
**DEMOGRAPHY WORKING PAPER 2000/1**

# PRODUCTION OF POST-CENSAL SMALL AREA ESTIMATES OF HOUSEHOLDS
# A PRELIMINARY INVESTIGATION

**David Bartie**
**Small Area Population Unit**
**Australian Bureau of Statistics**
**February 2000**

## Comments Sought

In 1999 the Small Area Population Unit (SAPU) of the Australian Bureau of Statistics (ABS) undertook a preliminary investigation into the production of small area household estimates in post-censal years.

This paper discusses the household estimates presently produced by the ABS and investigates several options for the production of post-censal small area household estimates. The paper concludes by suggesting a number of recommendations for future work.

ABS welcomes any comments you may wish to make about this paper.

Written comments should be addressed to:

David Bartie
Small Area Population Unit
Australian Bureau of Statistics
GPO Box 2272
Adelaide SA 5001

Or david.bartie@abs.gov.au

Telephone comments can be made to David Bartie on (08) 8237 7385.

## Contents

# 1. Introduction

The Australian Bureau of Statistics (ABS) currently produces estimates of households at the Statistical Local Area (SLA) level in Population Census years and at the capital city/balance of State level in non-Census years. Household estimates serve a number of important needs, both within the ABS and in the broader community.

The Small Area Population Unit (SAPU) has been commissioned to determine the feasibility of producing post-censal household estimates for smaller areas.

This paper documents SAPU's work to date on the project. Sections 1, 2 and 3 offer some general background information on the topic, discussing issues such as the demand for household estimates, differences between households and dwellings and methods currently used to produce small area household estimates both within and outside the ABS. Section 4 presents a more detailed discussion of the various methods considered. Overall, 6 different methods are considered, with some being judged more suitable than others. At this stage it still fairly unclear which of these methods, if any, represents the best option for the production of small area household estimates. Consequently the paper concludes by setting out a number of recommendations for further work.

Unless otherwise specified, for the remainder of this paper the term 'State' refers to Australia's States and Territories.

## 1.1 Current Demand for Household Estimates

Estimates of households and the distribution of the household population are useful for a wide range of external organisations involved in a variety of different areas. Providers of health, housing, family and community services tend to focus the bulk of their efforts towards the household, with factors such as unemployment, poverty, income distribution and housing needs often linked with the household rather than the individual. In addition, the household is the base unit of consumption for a wide range of services such as electricity, water and telephone services as well as for commercial products such as fridges, cookers, televisions and home computers. Analysis of household distribution and composition provides information about the regions in question and about society in general, giving an insight into factors such as family size, marriage rates and the ages when children leave home. Household estimates/projections are also useful for planning purposes, ensuring that future growth can be more accurately estimated and infrastructure provided in a timely and efficient manner.

An extensive review of demography statistics conducted in July 1996 (ABS, 1996d) included widespread consultation with external users and highlighted the fact that many external organisations such as State planning agencies are more interested in statistics relating to households rather than individuals, since household data is often more relevant in terms of urban and welfare planning and the provision of services. A desire for frequently updated estimates at the SLA level was also a common theme amongst current/potential users.

Recent discussion between the ABS and representatives of the Victorian Department of Infrastructure identified a number of key outputs in relation to household statistics, specifically household composition data, counts of the total number of households, and a measure of the average household size.

In early 1997, in a response to an ABS proposal relating to family estimates, the West Australian Ministry for Planning indicated that "estimates at the SLA level will be a valuable addition to the current understanding of demographic forces at the small area level and the methodology used to estimate household population by age and sex at that level would be most welcome."

Within the ABS, there is also potential for the further application of household estimates as benchmarks if estimates can be produced at the SLA level.

### 1.2 What are the Differences Between a Household and a Dwelling?

In the Census, dwelling information is obtained separately for each household. Therefore, since a census form is completed for each household, the total number of occupied private dwellings is equal to the total number of households.

Despite this, key differences between households and dwellings are acknowledged. In the 1996 Census Dictionary, a dwelling is defined simply as "a building or structure in which people live". Dwellings are categorised according to whether they are occupied or unoccupied at the time of the Census, and whether they are private or non-private dwellings. A private dwelling is normally a house or a flat, but can also be part of a house, a caravan, a houseboat or even a tent. Non-private dwellings are those that provide a communal or transitory type of accommodation such as hotels, motels, guest houses, prisons, defence establishments, hospitals and aged care facilities.

The Census Dictionary defines a household as follows:

*A household is defined as a group of two or more related or unrelated people who usually reside in the same dwelling, who regard themselves as a household, and who make common provision for food or other essentials for living; or a person living in a dwelling who makes provision for his/her own food or other essentials for living, without combining with any other person.*

Consequently, there are a wide variety of household types including family and non-family households, single-parent households, same-sex couple households and one-person households. A household comprises of persons who are usually resident in a private dwelling. Persons usually resident in non-private dwellings are not included in household estimates.

Thus, while a dwelling is defined purely in terms of the type of building, a household is defined according to the number and characteristics of the residents.

Analysis of Labour Force Survey data for 1991 confirmed the fact that most private dwellings contain only one household, with the incidence of multiple household dwellings only 0.7%.

## 2. Household Estimates Currently Produced by the ABS

### 2.1 Census Years

In the Review of Demography Statistics (ABS, 1996d), the ABS made a commitment to produce household estimates for census years at the SLA level. A set of provisional estimates was subsequently produced from the 1996 Census. This involved a number of steps -

1. For each SLA and age group (0 to 14 years, 15 years and over), the percentage of persons living in private dwellings was calculated. These percentages were then applied to the respective estimated resident population counts to obtain an estimate of the number of people living in private dwellings as at 30 June.

2. Each occupied private dwelling counted in the Census was classified according to the number of residents aged 15 years and over and aged under 15 years living in that dwelling. Thus, for each SLA, a matrix was produced showing number of private dwellings by composition (0 to 6+ persons under 15 years and 1 to 6+ persons 15 years and over). Given the low incidence of multiple household dwellings, this was assumed to be the equivalent of a household composition matrix for that SLA.

3. From the household composition matrix an estimate of the population by age group in each household type was obtained by multiplying the number of dwellings in each group by the corresponding number of residents. These estimates were then separately prorated so that the total household population equalled the estimate of the number of persons living in private dwellings, calculated previously (Step 1).

4. Estimates of the number of households by household composition could then be produced by dividing the number of persons in each group by the appropriate household size. For example if there were 120 people living in households with 2 persons aged 15 years and over and 1 person aged less than 15 years, the total number of households of this type equals 40 (120 ÷ 3).

This method therefore aligns with the ABS definition of "households" by automatically excluding visitors and residents of non-private dwellings and including absentees who are usually resident in the household.

This method produces good estimates of households because it focuses on the household population in the respective SLAs. This allows for relatively straightforward adjustments for Census undercount, residents temporarily overseas and backdating of data to 30 June (from the Census date of 6 August). However, there are some minor concerns. These relate to the suspected occurrence of double-counting in the Census in cases when a usual resident is temporarily absent on Census night and the degree of household undercount which occurred in the Census, specifically whether certain household types are more prone to undercount. Following each census, undercount is estimated for persons, dwellings and households. Previously, only estimates of total household undercount have been published. For the 2001 Census it is planned to make available estimates of household undercount according to household type (number of persons aged 15 years and over and number of persons aged under 15 in residence). This will help to confirm whether certain types of households are more likely to be missed. Another concern with estimates of households in Census years is the inaccuracy associated with estimation of populations of large households. This relates to the construction of the household composition matrix itself - larger households may only be classified by the minimum number of people in a particular age group who are residents. For example, a dwelling may be listed as having "6+" residents aged less than 15 years. Some inaccuracies may subsequently arise when using this information to derive population estimates.

**2.2 Non-Census Years**

During the post-censal period, the ABS produces household estimates at the part of State level (capital city/balance of State), with households classified according to the number of persons aged 15 years and over and the number of persons aged under 15 years in residence. “Household Estimates Australia, 1986, 1991-94" (ABS, 1996c) describes the methodology used, which is based on the household propensity method first proposed by Ironmonger and Lloyd-Smith. Household propensities refer to the probability of particular types of people to reside in households of a given type.

1. Using the results of the Census, the household population is distributed into particular household types, classified by the number of adults and the number of children in residence. Based on the proportion of the population in each type of household, a household propensity matrix is created (0 to 6+ persons aged under 15 years of age and 1 to 6+ aged 15 years and over).

2. Similarly, a household propensity matrix is created for the same period using the Labour Force Survey (LFS) for that period. The LFS is a monthly sample survey of between 30,000 and 35,000 households.

3. A Census adjustment factor is calculated by dividing the results of (1) by the results of (2).

4. The household propensities for the post-censal period are calculated using data from the appropriate LFS and then applied to population estimates for the period in question to produce the household distribution of the population. Prior to this the population estimates are adjusted to omit those people who are usually resident in non-private dwellings. The results are then applied to the Census adjustment factor. Thus, in simple terms, the results reflect those that might have been expected had a census been conducted at that time.

5. From the household distribution matrix the number of households of each type is calculated based on the number of persons aged 15 years and over resident in households of that type. For example, if 50 persons aged 15 years and over are shown to be resident in households containing 2 persons aged 15 years and over and 2 persons aged less than 15, the total number of households of that type is 25 (50 ÷ 2).

Due to the fact that the LFS is conducted monthly, up-to-date propensity information is available throughout the post-censal period. The LFS is designed to produce reliable estimates at the national, State and Labour Force Region (LFR) levels. There are currently 77 LFRs across Australia and they can consist of anything from a handful of SLAs in the centre of the larger metropolitan areas to entire Territories in the case of NT and the ACT. The Household Estimates Review (ABS, 1999b) has recommended that options be investigated for the smoothing of LFS propensities, since there is evidence that the propensities may display considerable volatility over time. The need for smoothing increases as the LFS sample size decreases for a given household composition and area. However, despite this, the review concluded that the propensity method offers "a sensible way of producing household estimates" and noted that recommendations for the improvement of ABS household estimates should focus on the way the method is implemented rather than on changes to the method itself.

## 3. Estimates of Households Outside the ABS

Much investigation has been conducted, particularly via the internet, of academic, government, and statistical sources external to the ABS to determine if estimates of households at the small area level have been produced elsewhere and to determine whether such methods could be adapted by the ABS to produce household estimates below the part of State level.

It appears that most estimates and projections of households available from overseas statistical agencies are produced using some form of the headship rate method. This method involves calculating the proportion of people in a particular population who are the heads of households and using these rates, in conjunction with an independent estimate of population, to estimate the number of households at a given point in time (Spicer, Diamond and Ni Bhrolchain, 1992). Household estimates available from other external sources also commonly use this method. The ABS ceased using the household head classification, which was felt to have become unrepresentative of modern society, at the time of the 1986 Census. Data output resulting from the use of this methodology is often limited in that rarely is data available concerning the "other" members of the household. Thus the use of such data for analysis and planning purposes is subsequently limited. In addition, since headship rates are normally derived from a census there is no easy way of updating these rates during the post-censal period, particularly for smaller areas.

The other method most commonly associated with the production of household estimates outside the ABS is the housing unit method. Technically more a measure of dwellings than households, this method involves the update of dwelling/household information from the most recent census using available data relating to building construction and demolitions.

While the headship rate method is normally used to produce estimates for larger areas, the housing unit method is commonly used to estimate for smaller areas. This is partially a reflection of the fact that organisations primarily concerned with smaller areas such as local councils, urban development agencies and universities, often have easy access to the necessary local area data relating to building construction and demolition. The main issues associated with this method are determining the lag between a dwelling being approved, being constructed and subsequently becoming inhabited,

and the availability of good quality construction and demolitions data or data which can be used as a proxy for these events, such as building approvals and electricity connections/disconnections.

As is the case in the ABS, a number of external agencies currently or potentially involved in the production of household estimates use or are moving towards the use of geographical information systems (GIS) and the geocoding of household information. Geocoding refers to the process of assigning a latitude and longitude to a statistical unit such as a dwelling or business location. In terms of households data, geocoding of administrative data sources could theoretically in future lead to a dataset containing precise locational information for each private household as well as associated data relating to each household such as size and structure. The Australian Electoral Commission (AEC) reported in 1996 that "within 5 years" it intends to geocode all "habitats" (places where people live) in Australia, while organisations such as Telstra and Australia Post also geocode the address information they collect. Organisations such as the Key Centre for Social Applications of GIS (GISCA), based at the University of Adelaide, already make extensive use of geographical information systems in relation to demographic data. While the application of this technology is still at a developmental stage, it has the potential in future to be an extremely valuable source of household data.

A brief summary of the household estimates methods adopted by the major overseas statistical agencies follows.

### 3.1 Canada

Prior to 1994, household estimates from the Canadian Census were published at the provincial level and were also available for large metropolitan areas. Since then, Statistics Canada have ceased publishing household estimates although still produce estimates for internal purposes and upon request. During post-censal periods, household estimates/projections are produced at the provincial and sub-provincial levels (roughly equivalent to States and Statistical Divisions in Australia) according to the headship rate method, classified according to age-group, sex, marital status and household type (family and non-family).

Currently, Statistics Canada have no plans to produce household estimates at the small area level.

### 3.2 New Zealand

Statistics NZ produces post-censal estimates of households at the national level by updating census counts based on measures of building construction and demolition. Estimates are of the total number of households only. In addition, a wide range of information on households and families is available from the annual Household Economic Survey, with the results of this extrapolated to the population as a whole. Household projections are produced for regional council areas (roughly equivalent to Statistical Divisions in Australia) using a householder (headship) rate method.

At this stage, there are no plans to produce household estimates for smaller geographic areas, although Statistics NZ have indicated that the method used for preparing the national level estimates could be easily adapted to produce estimates for smaller areas. Indeed, household estimates have been prepared on request by Statistics NZ for specific local authorities using a similar method incorporating building construction and demolition data.

### 3.3 United Kingdom

The UK Office of National Statistics produces annual household projections at the standard region level (roughly equivalent in terms of population size to States in Australia) by updating current household headship rates based on past trends and applying these to population projections. Annual estimates of households are produced at the local authority level (roughly equivalent to Statistical Divisions in Australia) by applying household headship rates from the census and updated from an annual labour force survey to the latest population estimates.

Recent efforts have focused on improvements to the current methodology, for example via the inclusion of economic factors, rather than on the production of household estimates for smaller geographic areas.

### 3.4 United States of America

The US Bureau of Census produces annual estimates of households at the State level using a combination of census and administrative data (building permits and electricity customers). "Householder formation rates" by age of householder are derived from the census and then applied to State population estimates. In addition, a family based approach is also used, with results of the annual Current Population Survey used to estimate the numbers of householders by family type. These results are combined with estimates of population derived from census counts updated by births, deaths and migration data to estimate the total number of households. Household headship rates based on the most recent census are used to produce household projections. The US Bureau of Census also appears to conduct a wide range of localised household studies particularly in relation to ethnicity and census undercount. As part of these studies detailed household information is collected by interviewing local residents, however the study areas are usually very small.

## 4. Options for the Production of Small Area Household Estimates in Post-Censal Years

The method currently used by the ABS produces estimates of households at the part of State level. Detailed household information is available from the Census and due to the fact that the LFS is conducted monthly, up-to-date propensity information is available throughout the post-censal period. However, due to the limitations in the size of the LFS sample it cannot be used to produce updated propensity information for small areas such as SLAs. Recent working papers such as "Future Directions for ABS Household Surveys" (ABS, 1998a) and "Strategic Directions for Social Statistics 1999-2000" (ABS, 1998b) have stressed this point, with the former paper stating that "national household surveys just cannot deliver small area data". Survey "estimates" at the SLA level have previously been produced in the ABS, most notably from the Survey of Disability, Ageing and Carers, using a synthetic estimation technique. However, it was conceded that, particularly for very small SLAs, such estimates might be quite unrealistic and it was intended that they only be used as "building blocks" for larger areas such as Statistical Divisions. Previously, the LFS has also been used to produce SLA unemployment data. This method incorporated the use of postcode data from the Department of Social Security relating to the location of people collecting benefits, but was later discontinued in order to re-evaluate the method. It therefore appears doubtful that either of the techniques used in the case of the disability survey or in the production of unemployment data would be appropriate for the production of household estimates.

SAPU has considered a variety of different options for the production of small area household estimates in non-Census years. No matter what the method(s) finally selected, it is intended that the estimates sum to the already available part of State estimates. As a result, post-censal changes in propensities at the LFR level will be partially taken into account. An option that may be worth investigating is whether the currently available household estimates can be disseminated at the LFR level, rather than solely at the part of State level. Given that the estimates are produced using the LFS it is suspected that this may be feasible. If this is the case it would then be possible to sum SLA estimates to LFR totals, resulting in improvements in accuracy.

The various options considered by SAPU to produce small area household estimates will now be discussed in detail.

### 4.1 Application of Census Propensities Throughout the Post-Censal Period

Given the fact that Census propensities cannot be easily updated in non-Census years, one of the simplest methods for producing SLA household estimates is to assume that Census based propensities remain constant during the following post-censal period. This was one of the options considered by Ironmonger and Lloyd-Smith in their original paper (Ironmonger and Lloyd-Smith, 1992). They calculated propensities for particular age and sex groups from the 1981 and 1986 Censuses for Australia as a whole and looked at the degree of difference between the two. The result of this was that they noted "substantial shifts" in propensities during the period and therefore concluded that the assumption that Census based propensities would remain constant was unrealistic.

The fact that Ironmonger and Lloyd-Smith calculated propensity information for such a large number of age groups (five year age groups from 0-4 to 75+) and household types (37 in total) may have been a factor in the degree of variability they found. The ABS would be unlikely to require data at such a level of detail and it is possible that less detailed propensities would not be so prone to fluctuation. However, this should be balanced with the fact that Ironmonger and Lloyd-Smith only looked at propensities across Australia as a whole, while for this project the ABS is concerned with SLA propensities. It can be expected that at this level the propensities are likely to be prone to increased variability over time. In particular, it is suspected that propensities in rapidly growing SLAs are more variable than those in SLAs which are more stable. Since it is not unusual for growth at the SLA level to occur in spurts, it is therefore doubly important to have access to regularly updated propensity information. Given these issues, SAPU considers it important that mechanisms for the update of propensities during the post-censal period be investigated.

### 4.2 Use of Ironmonger & Lloyd-Smiths' Method of Updating Propensities

Upon reaching the conclusion that the constancy of household propensities during the post-censal period could not be assumed, Ironmonger and Lloyd-Smith proceeded to consider a method for the update of Census-derived propensities (Ironmonger and Lloyd-Smith, 1992). The method they used involved observation of the change in propensities during the previous inter-censal period. They then assumed that this trend would continue in the subsequent period and updated the new set of Census based propensities accordingly. Due to the fact that propensities were calculated separately for adults and children, when they were applied to the population estimate, the final numbers of adults and children in particular household types was not consistent. A modification of the data was subsequently conducted to deal with any "excess" adults which involved their removal to households without children.

In theory, a similar method could be adapted for the production of SLA household estimates. However a number of potential problems are immediately apparent. Without modification, this method would require the production of Census-derived propensities for every SLA in Australia for at least two Census years. A major factor in this would be conversion of the Census data to a consistent set of SLA boundaries. Due to the extent of SLA boundary changes during the 1991-96 period, this would be a potentially difficult and time-consuming task (the issue of boundary changes is further discussed in section 5.1). Another issue is that while this method offers a means to update propensities, there are doubts whether the method of updating is responsive enough to reflect changes in the real world. As mentioned previously, growth at the SLA level commonly occurs in spurts, so using trend information from the previous inter-censal period may be inaccurate in the not uncommon situation that a moderately growing SLA suddenly goes through a period of rapid growth. Similarly, it can be expected that rapidly growing SLAs will eventually experience a gradual slowdown. If, as suspected, change in growth rates is a factor in change in propensities, use of propensity change trends from the previous inter-censal period may not be a sound approach.

### 4.3 Use of a Regression Methodology Similar to that Used for SLA Population Estimates

Currently post-censal population estimates for all SLAs are produced annually by SAPU based on a regression methodology. Briefly, this method establishes relationships between population and a variety of indicator data, such as Medicare counts and building approvals, which allows change in population to be modelled.

On first impressions, it appears feasible that a similar method could be used to estimate SLA household numbers. There is the option of producing a single model which will estimate the total number of households in a particular area, or developing a number of models to separately estimate different household types. It can be expected that a stock of dwellings measure such as building approvals or electricity connections, combined with other data such as Medicare counts or possibly population estimates, would be a very good indicator of households.

However, a potential problem is the suspected degree of correlation between a dwellings indicator and the number of households in most SLAs, meaning that this indicator could "crowd out" all other variables in the regression model. "Crowding out" arises when a particular explanatory variable and the response variable are so strongly correlated that when the regression models are produced, the coefficients associated with other explanatory variables become insignificant. Thus, for all intents and purposes the estimates derived from such a model are simply a derivative of one indicator variable. This problem of crowding out can therefore be expected to effect regression models which use a dwellings measure to estimate households. At the same time, it is doubtful whether a model which did not incorporate a dwellings component could produce an accurate estimate of households. It is assumed that a determination of the number of houses and flats in a particular area will be important in the estimation of households and there is no obvious way to measure this other than through the use of building approvals data.

One alternative option is to simply use population estimates as an indicator variable, since it can be expected that the population count and structure, will be strongly associated with the number and type of households in a particular SLA. It can also be argued that a result of dwellings being used in the models to produce population estimates, the subsequent use of these estimates to model households means that to some degree, dwellings will still be incorporated.

Further testing is therefore required to firstly determine whether indicators other than dwellings can be successfully used in a regression model to estimate households and whether the subsequent use of population estimates in such a model is valid.

### 4.4 Simple Adaptation of Dwelling Stock Measures

Given the potential problems associated with the use of regression models and with the update of Census propensities in non-Census years, it is possible that the most accurate estimates of households may in fact be produced via the use of dwelling stock measures.

Data relating to number of dwellings is available from a variety of sources. Within the ABS, data on building approvals is published every month. This data is available for Collector’s Districts (CDs) which can be aggregated to form SLAs. Approvals are classified both by the type of building and type of work involved. A stock of dwellings figure by type (houses/flats) can therefore be produced for every SLA at any point in time by adding to Census counts the number of approvals that have occurred in the intervening period. This housing unit method has been discussed previously and is used fairly extensively by external agencies to produce estimates of dwellings, households and subsequently, population.

The main disadvantage in using building approvals data in this manner is that the ABS does not collect statistics relating to demolitions, so theoretically the dwelling stock cannot reduce. There have long been problems associated with the difficulty and costs of identifying and collecting demolitions data. The 1994 Building Statistics Review (ABS, 1994b) highlighted the fact that demolitions data was potentially available for most States, however the accuracy and coverage of this data was at that stage difficult to assess, and the fact that data on demolitions is collected by a number of different bodies using a number of different techniques means that comparison between States might not be

valid. In addition, the review concluded that the availability of demolitions data for Qld and NT was unlikely given the fact that neither of these States have legislation requiring the reporting of this type of information. More recently, the Review of the Functional Classification of Buildings (ABS, 1999a) recommended that the “type of work” classification used by the ABS be expanded to include "demolition of existing buildings".

Other problems associated with the use of building approvals data are determining the lag between a building being approved and completed and accounting for those approvals which never result in a completed dwelling. However, given that these latter problems have been successfully managed when approvals data has been used elsewhere, namely in the production of population estimates, they are less of a concern than is the issue of demolitions.

Preliminary work has been conducted by SAPU to determine the accuracy of dwelling stock counts produced by the update of Census data using approvals. This involved extraction of dwelling counts from the 1991 and 1996 Censuses as well as building approvals data for the intervening period. The 1991 Census counts were then updated using building approvals and the results compared to the counts from the 1996 Census. For simplicity, this analysis was confined only to NSW, Qld and NT. SLA boundaries in all other States underwent significant revision during the 1991-96 period and the process of converting historical approvals data to the new boundaries is potentially an additional source of variability. Also, SLAs where the 1996 Census count was less than 20 for a particular dwelling type have been omitted to minimise the impact of very small SLAs. A summary of this analysis is presented in the following table.

| | **State/Territory** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **NSW** | | | **Qld** | | | **NT** | | |
| | Houses | Flats | Dwell. | Houses | Flats | Dwell. | Houses | Flats | Dwell. |
| Proportion of “accurate” SLAs | 78.5 | 46.8 | 83.5 | 69.7 | 25.4 | 66.1 | 59.3 | 41.3 | 55.0 |
| Proportion of “inaccurate” SLAs | 3.2 | 29.8 | 4.3 | 10.6 | 54.1 | 14.1 | 35.6 | 41.3 | 33.3 |

In the above table "accurate" applies to those SLAs where the updated estimate was within 5% of the Census figure. "Inaccurate" applies when the discrepancy was more than 10%.

It should be noted that a number of factors may be influencing the observed discrepancies. Firstly, as reflected in the fact that very small SLAs have been excluded from the analysis, smaller SLAs are likely to be prone to higher discrepancies. In such cases a difference between updated and observed counts of only a few dwellings may still translate to a high percentage discrepancy. This may be a major factor in explaining the poorer results for Qld and NT, since in these States average SLA size is relatively small. The fact that not all building approvals result in a completed dwelling can be expected to also be a factor as well as the previously discussed point that demolitions are not taken into account. The inflation of the updated figure which occurs as a result of these factors may be the reason why in all cases a much higher proportion of SLAs were overestimated than were underestimated. Another consideration is that the assumption that the Census figure represents the "true" value may not be entirely accurate. Sources of error associated with the Census count would include Census undercount and the fact that in isolated cases the "type" of a dwelling may have been misinterpreted by Census collectors.

In recognition of the problem of a discrepancy between approvals and completions, further analysis was carried out using adjusted approvals data. Approvals data was adjusted by comparing the change in dwellings between the Census dates with the total number of approvals during the period and forcing the approvals count to the Census dwelling change. This is a similar process to that which would occur when any SLA household estimates produced using approvals data are forced to the part of State household estimates. The adjusted figures for each SLA were then compared with the corresponding 1996 Census figure. The results of this analysis are summarised in the following table, with the bracketed figures indicating the "forcing factor" applied to the approvals data for each dwelling type (houses, flats and total dwellings). For example in NSW, a factor of 0.56 was applied to

force house approvals to the change in the number of houses between the Censuses. As before, very small SLAs where the 1996 Census count was less than 20 for a particular dwelling type have been excluded.

| | State/Territory | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | NSW (0.56, 0.84, 0.68) | | | Qld (0.87, 0.77, 0.84) | | | NT (0.64, 0.98, 0.75) | | |
| | Houses | Flats | Dwell. | Houses | Flats | Dwell. | Houses | Flats | Dwell. |
| Proportion of "accurate" SLAs | 81.7 | 50.3 | 91.5 | 72.6 | 29.5 | 68.2 | 50.8 | 41.3 | 50.0 |
| Proportion of "inaccurate" SLAs | 3.2 | 24.6 | 4.8 | 10.8 | 51.6 | 12.8 | 32.2 | 41.3 | 35.0 |

It can be seen that the forcing of approvals data results in some improvements in overall "accuracy" of the updated dwelling counts for NSW and Qld. However in the case of NT, the overall "accuracy" is less following forcing. From this is can be assumed that there is a greater variety in discrepancies at the SLA level in NT than is the case for the other States.

As is the case in most examples of the housing unit method outside the ABS, it is possible that a range of proxy variables such as electricity, telephone and water service connections could be used in preference to building approvals data. This data has the advantage that by looking at both connections and disconnections, building demolitions can be accounted for. A good evaluation of the use of indicator data to estimate dwellings appears in the paper entitled "Evaluating the Housing Unit Method" (Smith and Cody, 1999). Although primarily a discussion of techniques for estimating population, methods for producing accurate estimates of dwellings are also discussed. In this paper, dwelling estimates are produced by updating census data using a range of indicator data such as telephone connections, certificates of occupancy and property tax records. It was found that electricity connections and building permits data led to the best results.

GISCA is currently conducting a study to determine the feasibility of using data on electricity meter removals to estimate building demolitions in South Australia, as well as looking into potential sources of demolitions data in other States. The main issue here is the ability of the ABS to obtain these types of proxy data. The data would have to be readily available across Australia at the SLA level or a level which could be easily converted to SLAs. Prior investigations have indicated that electricity connections data, for example, is currently only available for the ACT, although it might be available in the near future for some other States. An extensive evaluation of indicator data was recently conducted as part of an ABS study into estimation of service populations (ABS, 1999c). While this deals primarily with person counts, much of the findings in regard to the suitability and quality of indicator data are also applicable to estimates of dwellings. It was found that most of the indicator data considered fell short of particular availability, quality and/or coverage criteria.

SAPU has obtained electricity connections data for the ACT at 30 June 1991 and 1996 and compared these to 1991 and 1996 Census counts. The result was that the overall difference between residential electricity connections and total dwellings in both years, excluding SLAs with less than 20 dwellings, was around 2.5%, with around 2,500 more connections than dwellings. However at the SLA level the average discrepancy between the two counts, excluding SLAs with less than 20 dwellings, was much higher, being 9.6% in 1991 and 8.1% in 1996. Thus, despite in theory measuring the same thing, it can be seen that in this case counts of electricity connections and dwellings are quite different. It should be noted however that this degree of discrepancy may be partly due to the small average SLA size in the ACT. Updating the 1991 Census count using residential connections in a similar manner as with approvals gives an overall discrepancy of 0.4% for the ACT as a whole. However, once again the average discrepancy at the SLA level is much higher at 11.4%. Preliminary analysis of some of the larger discrepancies between Census dwellings and electricity connections indicates possible differences in the way private residential dwellings are classified in each data source. For example in the case of some SLAs where the Census has identified a large number of non-private dwellings, it appears that some such dwellings may in fact be included in the counts of private residential electricity connections. Further investigation is warranted

but it is clear that without modification there may be problems in using electricity connections as a proxy for dwelling stock and approvals/demolitions.

There are a number of independent sources of stock of dwellings data currently available outside the ABS. The ACT Government publishes quarterly counts of the number of dwellings for each SLA/LGA in the territory. This information is compiled using data from the ACT Office of Asset Management. In comparison with the Census, ACT government dwelling counts were just 0.1% lower in 1991 and 0.1% higher in 1996. However as was the case with electricity connections data, the average discrepancy for SLAs with at least 20 dwellings was considerably higher, at 13.7% in 1991 and 14.9% in 1996. Once again, despite the exclusion of very small SLAs from this analysis it is fair to assume that much of the discrepancy is due to the overall small average size of SLAs in the ACT. As was the case with electricity connections the figures may also be affected by differences in the definition of private residential dwellings.

The State Government planning agency in South Australia, Planning SA, maintains a detailed GIS containing precise geographical and other information about dwellings in the State. This is updated using information from the Valuer-General. As soon as an application for land subdivision occurs this information is placed on the system, with a range of dwelling information added once a valuation of subsequent development occurs (triggered by the connection of water). Similar systems exist in WA and Tasmania with other States expected to follow.

Another potential source of dwelling stock data is the national address file of Australia Post. This contains information on every postal delivery point in Australia, with each identified by postcode, street name, number and type (business/residential). New addresses are added based on the observations of postal delivery officers and following the lodgement of mail redirections. In 1995, a study conducted by the Qld Office of the ABS compared residential delivery point data from Australia Post with ABS estimates of private dwelling numbers at both the postcode and the LGA level. The results varied depending on the particular area in question but were described as generally "quite promising". Reasons for discrepancies between the two datasets were identified as possible over-estimation of the dwelling stock by the ABS since demolitions are not taken into account, and possible under-estimation of the dwelling stock by the Australia Post data as a result of some multi-household units having only one delivery point. Another problem with this data is the widespread use of post office boxes which may be located some distance from the dwelling in question. This is particularly the case in rural areas.

There are therefore a number of options incorporating a more direct use of buildings data to estimates household numbers. The crucial issue here is that dwellings and households are not the same thing. As mentioned previously, a dwelling may contain more than one household and although analysis of 1991 LFS data put the incidence of multiple household dwellings at a very low level (0.7%), their existence should not be totally discounted.

Another issue is that if household estimates are produced from this method they will not be directly relatable to population estimates. The ABS produces population estimates based on a regression model which normally incorporates a "dwelling" component, such as house approvals, and a "person" component such as Medicare enrolments. In some cases it will therefore be the case that an SLA's resident population estimate will rise or fall due to Medicare rather than building approvals. Such changes in population will not be taken into account if household estimates are produced directly from building approvals. This is certainly not desirable and therefore population estimates would have to be incorporated in some manner if it was decided to proceed with this method. An additional problem is that any household estimates produced directly from buildings data would be limited in detail since only totals could be produced, with no breakdown of the household structure. As shown by the analysis comparing Census dwelling counts updated by approvals with the corresponding values from the next Census, there may be a fair degree of inaccuracy in this method at the SLA level. However as mentioned, such discrepancies may in part be a reflection of issues such as small SLA size, Census undercount, difficulties associated with classifying building type and the absence of demolitions data.

Despite these problems, due to deficiencies in the other methods of SLA household estimation, it may in fact be the case that use of a stock of dwellings figure to estimate households produces the most accurate results.

### 4.5 Use of Indicator Data to Estimate Change in Propensities

Section 4.3 has discussed the problems associated with the use of a regression methodology to produce SLA household estimates, while sections 4.1 and 4.2 have described the problem of updating Census propensities in non-Census years. Rather than using a regression model to directly produce household estimates, it may be possible to use a model to estimate the change in Census propensities in subsequent years. The crowding out problems associated with the more direct regression approach would be avoided since indicators such as building approvals will not be so strongly correlated with the dependant variable. Also, change in propensities would be based on current data rather than on trends from the previous inter-censal period. The methodology to be used for this process will now be described in more detail.

The first step would be to derive SLA household propensity information from the 1986, 1991 and 1996 Censuses. As mentioned in section 4.2, this task will be complicated by the requirement that all data must be on a consistent set of SLA boundaries. Following the extraction of this information, the degree by which the propensities change over time will then be able to be assessed. Analysis would then proceed to determine whether change in propensities at the SLA level can be associated with change in other variables. It is hoped that such relationships will enable propensity changes to be predicted.

At this stage it is unclear as to which indicator variables (if any) would be most suitable. This will become clear once the propensity information had been extracted and a variety of models have been tested. As before, it is important that the indicator data be readily available during the post-censal period at the SLA level or at a level that could easily be converted to SLAs. Since this method relies on the assumption that propensity change is in some way associated with growth it is suspected that many of the indicators used to model SLA populations could be used. It can also be expected that factors such as the change in the proportion of adults and children in an SLA's population could be a good indicator of shifts towards different types of household structure in that SLA.

Assuming that variables can be found which are satisfactorily able to indicate change in propensities, a regression model would be developed which takes account of the change in these variables to predict the change in propensities over the same period. Constraints will have to be placed on the process so that the predicted values for each of the various household types in a particular SLA do not sum to more than 100%.

In addition, to improve the accuracy of the estimates produced via this method it is proposed that SLAs be stratified in some manner, with the possibility of applying a different model to each stratum. A number of different options for stratification could be considered including those based on geographical considerations (SLA grouped by SD, SSD, LFR or by some other locational attribute) and levels of growth (change in estimated resident population or increase in dwellings stock). Another option is to stratify on the basis of "propensity stability". The absolute value of the change in propensity could be summed across the whole of the propensity matrix for each SLA to arrive at a figure which summarises the overall change in propensities during the previous inter-censal period. This may be more meaningful than stratifying based on location.

The points made in section 4.2 concerning the validity of applying trends from previous inter-censal periods is also an issue here. In addition this method assumes that propensity change is in some way related to growth. In stable, established SLAs there is no reason to expect that household propensities will experience significant change from one Census to the next. In contrast, in more rapidly growing or declining areas propensities are likely to be more variable as movements of population impact on the underlying age, sex and family structure of that area's resident population. It is also the case that change in propensities over time is a reflection of changing lifestyle patterns. In recent years average household sizes in Australia, and in much of the rest of the developed world, have been steadily declining, largely a reflection of declines in birth rates and the subsequent impact in average family size. While this methodology does not directly take such "lifestyle changes" into account, it is suspected that the effect of this on propensity changes is more gradual and does not occur so quickly as to significantly influence SLA propensities from one Census to another.

Further work is required to determine the feasibility of this method, particularly in respect of assessment of indicator data and of the statistical soundness of the technique, however it does

initially appear to be quite promising and offers a means of updating Census based propensities in non-Census years.

### 4.6 Application of a "Fast Growth" Propensity

Section 4.1 discussed the option that post-censal household estimates be produced by assuming that the propensities derived from the previous Census remain constant throughout the period. The main problem with this method was that the suspected impact of growth on SLA propensities was not taken into account. Section 4.5 proposed one method by which the impact of growth could be estimated. A compromise between these two alternatives is to assume that propensities remain fairly constant for "stable" SLAs and to apply a "fast growth" propensity to those SLAs (or "parts" of SLAs) which are rapidly growing.

At the time of the Census, SLAs would be stratified according to location, and perhaps other factors such as the proportion of persons aged 15 years and over and persons aged under 15 years residing there. Based on a simple summary measure of growth such as change in estimated resident population or dwelling stock in the previous year, the SLAs in each strata would be classified as "stable" or "rapidly growing". By looking at those SLAs which are rapidly growing, a typical "fast growth" propensity can be produced for each stratum. This would most likely be an "average" of the propensity information of all the rapidly growing SLAs in the stratum.

An additional factor to be considered in relation to household estimates is whether rapid growth is due to high births or high in-migration, since it can be assumed that each would affect household propensities in different ways. For example, high births can be expected to impact on the childhood propensities while high in-migration will have more of an impact on adult propensities. Consequently, for some strata there may be a need to produce two types of "fast growth" propensities, one for births and one for in-migration.

At the time at which household estimates are required, SLAs will be stratified in a similar manner to previously and then reclassified as "stable" or "rapidly growing" based on up to date indicator data. For those SLAs classified as stable the propensities from the previous Census would be applied to current population estimates to produce household estimates for the period in question. For those SLAs classified as rapidly growing the Census based propensities would be updated using the appropriate "fast growth" propensity. In such cases the Census based propensity will continue to apply but the "fast growth" propensity will be applied to what is deemed to be the "extra" or "fast growing" proportion of that SLA's population. Thus the SLA population will be split into "established" and a "new growth" components with a different propensity applied to each.

This method could also be expanded to deal in a similar manner with those SLAs classified as "declining".

## 5. Other Issues to be Considered

### 5.1 SLA Boundary Changes

Over the past five to ten years, there have been several phases of major SLA boundary revision in Australia. The result is that for most States a large proportion of current SLA boundaries are considerably different to what was the case in 1991. SAPU has gained considerable experience in dealing with boundary changes as part of its role to produce SLA population estimates, and it can be expected that most of the methods currently used to deal with boundary changes could be successfully applied in the case of household estimates. Despite this, SLA boundary changes can be expected to have a considerable impact on the production of household estimates and difficulties associated with changes in boundaries should be taken into account when considering the methods described above.

### 5.2 Amount of Detail Included in each Propensity Matrix

The question of how detailed each propensity matrix should be is a balance between the desire for a comprehensive, detailed set of estimates and the need for the level of detail to be practical. The work of Ironmonger and Lloyd-Smith (Ironmonger and Lloyd-Smith, 1992) showed that it is practical to produce highly detailed propensity matrices at the SLA level from Census data, but it is doubtful whether such a level of detail would be feasible when estimating for non-Census years. Household propensities at the part of State level are published for 16 different categories of household (1 - 4+ persons aged 15 years and over, 0 - 3+ persons aged under 15 years). The simplest option for post-censal SLA household estimates would be to produce totals only. It may be viable to use a similar number of categories as in the part of State estimates but the feasibility of this level of detail will not be able to be gauged until work begins on the testing of the proposed method(s).

SAPU is keen to investigate the possibility of incorporating an age component into household propensity matrices. This stems mainly from the current situation where it is possible to determine the number of households in a particular area containing 2 persons aged 15 years and over and 0 persons aged under 15, but there is no information about whether such households contain a young couple or a retired couple. Incorporating a simple age component into the propensity matrices ie 0-15, 16-45, 45+, would help to alleviate this problem. Once again, further testing will reveal if this is in fact feasible. In any case, it may be wise at this stage to follow the conventions set in the part of State method.

### 5.3 Underestimate of Households in the Census

As mentioned previously, it should be remembered that the Census itself may not provide a totally accurate count of the number of households in a particular SLA. Investigations by Ironmonger and Lloyd-Smith (Ironmonger and Lloyd-Smith, 1992) indicated that the 1981 and 1986 Censuses had the tendency to over-estimate the number of households with two persons aged 15 years and over and two persons aged less than 15 years at the expense of households with no persons aged less than 15, and to under-estimate the number of households with one person aged 15 years and over.

## 6. Further Investigation

This paper has discussed a range of methods for estimating post-censal SLA household estimates and has raised many questions. A significant amount of work involving investigation and evaluation would be required before a reliable set of estimates is produced. Options for further investigation could include the following.

1. Obtain SLA household propensity information and data relating to the number and type of households from the 1986, 1991 and 1996 Censuses. Initially this should be restricted to those States with minimal boundary changes during the period.

2. Conduct a detailed analysis of change in SLA propensities over time. Specifically, focus on determining the degree of propensity change over time and whether the variability of propensities appear to be related to the level of growth.

3. Assuming that a link between propensity change and SLA growth can be established, derive a "fast growth" propensity for particular SLA strata (4.6). Test the effectiveness of this technique by producing estimates for 1996 and comparing with the Census.

4. Determine the methodological and practical viability of using indicator data to model change in propensities (4.5). Test the effectiveness of this technique by producing estimates for 1996 and comparing with the Census.

5. Using data from previous inter-censal periods, investigate the use of a regression methodology to directly estimate households.

6. Further investigate the method of updating Census counts of dwellings and households using approvals data and compare with 1996 Census counts.

7. Determine the feasibility of producing current part of State household estimates based on the Labour Force Survey at the LFR level.

## 7. References


Australian Bureau of Statistics (ABS), 1994a : *Provisional Estimates of Households 1986-1994,* Demography Working Paper 94/3. Available on request.

Australian Bureau of Statistics (ABS), 1994b : *Building Statistics Review,* Report prepared by Economic Statistics Group, Australian Bureau of Statistics.

Australian Bureau of Statistics (ABS), 1995 : *Market Research into the Proposed Households and Families Estimates and Forecast Service,* Report prepared by Market Research Group, Business Development Unit, Australian Bureau of Statistics.

Australian Bureau of Statistics (ABS), 1996a : *1996 Census Dictionary, Census of Population and Housing,* Catalogue No. 2901.0. Australian Bureau of Statistics, Australian Government Publishing Service, Canberra.

Australian Bureau of Statistics (ABS), 1996b : *Dwelling Increase Between Censuses: A Comparison of Different Sources,* Paper prepared for 1996 State/Commonwealth Population Workshop. Small Area Population Unit, Australian Bureau of Statistics.

Australian Bureau of Statistics (ABS), 1996c : *Household Estimates Australia 1986, 1991-94,* Catalogue No. 3229.0. Australian Bureau of Statistics, Australian Government Publishing Service, Canberra.

Australian Bureau of Statistics (ABS), 1996d : *Review of Demography Statistics 1996*, Demography Working Paper 96/2. Available on the ABS website at www.abs.gov.au, select Themes/Demography/ABS Demography Working Papers.

Australian Bureau of Statistics (ABS), 1998a : *Future Directions for ABS Household Surveys*, Working Paper prepared by Population Statistics Group, Australian Bureau of Statistics, Canberra.

Australian Bureau of Statistics (ABS), 1998b : *Strategic Directions for Social Statistics 1999/2000*, Report prepared by Population Statistics Group, Australian Bureau of Statistics, Canberra.

Australian Bureau of Statistics (ABS), 1999a : *Review of Functional Classification of Buildings*, Report prepared by Manufacturing and Construction section, Australian Bureau of Statistics.

Australian Bureau of Statistics (ABS), 1999b : *Household Estimates Review*, Report prepared by Methodology Division, Australian Bureau of Statistics, Canberra.

Australian Bureau of Statistics (ABS), 1999c : *Service Population Pilot Study,* Demography Working Paper 99/3. Available on the ABS website at www.abs.gov.au, select Themes/Demography/ABS Demography Working Papers.

Australian Capital Territory Government, 1998 : *Housing Development in Canberra - June 1998 Quarter,* Office of Asset Management, Canberra.

Department of Housing and Urban Development, 1997 : *Metropolitan Adelaide Development Program 1996-2000,* Report prepared by the SA Department of Housing and Urban Development for the interdepartmental Urban Development Coordinating Committee, Government of South Australia.

Ironmonger, D.S. and Lloyd-Smith, C.W., 1992 : *Projections of Households and Household Populations by Household Size Propensities,* in Journal of the Australian Population Association, vol.9 (2)

O'Leary, J., 1998 : *Estimating the Number of Households in Victoria - 30th June 1996,* Working paper prepared for 1998 Australia-NZ Population Workshop, Research Unit, Department of Infrastructure, Victoria.

Smith, S.K. and Cody, S., 1999 : *Evaluating the Housing Unit Method: A Case Study of 1990 Population Estimates in Florida,* Working paper prepared for 1999 Population Estimates Methods Conference, Bureau of Economic and Business Research, University of Florida.

Spicer, K., Diamond, I. and Ni Bhrolchain, M., 1992 : *Simulating the Effect of Demographic Events on the Household Composition,* in Journal of the Australian Population Association, vol.9 (2).

U.S. Bureau of the Census, 1995 : *Methodology for State Housing Unit/Household Estimates.*

Yadava, K. N. S. and Yadava, S.S., 1996 : *Simultaneous Projection of Population and Households: A Comparative Study of Methodologies Used,* in Communications in Statistics Journal, vol.25 (3).