Methodological News

A Quarterly Information Bulletin

Australian Bureau of Statistics

ABS Methodology and Data Management Division

June 2013

Articles

Future seasonal adjustment infrastructure

Keeping seasonal adjustment infrastructure efficient, methodologically modern and quality assured is always a focus for the Time Series Analysis section (TSA), but more so in light of the current emphasis on building systems that can be used across many statistical production areas of the ABS, handle future metadata requirements and easily interact with other systems. An essential part of any roadmap to new seasonal adjustment infrastructure is ensuring that future inputs, seasonal adjustment 'engines' and outputs are handled efficiently and are adaptable. Further, evaluations of what existing seasonal adjustment infrastructure still has to offer versus other alternatives, be they pure 'off the shelf products' or 'best of both world cross breeds', need to be conducted. This necessitates a more flexible 'sand pit' environment for external software feasibility studies. A series of Methodological News articles over coming issues will outline key aspects of the future seasonal adjustment infrastructure plan.

This article will focus briefly on current time series metadata, referred to as 'series knowledge'. Series knowledge contains, besides the obvious start/end date and periodicity variables, a detailed record of analysis parameters required for appropriate seasonal adjustment and trending, aggregation structures to achieve seasonally adjusted and trend Australian level series from their respective states and industries (for example) and uniquely, existing software SEASABS has the ability to store past analyses in the form of archives.

Exploring such metadata, determining what is no longer relevant, determining what will be needed in the future and successfully translating it from text file format to an international standard like SDMX/DDI are all important first steps in ensuring greater compatibility between seasonal adjustment infrastructure and other systems. Standard time series metadata allows seamless connectability between systems be they processing based (e.g. FAME or SAS) or analytical tools (e.g. Demetra, Win X13 or R) and this is a key aspect of streamlining future time series processes. New metadata is envisaged to include greater representation of related time series across collections, improved quality assessment parameters, associated bonafide quality gate flags (which can be also monitored over time) and parameters from new knowledge / business rules to successfully govern envisaged batch capability (seasonal adjustment of a group of time series).

A series knowledge 'translator' (from existing text file format to SDMX/DDI) is expected to evolve as part of the move towards a standard sand pit evaluation environment through current collaboration between TSA and the ABS data management and technical application areas. Such a sandpit is to facilitate evaluation of not just time series software, but also other software that other parts of the Division wish to bring in to the ABS environment, aiming to significantly streamline this process within the ABS.

Further Information

For more information, please contact Lisa Apted (03 6222 5932, lisa.apted@abs.gov.au)

Evaluating Demetra +

The Time Series Analysis section (TSA) has recently completed a preliminary evaluation of Demetra+, seasonal adjustment software used and provided by the European Statistical System and free to download from the Internet. Demetra+ is one of the 'off the shelf' time series products thought to have a place in future ABS time series infrastructure. The preliminary evaluation was undertaken to assess initial broad suitability for incorporation into ABS time series functionality, and focussed on:

- ensuring the safety of the software being brought into ABS computing environment
- identification of problems with installation or analysis
- identification of special requirements with installation and user support
- assessing that the required functionality to perform seasonal adjustment comparable to the existing ABS time series system exists
- checking the stated capability of Demetra+ to perform seasonal adjustments on a whole set of time series
- confirming all available methods for importing data and metadata into Demetra+
- exploring methods and formats when exporting output out of Demetra+.

Findings from the preliminary evaluation have been promising. Demetra+ incorporates the two most popular seasonal adjustment methods: X-12-ARIMA from the United States Census Bureau and TRAMO/SEATS from the Bank of Spain. It provides TSA with an opportunity to explore the additional option of using a model based seasonal adjustment method in addition to its current moving average based approach when decomposing a time series. It also facilitates the comparability of the results and diagnostics from both methods.

During the remainder of the year, TSA staff are planning to perform further feasibility studies as well as an in-depth investigation into the utility of this tool for seasonal adjustment of ABS time series. Comparisons with existing seasonal adjustment infrastructure will also be made to identify any gaps or other differences in output and diagnostics that might arise. The focus is not only on Demetra+ functionality, but also how it will mesh with existing and emerging infrastructure.

Further Information

For more information, please contact Lujuan Chen (02 6252 6342 lujuan.chen@abs.gov.au)

Improving Survey Efficiency at the ABS with Responsive Design – A Simulation Study

The cost and effort of maintaining high response rates in ABS surveys has been increasing in recent years. How can the ABS adapt in a changed environment where we are observing declining response rates and increasing enumeration costs?

Traditionally, response rates are used as a measure of survey quality and are used to achieve precision targets. However, a high response rate does not guarantee low non-response bias. Responsive design can improve survey efficiency by:

- providing us with an informed understanding of the trade-off between survey cost and quality, using additional indicators of survey quality
- enabling us to adapt survey data collection strategies in response to improved survey performance intelligence.

Recent research (Shlomo et al 2012) has investigated two survey quality indicators to complement response rates: representativity indicators (or R-indicators) and maximal absolute bias, an upper limit to the survey non-response bias. These indicators require the response probabilities of all sample units, which are predicted using a response model. Non-response is assumed to be missing at random with respect to the survey data items, conditional on the covariates used in the response model.

We conducted a simulation study to explore the effects of responsive design on an ABS household survey. The simulation study investigated an alternative non-response follow-up targeting strategy and its effect on response rates, the R-indicator and maximal absolute bias. Under simulation conditions, we can calculate survey non-response bias and root mean square error (RMSE) for a number of survey items and use them as ultimate quality measures for assessment.

Our targeting strategy consisted of:

- enumerating all households for the first two calls or interviewer waves
- restricting interviewer follow-up to half the remaining non-responding households from wave three up to a maximum of ten waves
- exploring three follow-up strategies: targeting the easy respondents, the difficult respondents or a random half.

All targeting strategies led to reduced survey costs and lower response rates. However, targeting follow-up to households that were less likely to respond resulted in the lowest survey bias and response rate. Overall, targeting difficult respondents resulted in reduced RMSE compared with the other two groups. The maximal absolute bias was informative for the difficult and random respondents but too high for the easy respondents.

Further work will incorporate Census data as covariates to improve the response model and use follow-up strategies based on real survey procedures.

References

Shlomo, N, Skinner, C & Schouten, B. (2012) 'Estimation of an indicator of the representativeness of survey response', *Journal of Statistical Planning and Inference*, vol. 142, pp. 201-211.

Further Information

For more information, please contact Lan Kelly (08 8237 7643, lan.kelly@abs.gov.au)

Targeting non-response follow-up in a household survey – a geographic approach

Visiting households to obtain survey responses is very expensive, so it is appropriate to apply follow-up effort where it will have the most benefit. In a Special Social Survey (a large survey on a particular social topic), data collection extends over a long period, so it may be possible to use data from early-responding households to target further follow-up.

The responsive design team in Statistical Services Branch has considered and extended the existing methods for assessing bias reduction due to follow-up. One finding is that follow-up reduces bias when it improves the response rate for those types of dwellings or persons that would otherwise be under-represented in the sample. Conversely, increasing response among the well-represented types could actually increase bias. The team has proposed a strategy in which a basic level of follow-up is applied to all areas, but extended follow-up is targeted to avoid areas of types that are already well-represented after basic follow-up and that have few non-responding dwellings.

This approach is being tested using a case study approach based on data from the 2007-08 National Health Survey. In this study the impact on estimates of reducing the number of contact attempts or callbacks in various ways has been evaluated. The project demonstrates that there is a relationship between some survey estimates and the numbers of callbacks. It also shows that some characteristics of the areas being interviewed, as measured in a recent Census, are related to response in a way that is not adjusted for adequately by the survey weighting.

This leads to specific targeting strategies that favour extended follow-up in the low-response types of areas. The case study evaluates a number of such strategies and their effect on key estimates. The objective is to define a strategy that provides a sample that is similarly representative to that which is obtained by extended follow-up, or better, with a lower cost.

The next step for this work is to apply it to live surveys. This will involve working with survey areas and the population survey operations area to define appropriate follow-up scenarios for basic and extended follow-up, and to decide which scenario should be applied in each sampled area of the survey. We will also need to put in place monitoring to ensure that the resulting sample is representative. The aim will be to make major cost savings by reducing interviewer effort while producing valid and reliable estimates.

Further Information

For more information, please contact Philip Bell (08 8237 7304, philip.bell@abs.gov.au)

Family Tree Method

Data from the Australian Census of Population and Housing are typically supplied to users in the form of flat files wherein each record pertains to a single individual. Dwelling, family and person identifiers are provided so that users may reorganise the data hierarchically, and this arrangement appears to satisfy the requirements of most users. However, many important research questions require more detailed information on family interrelationships. Examples include the analysis of fertility rates, social mobility across generations, patterns of intermarriage and ethnic identification, and social inclusion/exclusion of households.

On every Australian Census household form, a reference person is identified and other household respondents are asked to define their relationship to this person. Based upon this reported direct relationship and responses to other Census questions, it is often possible to deduce information on the relationships between other family household occupants – in particular spousal, parent-child and grandparent-grandchild relationships. An advantage of the Australian Census is that the coding of relationship in household is based on the nuclear family concept, i.e. each separately identified couple relationship and/or parent-child relationship forms the basis of a family.

Due to confidentiality reasons, Australian data users are only granted access to confidentialised person-level sample Census files, e.g. 1% sample file via CD-ROM and 5% only through Remote Access Data Laboratory (RADL). For analyses requiring family interrelationships, non-ABS users have to make a special requests to the ABS for producing customised data, which requires significant resources for complex programming.

Drawing on international experience from the Integrated Public Use Microdata Series (IPUMS) project, Analytical Services Unit have been developing an extension of the 'own-children' method to identify family interrelationships. With this method, the Census de-identified data can be re-organised in a manner such that each record is appended with demographic and socioeconomic information from parents, grandparents, and/or spouse, where applicable. This method, known as the 'Family Tree' method, will allow users to utilise family interrelationships in their analysis. The re-organised data can be easily handled using a standard statistical package. The name of this method should not be confused with genealogy, which is an entirely different thing.

Analytical Services have recently used the method to look at intermarriage (inter-partnering), mixed parentage and inter-generational transmission of Indigenous status to help explain intercensal changes in the count of Aboriginal and Torres Strait Islander people. They are also looking into how it can be used to improve match rates in data linking exercises. Children have limited information to differentiate each other, while parents' information can be used as additional linking variables to distinguish between children with very similar characteristics.

Further Information

For more information, please contact Guangyu Zhang (02 6252 5791, guangyu.zhang@abs.gov.au)

A more rigorous approach to case-study methodology in Annual Integrated Collections

The ABS' Annual Integrated Collections program (AIC) collects data used in the production of Australia's National Accounts. The primary form of data collection is the Economic Activity Survey (EAS), an annual probability survey that samples from most industries in Australia via self-completed mail-out/mail-back methods, augmented by secondary 'flexible' surveys.

However, these surveys have a long development time. Additionally, their broad scope and self-enumeration make them unsuitable for collecting fine-level commodities data, e.g. a breakdown between 'Computer systems, hardware and software design and development services' and 'Computer support services'.

To meet this need AIC sometimes uses a 'case study' approach, selecting approximately 20 of the largest businesses within an industry and collecting data via personal interview. If a case study can achieve near-total coverage of the whole industry (e.g. the top 20 petroleum and coal manufacturers account for 98.5% of industry turnover and 75% of employment) then this is likely to produce good results.

It's less obvious whether case studies are suitable for industries which aren't dominated by the top 20. Personal interviews and small sample size allow for high-quality data collection, minimising respondent error. But the selection method risks introducing bias when the difference between large and small industries goes beyond just scale – e.g. large businesses may use capital-intensive production methods that give higher productivity per employee, in which case pro-rating their results to small businesses will give misleading results.

To provide guidance on this issue, Business Survey Methodology section used a simulation approach based on recent EAS survey data. EAS estimates are based on a weighted sample covering both large and small units. These were compared to estimates derived by treating the top 20 units as a mock case study and pro-rating them to the whole industry by turnover (data available for all businesses).

These simulations showed that the accuracy of case studies depends not only on the level of coverage, but also on the industry and on the variable of interest.

For example, a simulated case study of the employment placement/recruitment/labour supply services industry gives coverage of 27% of total industry turnover. Comparing to AIC estimates indicates that case-study error is approximately 12% for expenditure on labour costs/wages and 5% for total income. (In this industry, total income is very closely correlated with turnover, so pro-rating case study data by turnover gives excellent results.)

By contrast, a case study of the computer systems design and related services industry achieves coverage of 36% but the errors for these same items are 15% and 9% respectively. Even though coverage is higher, results are poorer.

Simulations in one of the mining industry groups showed that case studies on the same group of units produced errors of 11% for fuel tax reimbursement but 35% for expenditure on employment agencies, perhaps reflecting that miners have similar patterns of fuel usage but may have very different recruitment strategies.

The methods developed here can be used to assess the viability of future case studies, providing EAS survey data contains a suitable proxy for the variables of interest.

Further Information

For more information, please contact Geoffrey Brent (03 9615 7685, geoffrey.brent@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