Methodological News
A Quarterly Information Bulletin

ABS Methodology and Data Management Division

Australian Bureau of Statistics

Articles

www.abs.gov.au

Update from the Data Linking Team

The data linking team in the Analytical Services Unit (ASU) provide methodological support to a number of internal clients conducting data linking projects. These projects include the linking of a 5% sample of 2006 Census records to the 2011 Census to form the first wave of the Statistical Longitudinal Census Dataset (SLCD), the linking of deaths registrations to the 2011 Census to investigate Indigenous mortality, the linking of school enrolments data to the 2011 Census and the linking of the Migrants Settlements Database to the 2011 Census.

Research performed by data linking team members has resulted in two significant methodological improvements that have been implemented for the 2011 linking projects. The first of these is the use of the Expectation-Maximisation (EM) Algorithm to estimate the parameters of the models used for each linking run. The second is the use of sampling methods to determine the weight cut-offs for acceptance and rejections of record pairs as links. The data linking team believes that these methodological improvements will result in increased accuracy of the linked data compared to similar linking projects which were conducted with the 2006 Census data.

Linking projects which use name and address information as linking variables are referred to as Gold Standard projects. All name and address information from the Census is destroyed at the end of the Census processing period. For this reason, these projects need to be completed before the Data Processing Centre (DPC) shuts down in early December. The data linking team and our client areas are all currently very busy working to this deadline. There are also linking projects which do not use name or address information. These are known as Bronze Standard, and will continue after the DPC shuts down. Once all the linking has been performed, the data linking team will move on to the analysis of the linked data sets.

Further Information

For more information, please contact Phillip Gould (02 6252 5315, phillip.gould@abs.gov.au) or Carrie Samuels (02 6252 6186, carrie.samuels@abs.gov.au)

Modelling Industry Water Consumption in Non-Survey Years

The ABS produces data on the supply and use of water by industry and households in Australia through its annual Water Accounts Australia publication (ABS Cat. no. 4610.0). Annual surveys are conducted for the Water Supply and Agriculture industries, however the key survey on water use by all other industries (Energy Water and Environment Survey - EWES) is conducted every three years only. This project developed a modelling method for estimating water use by industry, state and water source for the non-survey years.

This project used the 2008-09 EWES data which was linked to the 2008-09 Business Activity Statement Unit Record Estimates (BURE) data to derive estimates of water use for 2009-10 and 2010-11. The modelling

methodology was similar to that used for the energy consumption modelling project undertaken by the ABS previously. The model used was a log-linear model which modelled water consumption at the firm level as a function of firm turnover and the industry it belongs to (i.e., industry dummy).

The choice of the explanatory variables was dictated by the availability of the variables in BURE which would be the main data source for producing estimates in the non-survey years. Additional steps as part of the estimation process involved adjustment for log transformation bias, adjustment for zero-energy reporting by firms and benchmarking of the adjusted model estimates to EWES 2008-09 estimates to align for scope differences between EWES and BURE. The final estimates for 2009-10 and 2010-11 were in line with trend growth in turnover from 2008-09.

Further Information
For more information, please contact Kay Cao (07 3222 6467, kay.cao@abs.gov.au)

Automated Field Interviewer Workload Allocation Tools in Action

Population Survey Operations section (PSO) maintain a workforce of field interviewers to approach Australian householders to conduct surveys. Operating on a 6 week - 3 month in advance schedule, PSO release allocations to each interviewer of upcoming survey specific "workloads" of household locations for personal and/or telephone interviews. Household surveys can either be regular

(Monthly Population Survey) or irregular one-off and lasting for one or several months (Special Social Surveys). The survey specific workloads are expected to be conducted in particular 1, 2 or 3 week periods.

Interviewer workload allocation is subject to constraints such as the number of available interviewers, dates when interviewers are available for duty, geographical location of interviewers and workloads, training for particular surveys, and the enumeration period within which a survey collection must occur. In late 2011 operations research consultant Olena Gavriliouk produced and released the Automated Workload Allocation Tool (AWAT) for field trial with PSO. This release included an interviewer travel model and predicted interviewer weekly work hours that have been well used by PSO to finalise workload allocation decisions.

In 2012 Olena has further expanded the sophistication of AWAT based on the live field experience, to deal with the iterated advance allocations that PSO needs to do survey and interviewer management. The AWAT project has also involved the provision of database based tools providing (i) geospatial views of interviewer and workload locations with pop-up information summaries, (ii) cross tabulated views of weekly interviewer assignments with predicted work hours and monthly summaries, and (iii) editing tools to immediately enter field staff changes for prompt reuse of AWAT.

AWAT is currently being used directly by PSO staff to produce first-cut weekly field assignments for interviewers for many months in advance and to provide information on interviewer capacity planning.

AWAT is producing efficient allocations and saves many staff-hours which in the past were devoted to the manual process of allocation.

Further Information

For more information, please contact Olena Gavriliouk (03 9615 7413, olena.gavriliouk@abs.gov.au) or John Martin (02 6252 7006, john.martin@abs.gov.au)

Building a Predictive Model for the Geographic Spread of E-form Take-up

Given the widespread use of the Internet by Australian households, the ABS is moving towards providing an electronic form (e-form) as the preferred means for households to respond to ABS surveys and to the Population Census. In designing such a collection, fewer resources (such as interviews or Census collectors) can be assigned to geographic areas that are likely to have higher e-form take-up, thus reducing costs.

In this project, the geographic spread of e-form take-up in the 2011 Census has been modelled against known characteristics of small areas (Census workloads known as CLWs). Besides indicators of what state each CLW is in and how remote each CLW is, the predictors in our model included proportions of types of people (Aboriginal and Torres Strait Islanders, non-English speakers, school completers, women), proportions of people that fell into pre-defined income and age categories, and proportions of different types of dwellings (people living in buildings

with 3 or more floors, private dwellings, different dwelling types, lone person households, older households) within each area.

Our final model contained 36 variables (plus an intercept), and we used the same set of variables to predict such things as e-form take-up for the entire population of a CLW, as well as e-form take-up just for people who have (or don't have) broadband access.

This model is expected to be useful for planning a wide range of collections, even those with different e-form take-up rates than the 2011 Census. The model gives a way to distribute a specified high-level take-up rate across detailed geographic areas. Predictions from the model will be produced for Mesh Blocks (fine areas usually comprising about 50 dwellings) so that they can be aggregated to geographic areas used in planning surveys and the next Census.

Further Information

For more information, please contact Philip Bell (08 8237 7304, philip.bell@abs.gov.au) or Daniel Fearnley (02 6252 5446, daniel.fearnley@abs.gov.au)

Land Management Practices Survey Design

This year the Rural, Environment and Agricultural Statistics Branch (REASB) is conducting the Land Management Practices Survey (LAMPS). This is a user-funded survey coming under the umbrella of the Federal Government's Carbon Farming Initiative. It will collect information on agricultural practices that relate to CO_2

abatement. The survey is scheduled to be repeated in 2014 and 2016. Outputs from the present survey will assist in defining what is current "common practice" as well as providing benchmarks for measuring changes in farming practices in response to government climate change initiatives.

The requirements of this survey differ from those of the annual Agricultural Commodity / Agricultural Resource Management survey in significant ways and REASB considered that a separate survey was needed in order to satisfy these requirements. These differences also made the design a challenge. One of the differences is that many of the data items collected are new and so there is little historical data available for key items of interest. The design was therefore done using nine items from among the data regularly collected in agricultural surveys which were considered close "proxies" for the items of interest.

A second challenging feature was the output requirements. Agricultural survey outputs are generally produced for nation, state and sub-state national resource management regions. For LAMPS, outputs are required at national and state levels and for agro-climate regions, with estimates for the latter being a priority. These regions pose a difficulty for stratification in that they may cross state boundaries and in some cases consist of non-contiguous sub-regions. Consequently, the geography component of the stratification was chosen to be state x agro-climate region so that estimates of acceptable quality could be produced for both of these geographies.

At each geographic level outputs were required for a mixture of ANZSIC classes and groups of classes, 26 in all. Combined with the number of design variables, this led to a number of constraints so large (6,318) that it was impossible to run the allocation code for the entire set. The first task was therefore to try to reduce this constraint set. This was done by eliminating constraints for which there was insufficient historical data. In addition constraints for ANZSIC class x agro-climate region cells were eliminated if the expected cell estimate was only a small proportion of the estimate at higher output levels, e.g. all of the Agriculture division for the particular region, or nationally for the particular ANZSIC class.   The final constraint set had 1,380 constraints which is comparable with the number in the annual Agricultural Commodity / Agricultural Resource Management survey.

The survey is currently in the field and methodology work is now focussed on supporting estimation for this survey and this year's Agricultural Resource Management survey (ARMS). These two surveys cover the same reference period and also have in common a number of data items which are conceptually the same. It was considered necessary to have these items in the LAMPS survey because they were intrinsic to the objectives of this survey and also because estimates of these items needed to be consistent with those of other items collected. However, this presents REASB with the challenge of meeting user expectations that two estimates of the same quantity for the same reference period should be the same even if produced from different surveys. One way of bringing estimates from the two surveys into closer alignment is to calibrate estimates from one against those produced by the other. Since LAMPS has a

significantly larger sample size (50,000) than ARMS (37,000) it is planned to calibrate ARMS estimates of business counts and total area of holding with those from LAMPS at the state level. This will be supplemented by other strategies still under discussion.

Further Information

For more information, please contact Jos Beunen (02 9268 4647, jos.beunen@abs.gov.au)

Version 2.9 of the SEASABS Suite Released

The Time Series Analysis section (TSA) successfully delivered the release of the latest version of the ABS seasonal adjustment software SEASABS Suite on 19 October 2012. The main new features and enhancements as an analysis and output tool are summarised as follows:

- Implementation of the temporal aggregation method to calculate seasonally adjusted quarterly estimates so as to increase consistency between equivalent monthly and quarterly seasonally adjusted estimates, ensure additivity of monthly and quarterly seasonally adjusted estimates, and to improve the efficiency of the ABS's seasonal adjustment analysis.
- A functionality that estimates and applies seasonal break corrections using regARIMA modelling automatically.

- The capacity to convert prior factor corrections when the decomposition model changes.
- An introduction of a new analytic tool to facilitate research analysis and seasonal adjustment quality checking of time series at a group level.
- Additional quality checks to flag any generated negative seasonally adjusted and/or trend values from a series of non-negative original values.
- A 'Before/After prior correction comparison' tool in SEASABS when considering whether to apply a prior correction to a series.
- An enhanced functionality to allow for multiple seasonal/non-seasonal sub-spans to deal with changing seasonality status within a time series.
- New Hot-keys for commonly used SEASABS commands.
- Improved graphing of trend breaks so as to show all breaks being applied.

These new functionalities streamline TSA practices and increase productivity as well as benefit SEASABS users by improved quality management. A standalone version of the latest release of SEASABS is available to external organisations under licence.

Further Information

For more information, please contact Lujuan Chen (02 6252 6342, lujuan.chen@abs.gov.au) or Philip Carruthers (02 6252 3057, philip.carruthers@abs.gov.au)

Time Series Methods Course Revision

Time Series training for methodologists has been substantially revised this year. The course aims to provide ABS methodologists with both theoretical and practical experience with seasonal time series, with particular emphasis on the real world issues and changes that require methodologists' attention and intervention. There were major updates to the course notes, and the course was delivered in a two week intensive format. The course consists of a set of lectures, course notes, tutorials, assignments and an exam.

The course covers: Time series decomposition; interpreting time series; Fourier analysis; filters; gain functions; prior corrections, including moving holidays and trading day corrections; ARIMA models; regression-ARIMA; and backcasting (to handle classification changes).

Some more new sections will be written and the course notes as a whole reviewed before re-releasing the notes on the ABS website during 2013.

Further Information
For more information, please contact Rachel Barker (02 6252 6183, rachel.barker@abs.gov.au) or Theresa Nunan (02 6252 6382, theresa.nunan@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