefits and any one-off lump sum payments. Such benefits may need to be more specifically taken into account. Alternatively we could explore the possibility of imputing government benefits based on the official rates of payments, eligibility criteria and income tests for certain types of benefits e.g. Family Payments. However, detailed modelling work done in the past has shown it is difficult to predict reported benefits very accurately.

8. FURTHER RESEARCH

The exploratory study to test the feasibility of the static ageing methodology is still incomplete. Only after full and detailed study and analysis will we be able to make a decision on the suitability or otherwise of the proposed methodology. Further analysis including forecasts of income from previous financial year and testing of the distribution of population and income at a more disaggregated level (such as age, family status, numbers, outlays) are planned.

While the procedure used for demographic ageing appears to have produced reasonably good estimates of the population, we could also look at alternative ways of deriving population reweights such as benchmarking the base year population to a limited set of socioeconomic groupings[16], in line with the procedure used in SIH.

As discussed earlier, the benchmark data sources and uprating factors for some income variables will need to be reexamined and necessary adjustments made to improve the forecasts and resulting synthetic estimates. It may be necessary to use other methods such as using regression techniques to derive the uprating factors for some income components instead of applying average growth factors in macroeconomic variables.

We intend to repeat the exercise for another period (say 1996–97 and 1997–98) to see how robust the methodology is in generating estimates of population and income for other periods.

16 Such as state by age by sex, state by number of children aged 0–4, 5–14, state by labour force status etc..

9. CONCLUSION

In this paper we briefly described the static ageing technique and how it could be applied to produce forecasts of income at unit record level for non-survey years. Static ageing essentially involves updating the sampling weights and income of the base year microdata to the required point in time, using a set of benchmark data. To test the usefulness and accuracy of the proposed methodology, an exploratory study is being conducted using past SIH data. If the proposed methodology proves feasible then Phase 2 of the project will involve producing forecasts of income for non-survey years in the future. Validating the results obtained from applying this technique and user acceptance of the results will remain a major challenge for this project.

Preliminary results from the exploratory study undertaken so far are briefly discussed in the paper. The full testing of the model results is still incomplete at this stage. It is only after full and detailed study and analysis will we be able to make a decision on the suitability or otherwise of the proposed methodology. Preliminary results suggest that the static ageing technique does reasonably well in terms of generating population estimates for 2000–01. It also does reasonably well in terms of producing estimates of income from wages/salaries, government benefits and private transfers. It, however, does poorly in terms of producing estimates of income from self-employment and investment. There is further scope for refining the method used for uprating income, including exploring alternative ways for deriving the uprating factors.

ACKNOWLEDGEMENTS

The author would like to thank Leon Pietsch, Graeme Thompson and Gemma Van Halderen, for helpful comments on the paper. Responsibility for any errors or omissions remains solely with the author.

REFERENCES

Bremner, K., Beer, G., Lloyd, R. and Lambert, S. (2002) *Creating a Basefile for STINMOD*, Technical Paper No. 27, National Centre for Social and Economic Modelling, University of Canberra.

Devine, J. and Wertheimer, R. (1981) *Ageing Techniques Used by the Major Microsimulation Models*, Working Paper 1453–01, The Urban Institute, Washington D.C.

Lambert, S., Percival, R., Schofield, D. and Paul, S. (1994) *An Introduction to STINMOD: A Static Microsimulation Model*, Technical Paper No. 1, National Centre for Social and Economic Modelling, University of Canberra.

Merz, J. (1994) *Microsimulation – A Survey of Methods and Applications for Analyzing Economic and Social Policy*, Discussion Paper No. 9, Universitat Luneburg.

Miller, J. and Leaver, D. (1993) *Income surveys and microsimulation – the ABS experience to date*, Proceedings of the Special International Association for Research and Wealth Conference on Microsimulation and Public Policy, Canberra.

Mitton, L., Sutherland, H. and Weeks, M. (eds) (2000) *Microsimulation for Policy – Challenges and Innovations*, Cambridge University Press.

APPENDIX. INCOME UPRATING FACTORS

	August 1999	*August 2000*	*Ratio*
Wage and salaries*			
Full-Time ($/w)			
Median of Q1	400	411	1.03
Median of Q2	529	560	1.06
Median of Q3	650	690	1.06
Median of Q4	840	878	1.05
Median of Q5	1,200	1,250	1.04
Part-Time ($/w)			
Median of Q1	55	56	1.02
Median of Q2	125	135	1.08
Median of Q3	235	245	1.04
Median of Q4	340	350	1.03
Median of Q5	500	528	1.06

* Based on August Employee Earnings, Benefits and Trade Union Membership Survey for each year.

	1998–99	*1999–00*	*Ratio*
Unincorporated Income (UI)			
Non-Farm UI* ($m)	8,712	9,091	1.04
Non-Farm Self-employed LF** (000)	1,010	1,032	1.02
Average NonFarm UI ($)	8,624	8,813	1.02
Farm UI* ($m)	1,134	1,381	1.22
Farm Self-employed LF** (000)	227	236	1.04
Average Farm UI ($)	4,988	5,860	1.17

* Based on national accounts data

** Based on Labour Force Survey data

Farm and non-farm income and labour force numbers are average of 4 quarters.

	1998–99	*1999–00*	*Ratio*
Investment Income			
Interest (National Accounts) ($m)	16,171	17,352	1.07
Dividends (National Accounts)($m)	9,163	10,198	1.11
Rent* (National Accounts)($m)	17,351	18,510	1.07
Others (CPI)	121.8	124.7	1.02

* Excludes imputed components

	1999–00	*2000–01*	*Ratio*
Government Benefits*			
Benefit Type			
Age Pension (Single, std rate) ($/f) (Jun)	372	402.1	1.08
NewStart Allowance (Single <21 yrs) ($/f) (Jun)	331.6	357.8	1.08
Family Payment (Child <13 yrs) ($/f) (Dec)	101.6	119.6	1.18
Others (CPI) (1989–90=100)	126.2	133.8	1.06

* Based on Centrelink payment rates

	1999–00	*2000–01*	*Ratio*
CPI			
Average over financial year (1989–90=100)	124.7	132.2	1.06

FOR MORE INFORMATION...

INTERNET **www.abs.gov.au** the ABS web site is the best place for data from our publications and information about the ABS.

LIBRARY A range of ABS publications is available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries..

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our web site, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice..

PHONE 1300 135 070

EMAIL client.services@abs.gov.au

FAX 1300 135 211

POST Client Services, ABS, GPO Box 796, Sydney 2001

FREE ACCESS TO PUBLICATIONS

All ABS statistics can be downloaded free of charge from the ABS web site.

WEB ADDRESS www.abs.gov.au

2000001523445
ISBN 0 642 48169 5

RRP $11.00

© Commonwealth of Australia 2006

Produced by the Australian Bureau of Statistics