

# Methodological News

A Quarterly Information Bulletin



ABS Methodology and Data Management Division

March 2014

## Articles

Imputing Rent for Australian Owner-Occupied Dwellings: A New Methodology	2
Treatment of Missing Data in Statistical Data Integration	2
Predictive Modelling to Support 2016 Census Data Collection	4
Industrialising Small Area Estimation at the ABS	5
The Road to Better Transport Infrastructure	6
Modelling the Short-Term Dynamics of Unemployment Using the ABS Longitudinal Labour Force Survey File	7
How to Contact Us and Email Subscriber List	8

## Imputing Rent for Australian Owner-Occupied Dwellings: A New Methodology

In 2008, ABS released its first experimental estimates of imputed rent for owner-occupied dwellings and subsidised rentals in the ABS household income statistics (ABS cat. no. 6525.0). Imputed rent for owner occupied dwellings is an important component of the international standards for household income statistics. Its availability has filled a significant data gap in the ABS household income measures and has extended the range of analyses that can be undertaken using household income statistics.

In computing the gross imputed rent estimates, the current methodology employs a hedonic modelling procedure, with a Heckman correction and an extrapolation adjustment for the higher end values. The estimates were derived using data from the Household Income and Expenditure Survey (HIES) 2003-04 and the Survey of Income and Housing (SIH) 2005-06. The established methodology was also used to produce imputed rent estimates in respect of SIH 2007-08 and HIES 2009-10 (see ABS cat. no. 6523.0).

In 2012, the Analytical Services Unit (ASU) in collaboration with the Living Conditions section undertook a feasibility study to examine if the current methodology for estimating gross imputed rent can be improved using administrative data. ASU developed a stratification-based estimation procedure as an alternative to the current methodology using linked data from the 2006 Census and the 2005-07 Valuers-General (VGs) data. The stratification is based on the

Socio-Economic Indexes for Areas (SEIFA), section of state (SOS), number of beds and average house values at the collection districts (CD) level. The new methodology is a more straightforward way of imputing the rental rate for owner-occupied dwellings during census years.

The results of the application of the new methodology are promising and the Living Conditions Section is currently testing it using the latest data, i.e. 2011-12 SIH, 2011 Census and 2010-12 VGs, at the mesh block level.

### Further Information

For more information, please contact Franklin Soriano (02 6252 5933, [franklin.soriano@abs.gov.au](mailto:franklin.soriano@abs.gov.au))

## Treatment of Missing Data in Statistical Data Integration

Statistical data integration of micro level data from different administrative and/or survey sources is an emerging priority for the ABS and wider National Statistical Service (NSS), as a means to investigate more complex and expanded policy and research questions that would not be possible using separate unintegrated data sources. The linking of micro level data from these source files can often lead to records with missing data in the linked files which needs to be addressed in order to enable reliable analyses to be performed on these linked files.

The two standard methods for the treatment of missing data are weighting and imputation. Weighting methods are the current preferred approach to the treatment of missing data in

statistical data integration, since weighting automatically maintains the relationships between the data items within the various source files, the theoretical framework for the calculation of variances is much more advanced and developing a weighting system usually requires fewer resources to build and less time to execute.

A two-phase approach has been proposed to handle weighting for the missing data, with the sampling mechanism as the first phase and the linking mechanism as the second phase. In some circumstances one of the datasets to be linked will conceptually represent the population of interest completely. In this situation the sampling mechanism in the first phase equates to a complete enumeration of the population.

The proposed weighting method is a two-step linking propensity calibration estimator. In the first step of the two-step calibration procedure, the estimated linking probabilities are obtained from fitting a logistic linking propensity model. The inverse of these estimated linking probabilities are used as the intermediate weights that feed into the second step of the calibration weighting procedure. In the second step of the two-step calibration procedures, the intermediate weights are calibrated to known population auxiliary totals.

The choice of auxiliary variables to be included in the first step of the two-step linking propensity calibration procedure should be based on those auxiliary variables which help explain the probability of linking. The choice of auxiliary variables to be included in the second step of the two-step linking propensity calibration procedure

should be based on those auxiliary variables which can improve the accuracy of the estimates (i.e. those auxiliary variables highly correlated with the survey variables) and/or those auxiliary variables which are needed to ensure consistency with known population totals.

This particular two-step weighting procedure was applied to the data linked between the 2011 Census and the Department of Immigration and Border Protection's Settlement Data Base (SDB). A paper on the proposed two-step linking propensity calibration procedure was recently presented to the Methodology Advisory Committee.

Further investigations are being conducted into the interplay between setting linking cut-off values and weighting linked records. The Bronze Low linked files (which uses a low threshold value to accept or reject links) are currently used, rather than the Bronze High linked files (which uses a high threshold value to accept or reject links). However, a major drawback of using the Bronze Low linked files is that there are more incorrectly linked records (i.e. accepting links to similar persons or businesses rather than the same person or business). The Bronze Low linked files also require substantially more resources to produce than the Bronze High linked files.

## Further Information

For more information, please contact John Preston (07 3222 6229, [john.preston@abs.gov.au](mailto:john.preston@abs.gov.au))

## Predictive Modelling to Support 2016 Census Data Collection

The Operations Research and Process Improvement (ORPI) section is developing a predictive modelling framework for the 2016 Australian Census of Population and Housing. The 2016 Census is introducing several significant changes to the data collection operation which aim to provide a more efficient process while maintaining the coverage of the Census. This new enumeration model removes the need for Census field officers to visit every dwelling in Australia. In a large proportion of areas across Australia, secure access codes for the online form will be mailed to households in the first instance. These will be followed by reminders sent by mail. Householders wishing to complete their Census form on paper will have a form mailed to them and can mail back their completed form. In the remaining areas of Australia, a more traditional delivery approach using Census field officers will be retained.

The primary objective of the predictive modelling framework is to predict the amount of field staff resources needed for the follow-up phase of collection. In the follow-up phase of collection field officers will travel to households which are still to respond. The predicted proportion of households requiring follow-up in each fine geographic region is a key input. In addition to modelling the proportion which respond prior to follow-up, the number of return follow-up visits and field officer travel time are being modelled to predict field staff requirements. During the enumeration period the predictions made during planning will be updated using information on the collection process. These

updated predictions will ensure the allocation of staff resources across areas is responsive to collection progress.

The extensive changes to the enumeration model presents a significant challenge to predicting respondent behaviour. The modelling framework combines data from the 2011 Census and from field tests of the new enumeration model. The rates of internet response and follow-up effort required in different regions in the 2011 Census will be used to identify demographic and geographic characteristics associated with response prior to follow-up. These factors will be included in a model predicting the relative level of response prior to follow-up across regions. Data from the tests will inform assumptions about how respondent behaviour will be different under the new enumeration model compared with 2011. For example, the test data will inform assumptions about how various prompts impact the amount of response received during each enumeration phase.

The components of the modelling framework have already been used to inform decisions about the size of the workload areas for the 2016 Census. The modelling framework and underlying assumptions will be refined following analysis of the August 2014 Major Test, and the refined model will guide the recruitment strategy for the 2016 Census.

### Further Information

For more information, please contact Julian Whiting (08 8237 7362, [Julian.whiting@abs.gov.au](mailto:Julian.whiting@abs.gov.au))

## Industrialising Small Area Estimation at the ABS

ABS surveys are designed to achieve accurate estimates using design-based methods at broad levels, for example, state and national levels in household surveys, or industry divisions in business surveys. At finer levels of geography, or detailed industry classifications, there is usually insufficient sample for reliable estimation. However there is a strong government demand for these small area statistics for planning, policy development and resource allocation. Small area estimation methods are one way in which this demand might be addressed.

In small sample situations, model-based methods are often able to provide estimates of acceptable accuracy. Statistical models are used to borrow information from other areas or time points via auxiliary data related to the target variable. This auxiliary data may include Census or external administrative data sources. In the ABS, model-based, small area estimates have been produced for a few external consultancies and been published as a 'complementary' product in a couple of cases. Typically the methods and processes involved are time-consuming and bespoke, but lead to high quality estimates.

It is envisaged that in the future small area estimates may be released as official statistics. A number of challenges arise, especially the need for systems and processes that are 'mass customisable'. This means the ability to output a large number of estimates quickly, as well as the flexibility to implement different methods for different situations. To this end we have initiated research to examine a number of alternative

model-based methods. A focus has been on weight-based approaches as these are particularly suited to production-efficiency. These include the model-based direct estimator of Chandra and Chambers<sup>1</sup> and the re-weighting method described by Schirm and Zaslavsky<sup>2</sup>. Work is also progressing to understand and map the business process and business requirements for a future small area system.

Other challenges to address include achieving a right balance between efficient 'black box' processes and the need for quality assurance and expert technical input. Good knowledge management built into intelligent systems will be important here. The ability to produce reliable, production-efficient mean square error estimators for small area estimates is also a key area of research.

The task of industrialising small area estimation is a challenge, in particular attempting to simplify and streamline a process that is by nature very complex is difficult. Work to date indicates good prospects of success.

<sup>1</sup> Chandra, H and Chambers, R. (2009). Multipurpose weighting for small area estimation. *Journal of Official Statistics*, **25**, 1-18.

<sup>2</sup> Schirm, Allen L. and Alan M. Zaslavsky. Reweighting Households to Develop Microsimulation Estimates for States. *1997 Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1997.

### Further Information

For more information, please contact Sean Buttsworth (02 6252 5174, [sean.buttsworth@abs.gov.au](mailto:sean.buttsworth@abs.gov.au)) or Peter Radisich (02 6252 6731, [peter.radisich@abs.gov.au](mailto:peter.radisich@abs.gov.au)).

## The Road to Better Transport Infrastructure

Detailed information on the movement of freight by trucks has not been collected by the ABS since 2000/01. Following a feasibility study in 2012 into options for the ABS to once again collect detailed road freight information, the ABS received confirmation of external funding for a Freight Movement Survey (FMS) in January 2013. While administrative data was investigated during the feasibility study, the agreed proposal involved a sample survey collecting information from rigid and articulated trucks only. The road data obtained will complement that already available for rail, sea and air. The FMS 2013/14 will fill a critical statistical information gap and be used in decision making regarding transport infrastructure.

The FMS 2013/14 went into the field at the start of November 2013 and will collect information from registered owners relating to trip details including origins, destinations, distance travelled, load weight and commodity carried. Vehicles are selected with the respondents being asked to fill in a diary for a selected week of the reference period. The reference periods will cover a 12 month period ending October 2014.

The FMS will be run in conjunction with the Survey of Motor Vehicle Use (SMVU) 2013/14 which will be conducted over the same reference period. The SMVU collects information on the usage of vehicles over a specific four month period within the reference period. Half of the vehicles selected in the FMS will also be selected in the SMVU and will subsequently have data pertaining to the trips for a week as well as

the usage of the vehicle over a four month period.

Respondent 'fatigue' was an issue with the previous FMS collection with trips reported in the latter half of the diary consistently being less than that reported in the first half of the diary period. The risk of this occurring for the current FMS have been mitigated by reducing the length of the diary period from two weeks down to one as well as the collecting of odometer readings from vehicles to ascertain whether the reporting in the diary is representative of the usage of the vehicle over a longer period.

The sample design of the FMS 2013/14 aimed to maximise the quality of the main data items of total distance travelled and tonnes kilometres (tonnes carried by distance travelled) at the national and state of registration levels. This should allow for acceptable quality for the major routes travelled across the country.

Investigations have been undertaken as to the best methodology to combine the FMS and SMVU components of the collections to produce the highest quality, coherent estimates from the information available.

### Further Information

For more information, please contact Brett Frazer (07 3222 6028, [brett.frazer@abs.gov.au](mailto:brett.frazer@abs.gov.au))

## Modelling the Short-Term Dynamics of Unemployment Using the ABS Longitudinal Labour Force Survey File

The Analytical Services Unit is currently undertaking an analysis using the newly constructed ABS Longitudinal Labour Force Survey (LLFS) file. The focus is on examining the transition from unemployment for Australians aged 20 years or older, over a period of three years, from the beginning of 2008 to the end of 2010. By including more than 1.8 million records from around 150,000 households observed over a period of up to eight consecutive months, the file is well suited for analysing short-term labour market dynamics.

The analysis is divided into two parts. The first is focused on non-parametric techniques to examine the number of periods spent in unemployment and includes raw hazard and survival functions, as well as life tables. The second part builds on the first and examines the transitions from unemployment using a discrete duration modelling approach. To account for the observed as well as unobserved heterogeneity, both the ordinary logit as well as the random effects logit models are implemented. As there are different exit states, the analysis applies the competing-risks framework to separately examine the transition from unemployment into the four exit states: any type of employment (defined as either full-time or part-time), full-time, part-time, or not in the labour force.

Amongst others, the preliminary findings indicate that the longer the unemployment spell, the less likely it is for people to exit to

employment, and the less likely it is for them to leave the labour force (contrary to the discouraged job seeker effect). This is an interesting result that warrants further investigation.

Research analysts interested in getting more information about the LLFS can refer to *Microdata: Longitudinal Labour Force, Australia, 2008-10* (ABS cat. no. 6602.0).

### Further Information

For more information, please contact Cristian Rotaru (02 6252 5098, [cristian.rotaru@abs.gov.au](mailto:cristian.rotaru@abs.gov.au))

## How to Contact Us and Email Subscriber List

Methodological News features articles and developments in relation to methodology work done within the ABS Methodology and Data Management Division. By its nature, the work of the Division brings it into contact with virtually every other area of the ABS.

Because of this, the newsletter is a way of letting all areas of the ABS know of some of the issues we are working on and help information flow. We hope the Methodological Newsletter is useful and we welcome comments.

If you would like to be added to or removed from our electronic mailing list, please contact:

Valentin M. Valdez  
Methodology & Data Management Division  
Australian Bureau of Statistics  
Locked Bag No. 10  
BELCONNEN ACT 2617

Tel: (02) 6252 7037

Email: [methodology@abs.gov.au](mailto:methodology@abs.gov.au)

The [ABS Privacy Policy](#) outlines how the ABS will handle any personal information that you provide to us.