Methodological News
A Quarterly Information Bulletin
Australian Bureau of Statistics
ABS Methodology and Data Management Division
March 2012

Articles

An Update on the Methodologies for Estimating Energy Consumption

The new data source for the Australian Energy Statistics – the National Greenhouse and Energy Reporting System (NGER) – does not require businesses using less than 200 terajoules to report their energy consumption. This results in a data gap in the estimate of total industry energy consumption by businesses. The Analytical Services Branch is currently working with the ABS Centre of Environment Statistics on a project aiming to develop a methodology for estimating statistics for this under-coverage.

The energy modelling literature review identified several possible modelling approaches. Of these, the most appropriate approach based on the available data, was to estimate business energy consumption using the static single equation approach with turnover and industry dummy variables (at 3-digit ANZSIC06 level) as the main explanatory variables. Data were sourced from the ABS 2008-09 Energy, Water and Environment Survey (EWES) and the Business activity statement Unit Record Estimates (BURE).

Different model specifications were tested including: (1) log-linear models, (2) a model with interaction dummy variables, (3) a model using derived price measure ($ expenditure per gigajoule), (4) a model with ANZSIC06 2-digit industry dummy variables, (5) models at division level, and (6) a model using energy expenditure data.

Several methodological issues were also highlighted during the estimation process. These included transforming data, dealing with zero observations of turnover and energy consumption volume, and correcting for log transformation bias.

Further Information

For more information, please contact Kay Cao on (02 6252 5264, kay.cao@abs.gov.au)

Recent Developments in the Remote Execution Environment for Microdata

The key components of the Remote Execution Environment for Microdata (REEM) are the development of a Survey Table Builder (STB) and an Analysis Service (AS). These services will access de-identified detailed household survey microdata held in the ABS secure environment, with confidentiality routines built into the table-building and model estimation processes to ensure that outputs are confidentialised in line with ABS legislative requirements and can be released as public use outputs.

REEM will also use internationally recognised standards for the exchange of data and metadata including the use of the Data Documentation Initiative (DDI), Statistical Data and Metadata Exchange (SDMX) and machine-to-machine interfaces (APIs). The adoption of these standards will enable the ABS to disseminate richer metadata to assist researchers in the use and interpretation of ABS microdata.

The development of STB for Household collections has commenced with staged development planned to occur over the next two years. The first stage of STB has now been released to selected users for evaluation. This release enables estimates of counts of households, families, income units and persons to be tabulated from household survey data. Currently, 2009 Survey of Disability, Ageing and Carers (SDAC) data is available in STB. Data from the 2010 Characteristics of Recent Migrants Survey (CORMS) data is also planned to be loaded to STB during the evaluation period. The evaluation period will run until the end of March 2012. After the evaluation is complete, there will be a full release of the first stage of STB. A number of additional survey datasets will be made available on STB during the period from April to June 2012.

The next stage of STB development will incorporate functionality to produce key estimates from continuous data items such as income from household surveys (expected for release in late 2012). Investigations have also commenced into the feasibility of releasing administrative and linked datasets into STB.

In parallel, the ABS has commenced development of the Analysis Service (AS). This Service will incorporate basic analysis techniques including exploratory data analysis and common modelling techniques (such as linear, logistic, multinomial and probit regression models). Data transformations, manipulation and data aggregation tools will also be available to allow users to create their own customised datasets from ABS datasets. The AS will provide a menu-based user interface for users to apply analysis requests as well as an Action Lists to provide users with a log of previous actions applied.

The first staged test release of the AS is planned for July 2012 for the 2010 CORMS data. User feedback will be sought from selected users following the June 2012 release, which will be used to prioritise enhancements and additional functionality for subsequent releases. Later releases will include functionality to handle complex hierarchical data structures and multiple response variables.

Further Information

For more information, please contact Victoria Leaver (02 6252 5445, victoria.leaver@abs.gov.au) or Daniel Elazar (02 6252 6962, daniel.elazar@abs.gov.au).

An Application of the Propensity Score Matching

The Analytical Services Branch recently investigated the usefulness of propensity score matching (PSM) in the context of the ABS Business Characteristics Survey (BCS). The research, which is based on the 2009-10 wave of the BCS data, aimed to produce an adequate matched sample for statistical analyses. Overall, the results show that the PSM can be a useful technique for creating matched samples from the BCS.

The methodology follows the propensity score matching approach suggested by Rosenbaum and Rubin (1983). By matching, the aim was to control for the impact of the variables that could contribute to the selection process, which are assumed to be

captured by the selected observed variables, and to create a matched sample with balanced observed covariates. In the context of this study, the PSM was used to match the firms that received government assistance to similar firms that did not receive any government support.

As part of the methodology, the study considered three matching algorithms, namely, the nearest neighbour, the caliper, and the 5 to 1 digit matching. After making sure that the PSM assumptions were met, the algorithms were subjected to various robustness tests and micro assessments. Some of these tests include the chi-square, the standardised bias, and the pseudo R-squared.

As an application on the matched sample, the study tested for the impact of the correlation within matched pairs by considering a random effects model– i.e., a probit generalised linear mixed model. The results were then compared to those of an ordinary probit model (ignoring the within-group correlation), so as to observe the impact of the correlation within matched pairs, and to those of a probit model on a non-matched sample, so as to examine the impact of the PSM on the regression results.

Further Information

For more information, please contact Cristian Rotaru (02 6252 5098, cristian.rotaru@abs.gov.au) or Sezim Dzhumasheva (02 6252 5298, sezim.dzhumasheva@abs.gov.au)

Improving Economic Activity Survey by using BAS Unit Record Estimates (BURE)

BAS Unit Record Estimates (BURE) are available for the key financial data items reported by businesses on the Australian Taxation Office's Business Activity Statement (BAS). BURE are produced on both a quarterly and financial year basis by applying a number of transformations to the raw BAS data. BURE standardise the reporting period of the raw BAS, remove the GST if it is included in the reported values and impute an estimate for missing returns and correct basic errors. BURE are available for each business for the BAS data items of Total Sales, Total Salary, wages & other payments, Non-Capital Purchases and Capital Purchases.

The extensive use of BURE data in the annual Economic Activity Survey (EAS) has led to a substantial improvement in the accuracy of produced outputs. BURE are used directly in the creation of survey estimates for approximately 700,000 of the smallest non-employing businesses. This reduces the size of the business population in scope of direct collection by 35%. This reduces the reporting burden for small businesses and improves the efficiency of the sample for the remaining population.

The most significant gains for the EAS come with BURE Total Sales and Wages & Salaries as benchmark variables in the estimation process, which is based on using the Generalised Regression estimator (GREG). Using GREG, as opposed to the Horvitz-Thompson estimator, has led to a reduction in the EAS sample from 30,000 to

20,000 businesses, without affecting the quality of the key financial estimates that have good correlation with benchmark variables. This represents a significant reduction in costs for the EAS as well as the burden placed on business providers.

Further savings and improvements in the quality of survey outputs are achieved through the improved efficiency of the editing processes for the EAS, with BURE data being used to evaluate the significance of survey reporting errors. In addition, BURE data is used in the imputation of non-responding businesses.

While the use of GREG estimation delivers savings at a broad output level, more intensive statistical modelling techniques that combine survey data and BURE data, can allow for finer level outputs to be created, which would not otherwise be possible without a larger sample. "Experimental' estimates for the Manufacturing industry have been created for a small set of data items at the state and fine industry level. These estimates are based on a sample of approximately 2,700 businesses, but have levels of quality similar to those, which were achieved previously by the Manufacturing collections with a larger sample of approximately 9,000 businesses.

Like any data source, BURE have a number of deficiencies. The main problem is the difficulty in using BURE for large business structures created using the ABS Economic Unit Model and which have complex relationships between the ABN(s) and businesses. With BURE only being available at the ABN level, the BURE records for complex businesses are produced using a number of approximations which unavoidably lead to a lower quality of data. There are plans for improving the quality of these estimates by using additional sizing information, which is currently collected during the profiling of Large Businesses.

The second problem is related to the fact that there may be a genuine reason for difference of reporting Total sales on the BAS statement and on the Economic Activity Survey form. Some providers complete their EAS form using their Profit and Loss statement (P&L). However, BAS data is reported with an emphasis on correctly reporting GST liabilities/credits. The Annual Integrated Collection (AIC) BSC statistical support team is planning to conduct investigations into units with atypical large BURE Sales and Non-Capital Purchases, to see if modified BURE for these units would be more appropriate for stratification, editing, estimation and data substitution purposes.

Apart from continuing work on developing a greater understanding of BAS data, its quality and its relationship with EAS data, there are intentions to expand the use of BURE data in the Economic Activity survey by 1. investigating options for improving the accuracy of AIC Gross Fixed Capital Formations estimates by evaluating the use of BURE Capitalised Purchases as auxiliary information during stratification and estimation, and 2.increasing the scope of the modelled estimates produced at finer output levels to a larger number of industries and data items.

Further Information

For more information, please contact Irina Pribil (03 9615 7566, irina.pribil@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