Australian Bureau of Statistics

1351.0.55.040

Research Paper

Investigations into Modelling Estimates of Energy Consumption Reporting Undercoverage

New Issue

Research Paper

Investigations into Modelling Estimates of Energy Consumption Reporting Undercoverage

Kay Cao, Jason Wong and Anil Kumar

Analytical Services Branch

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30 AM (CANBERRA TIME) FRI 01 MAR 2013

ABS Catalogue no. 1351.0.55.040

Produced by the Australian Bureau of Statistics

INQUIRIES

The ABS welcomes comments on the research presented in this paper.
For further information, please contact Mr Ruel Abello, Analytical Services Branch on Canberra (02) 6252 6307 or email <analytical.services@abs.gov.au>.

CONTENTS

LIST OF TABLES

ABBREVIATIONS

ABARES	Australian Bureau of Agricultural and Resource Economics and Sciences
ABS	Australian Bureau of Statistics
AES	Australian Energy Statistics
ANZSIC	Australia and New Zealand Standard Industry Classification
ATO	Australian Taxation Office
BAS	Business Activity Statement
BREE	Bureau of Resources and Energy Economics
BURE	Business Activity Statement unit record estimate
EAS	Economic Activity Survey
EWES	Energy, Water and Environment Survey
FES	Fuel and Electricity Survey
GJ	Gigajoule (1 GJ= 10^9 Joules)
NGER	National Greenhouse and Energy Reporting System
TJ	Terajoule (1 TJ= 10^{12} Joules)

ACKNOWLEDGEMENTS

The authors would like to thank the following people for their guidance or assistance with this paper: Ray Chambers and Robert Clark (University of Wollongong); Karen Connaughton, Peter Williams and Sean Lawson (ABS Energy Account Section); Irina Pribil and Kristy Naylor (ABS Statistical Services Branch); and Melissa Gare, Ruel Abello, Franklin Soriano and Peter Rossiter (Analytical Services Branch). Responsibility for any errors or omissions remains solely with the authors. The views in this paper are those of the authors and do not necessarily represent the views of the Australian Bureau of Statistics.

The results of these studies are based, in part, on tax data supplied by the Australian Taxation Office to the Australian Bureau of Statistics under the *Income Tax Assessment Act 1936* which requires that such data are only used for statistical purposes.

No individual information collected under the *Census and Statistics Act 1905* is provided back to the Australian Taxation Office for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes, and not related to the ability of the data to support the Australian Taxation Office's core operational requirements.

Legislative requirements to ensure privacy and secrecy of this data have been followed. Only people authorised under the *Australian Bureau of Statistics Act 1975* have been allowed to view data about any particular firm in conducting these analysis. In accordance with the *Census and Statistics Act 1905*, results have been confidentialised to ensure that they are not likely to enable identification of a particular person or organisation.

INVESTIGATIONS INTO MODELLING ESTIMATES OF ENERGY CONSUMPTION REPORTING UNDERCOVERAGE

Kay Cao, Jason Wong and Anil Kumar
Analytical Services Branch

ABSTRACT

A new data source for the Australian Energy Statistics – the *National Greenhouse and Energy Reporting System* (NGER) was introduced in 2009–10. The NGER does not require businesses using less than 200 TJ to report their energy consumption. This inevitably results in a data gap in the estimate of total energy consumption by businesses.

This paper considers the suitability of Single Equation models, System of Cost Equation models and Dynamic (time series) models for estimating the energy consumption data gap utilising energy consumption data from the Australian Bureau of Statistics 2008–09 *Energy, Water and Environment Survey* (EWES) and financial data from the *Economic Activity Survey* (EAS) and the 2008–09 *Business Activity Statement Unit Record Estimates* (BURE). The single equation model is then applied to the 2009–10 and 2010–11 BURE data to estimate the energy consumption of the businesses that fall below the NGER reporting threshold (200 TJ) for 2009–10 and 2010–11.

1. INTRODUCTION

The Australian Energy Statistics (AES) is currently produced by the Bureau of Resources and Energy Economics (BREE). Before 2009–10, the AES was based on data from the *Fuel and Electricity Survey* (FES), which was conducted by the Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES), and other sources. The FES is largely confined to the mining, manufacturing, communication, rail transport, electricity generation sectors and gas production and distribution sectors. In 2009–10, the FES is replaced by the National Greenhouse and Energy Reporting System (NGER) as the main collection instrument.

Under the NGER, only those businesses above the reporting threshold for greenhouse gases or energy use (which was 350 TJ in 2009–10 dropping to 200 TJ in 2010–11) are required to report their energy consumption. This means energy consumption of many Australian businesses, across different industry divisions, will not be captured by the NGER data as they fall under the reporting threshold. This undercoverage is required to be addressed to get a more complete picture of overall energy consumption by businesses in Australia.

Previous analysis[1] carried out by the ABS using the 2008–09 Energy, Water and Environment Survey (EWES) data estimated that the NGER would lead to about 10.6% undercoverage of aggregate energy consumption data in the reference year because of the exclusion of small to medium energy consuming businesses. The analysis suggested a combination of a limited industry gap survey and modelling approach for filling the data gap left by the NGER.

This paper investigates the modelling options for estimating the NGER undercoverage. The remainder of the paper is organised as follows. Section 2 presents a brief review of the literature on energy demand modelling. Section 3 describes the data sources, the derivation of required variables, data merging and cleaning, and data quality issues. Section 4 discusses alternative model specifications and regression results. Section 5 describes the steps in deriving the estimates of the NGER undercoverage for 2009–10 and 2010–11. Section 6 concludes and provides some recommendations.

---

1 *Estimation of Undercoverage of the National Greenhouse and Energy Reporting System (NGER) due to Business falling under the Threshold*, unpublished ABS working paper, December 2010.

2. LITERATURE REVIEW ON ENERGY MODELLING APPROACHES

2.1 Economics of energy

Energy is an important input into any modern production process. The demand for energy is derived from the need to use energy-using equipment to produce goods and services either in industry or household. Therefore, energy demand is closely related to the type of capital used, its efficiency in using energy to produce products and services and capital utilisation.

As it is generally not possible to change the stock of capital in a short time period, demand for energy could respond differently to changes in exogenous factors in the short-run compared to the long-run (i.e. longer time period when the stock of capital and its efficiency can be changed).

Demand for energy can also be affected by the degree of substitution between other production inputs and energy or substitution between different types of energy. For example, the common input factors used in a production process such as labour, capital and other materials can either be a complement to or a substitute for energy. Between energy types, for example, electricity can replace coal or gas in industrial processing, or renewable energy can substitute for depletable energy when technology allows.

As energy is also a commodity, apart from the technical factors (output level, capital efficiency and utilisation), prices of energy and other inputs may also play an important role in determining its demand. Changes in energy price do not only impact energy demand in the short-run but also facilitate changes in capital types and its efficiency which in turn impacts on energy demand in the long-run.

The above aspects of energy economics are important in shaping energy demand models. Modelling energy demand has flourished especially after the first world oil shock in the early 1970s. Over the years, together with developments in econometric methods, there have been different generations of models employed in modelling energy demand. The following section summarises three main categories of energy demand models.

2.2 Modelling approaches

Before going into the details of different types of energy models, it is important to note that there is no single 'right' approach to modelling energy demand (Ryan and Plourde, 2009). Energy models vary depending on aggregation level, e.g. firm vs industry vs national level. The choice of model forms can also vary according to the type of data, e.g. cross-sectional vs time series data. Models can also be either partial or general equilibrium models, with the latter taking into account both demand and supply factors. Additionally, depending on the focus on energy as a whole or a specific fuel type, models can use 'top down' (high level of energy aggregation) or 'bottom up' (breakdown of energy types) approach.

2.2.1 Single equation models

This is one of the early approaches to modelling energy demand. It is not based on any economic optimisation process and often only involves production output as the main explanatory variable. Griffin (1991) and Ryan and Plourde (2009), for example, discussed this early approach. A generalised model for this approach can be specified as follows:

$$E = \beta_1 + \beta_2 Y + \sum \gamma_i Z_i + u \tag{1}$$

where

E is energy consumption;

Y is production output;

Z_i are other explanatory variables, most commonly energy price and lag variables of E (in the case of dynamic adjustment models); and

u is a random error term.

A simple single equation model has the advantage when there are data constraints, for example, when there are only data on production output. A single equation can also be considered as a reduced form of a more complicated modelling approach when there are data and other constraints.

2.2.2 System of cost equation models

This type of model is based on the theory of production economics. Application to modelling energy demand was facilitated especially since the introduction of the translog (transcendental logarithmic) function by Christensen *et al.* (1973). Translog production function, for example, relaxes the range of substitution possibilities between inputs which does not require a unitary elasticity (as in the case of Cobb–Douglas production function) or constant elasticity of substitution (Griffin, 1991).

A common translog production (or cost) function involves inputs such as capital (K), labour (L), energy (E) and materials (M). To derive an optimal input demand equation, an optimal cost function is specified as:

$$\begin{aligned} \ln C^* = {} & \beta_0 + \beta_y \ln Y + \frac{1}{2}\beta_{yy}(\ln Y)^2 + \sum_{i=1}^{4}\beta_i \ln x_i + \\ & \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4}\beta_{ij}(\ln x_1)(\ln x_j) + \sum_{i=1}^{4}\beta_{yi}(\ln Y)(\ln x_i) \end{aligned} \tag{2}$$

where

C^* is the optimal production cost;

Y is output level; and

x_i represents prices of the i production inputs (K, L, E, M).

To derive an optimal input demand equation (in this case, energy), Shephard's lemma is applied (for detailed explanation and proof, see for example, Coelli *et al.*, 2005). It states that the partial derivatives of optimal cost function with respect to input prices give the corresponding conditional input demand functions, which are the economically optimal input levels to produce a given output quantity. Therefore, we obtain an input demand equation for energy as:

$$E^* = \beta_e + \beta_{ee} \ln x_e + \beta_{ye} \ln Y + \beta_{ke} \ln x_k + \beta_{le} \ln x_l + \beta_{me} \ln x_m \quad (3)$$

where E^* represents the share of energy cost in total production cost and x_k, x_l, x_e and x_m are the prices of K, L, E and M, respectively.

To obtain efficient estimates, often a system of cost and cost share equations is estimated (e.g. equations (2) and (3)). In addition, when there are data available over time, a time variable can also be introduced into the model to capture technological change and its impacts on input demand over time. Berndt and Wood (1975), one of the first papers utilising translog function approach to energy demand modelling, provided a good example of the use of this method.

2.2.3 Dynamic (time series) models

As energy demand is closely associated with the use of capital, the stock of which is often considered only changeable over a long period of time, energy demand behaviour is arguably different in the short-run compared to the long-run. Whenever there is time series data available, dynamic modelling approach should be considered.

Dynamic (or time series) models often involve the lag terms of the dependent and/or explanatory variables. For example, in the single equation approach discussed in Section 2.2.1 (equation (1)), an extension to include lag terms can be specified as:

$$E = \beta_1 + \beta_2 Y + \sum \gamma_i Z_i + \beta_3 E_{t-1} + u \quad (4)$$

where E_{t-1} is energy consumption in the previous period.

With recent developments in time series econometric techniques, various forms of time series models have been introduced into energy demand modelling. A summary of popular models such as Error Correction Model (ECM), Structural Time Series Model (STSM) and asymmetric price models can be found in Ryan and Plourde (2009).

2.3 Modelling approach used in this study

Given the availability of only cross-sectional data for this study, we focus on static (non-dynamic) models for energy consumption modelling (equation (1)). The dynamic modelling approach (equation (4)) will not be used. Furthermore, due to the unavailability of energy price data in the future years, the system of equation approach (equation (2) and (3)) will not be used.

For a pure demand analysis, a system of equations (2) and (3) is perhaps the most robust method to use as it will reveal further information on the substitution between energy and other inputs. However, using available data will require some modification of the usual production economic framework due to the unavailability of data on prices of capital and non-capital inputs. This is beyond the scope of this project but could be addressed in future research.

Due to the reasons stated above, in this study, we utilise the single equation modelling approach as specified in equation (1) to estimate the energy consumption gap for future years.

3. DATA

Data on energy consumption are sourced from the 2008–09 *Energy, Water and Environment Survey* (EWES). Two sources of financial data are considered (BURE and EAS). However, for the final estimates, only BURE data are used. These datasets also jointly provide other business information such as Australian Business Number (ABN), industry ANZSIC (2006) classification, etc. All datasets provide cross-sectional data where data at unit level were collected for the reference year (2008–09).

3.1 EWES (source of energy consumption data)

The 2008–09 EWES is the first of a three-yearly survey conducted by the Australian Bureau of Statistics (see ABS, 2010b). It provides information on energy, water and environment management practices for selected Australian industries. It also provides information on energy and water consumption (both expenditures and usage quantities) and generation of renewable and non-renewable fuels. Energy consumption data are broken down into electricity, gas, and other fuels.

The EWES excludes the following industries:

- Agriculture;
- Water supply, sewerage and drainage services;
- Finance;
- Insurance;
- Public administration;
- Defence; and
- Private households employing staff.

Summaries of aggregate energy consumption data from the 2008–09 EWES are presented in Appendixes A and B.

3.2 BURE and EAS (sources of financial data)

3.2.1 2008–09 data

The BURE data provide a user friendly way of accessing the key financial data from the Australian Taxation Office (ATO) Business Activity Statement (BAS). Raw BAS data can be reported to the ATO weekly (or more frequently), monthly and quarterly or annually which makes it difficult to compile estimates for a particular time period. The BURE data standardise the reporting period of raw BAS, remove the GST if included in reported values and impute an estimate for missing returns. This provides a comprehensive set of BAS unit record estimates for all businesses with active

Income Tax Withholding (ITW) or Goods and Service Tax (GST) role on the ABS common frame.

The BURE data are produced on both quarterly and financial year basis. In this study, we use the annual 2008–09 BURE data. It contains data of total turnover, salary and wages, capital and non-capital purchases.

An alternative source of business financial data is the EAS. The EAS is conducted annually (June year). A combination of the EAS data and data from ATO BAS is published in the Australian Industry series (ABS, 2010a). It provides detailed data on business income and expenditures.

There may be discrepancies in the corresponding variables between the BURE and the EAS data for the same business due to different variable definitions or data collection processes.

The total number of businesses in the 2008–09 BURE is 1,984,588, while the original sample size of the 2008–09 EAS is 22,649. A summary of turnover data from the 2008–09 BURE is presented in Appendix C.

3.2.2 2009–10 and 2010–11 BURE data

These datasets are used for energy consumption volume estimation in the corresponding years.

One important characteristic of the data that has implications for the energy gap estimation is the presence of zero-turnover observations. The following table shows this aspect of the data.

3.1 Zero-turnover in 2009–10 and 2010–11 BURE

	Zero turnover but some expenditures	*Zero turnover and no expenditure*	*Total zero-turnover units*	*Per cent of total population*
2008–09 BURE	102,105	105,011	207,116	10.3%
2010–11 BURE	104,157	110,094	214,251	10.5%

Units reporting zero-turnover make up about 10% of the total number of businesses in both datasets.[2] These units are distributed across all industries. About half of these units still report one or more expenditure items. This can be due to reporting arrangements the unit has with the tax office, i.e. the business is linked to another business which reports all turnovers. In this case, their energy consumption will be included in the consumption of the other business and still contribute to the total gap estimate.

2 Total BURE business population is 2,013,650 in 2009–10 and 2,047,697 in 2010–11.

Units with neither turnover nor other expenditures are considered non-operative (i.e. temporary nil or confirmed 'dead') Therefore, they can be excluded from the total energy gap estimate.

Another characteristic of these datasets is the exclusion of some Financial and Insurance Service industries (ANZSIC 62 & 63) and some Public Administration and Safety industries (ANZSIC 75 & 76). This results in the exclusion of their energy consumption from the final gap estimate.

3.3 Merging and cleaning the data

The 2008–09 EWES data were merged with the 2008–09 BURE data. Data from the EWES and the EAS were also merged to create an alternative dataset for examining the suitability of using EAS data for estimation purposes.

After merging, data (observations) that are considered not suitable for analysis[3] were excluded from model estimation. Further, due to the logarithmic specification, units with zero turnover and energy volume were excluded. The data used in model regression also excluded units with total energy consumption of above 200 TJ to restrict the sample to those falling below reporting threshold.

The excluded units are distributed across all industries. Estimation adjustment due data exclusion is discussed further in section 5.3.

The data cleaning steps for the merged EWES and BURE data are shown in table 3.2.

3.2 EWES/BURE data cleaning steps

	Sample size	Units excluded
Full EWES sample	14,404	–
Excluding ABS profiled unit	11,165	3,239
Excluding non-full-period operator	9,258	1,907
Excluding units consuming >200TJ	9,176	82
Excluding zero energy consumption*	6,391	2,785
Excluding zero turnover	6,302	89

Note: * adjustment for exclusion of zero consumption data is discussed in Section 5.3.

The merged and cleaned EWES/BURE and EWES/EAS data had 6,302 and 5,818 units, respectively.

Due to having a slightly larger sample size and also potentially including micro non-employing units, the BURE were used to predict the NGER gap. The discussion in the following sections mainly includes results using the BURE data.

3 These include ABS profiled units and units which were not full period operators.

3.4 Derivation of total energy consumption (GJ)

Based on the consumption volumes of electricity, gas and other fuels, we derived total energy consumption for each business (in a common unit – gigajoules or GJ), using a set of conversion factors. Each fuel has its own conversion factor which represents the amount of energy in GJ embedded in each volume unit. Details of conversion factors are presented in Appendix D.

Multiplying the fuel consumption volume with the corresponding conversion factor gives the amount of total energy consumption in GJ for each fuel. Aggregating this up for all fuels consumed by a business gives the total energy consumed in GJ for this unit.

4. MODEL RESULTS

In this section, we present the estimation results for different model specifications, and discuss the chosen model for gap estimation. The static single equation model as discussed in Section 2.2.1 (equation (1)) was employed to estimate the relationship between energy consumption and output by different industries.

4.1 Model specification

Due to the unavailability of energy prices in the future years, in this model we only used turnover as the main explanatory variable. Further, to differentiate between industries we also included industry dummies.

The estimated coefficient of turnover reflects the impact on energy demand from the production requirement. The inclusion of financial variables other than turnover (e.g. wages, capital and non-capital expenditures) is unnecessary due to the high correlation between turnover and these variables (both at raw and log scales).

The model is re-specified as:

$$E_i = \beta_1 + \beta_2 Y_i + \sum_{j}^{n-1} \gamma_j D_j + u_i \tag{5}$$

where

E_i is total energy consumption volume of unit i (GJ);

Y_i is production output (here, business turnover is used as a proxy for total output);

D_j are industry dummies (here we categorise industries using the ANZSIC 2006 at three-digit level);

n is the number of the three-digit ANZSIC industries; and

u_i is a random error term, which represents the difference between observed data and model estimates.

4.2 Estimation method

We estimated the above model, using the 2008–09 EWES/BURE data, for those units with total energy consumption of less than 200 TJ (i.e. those who are not required to report to the NGER).

In this paper, we report the results from the ordinary least squares method of estimation.

4.3 Results

To obtain the final model, during the model testing phase, we considered several variations of the model:

4.3.1 Log-linear model

The data, as described in Section 3 above, show a high degree of skewness (i.e. a large proportion of units with low energy consumption level). Often, in such cases, variables are transformed to a logarithmic scale (natural logarithm) to enable the error term of the model satisfying the normal distribution condition, and thus enhancing the robustness of the model regression.

The estimation results show that model estimates for variables in raw scale have very low goodness-of-fit (Adjusted R-square=0.0696) (see Appendix E).

On the other hand, the fitted log-linear regression model for energy consumption has an adjusted R-squared value of 0.4220. It is considered a reasonable fit that can be used to estimate the population value of average energy consumption.

Estimation results for the log-linear model, using the EWES/BURE data, are presented in table 4.3. This is the model used for the energy gap estimation.

The coefficient of turnover is statistically significant at 1% significance level. Many industry coefficients are also statistically significant, suggesting their energy demand is statistically different from the base industry after controlling for turnover.

4.3.2 Model with interaction dummy variables

Industry specific characteristics affecting energy consumption can also be captured by an interaction term between industry dummy and turnover. This is often referred to in the literature as *slope dummy* variable where industry characteristics affect energy intensity (energy to turnover ratio). We have considered two variations of the model above: (1) model with both industry and interaction dummy variables and (2) model with only interaction dummy variables.

A model with interaction dummy variables can be specified as follows:

$$E_i = \beta_1 + \beta_2 Y_i + \sum_{j}^{n-1} \gamma_j D_j Y_j + u_i \qquad (6)$$

Estimation results for this model in log form, using the EWES/BURE data, are presented in Appendix F. This model has similar goodness-of-fit to that of the model with intercept dummy variables (Adjusted R-square=0.4187). This suggests the two model specifications have similar prediction ability. The model with both industry and interaction dummy variables results in only a few significant coefficients and therefore is not used for the energy gap estimation (Appendix G).

4.3 Single equation model – regression results
Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	Standard error		Coefficient	Standard error
Intercept	–4.0925 ***	0.4825			
Log_bure_turnover	0.6329 ***	0.0127			
ind020 (=reference)			ind211	1.8870 ***	0.5465
ind030	–1.1133 *	0.6583	ind212	1.9358 ***	0.6310
ind041	–0.3984	0.5059	ind213	1.2700 **	0.6100
ind042	–0.2351	0.8353	ind214	1.1463	0.7108
ind051	–1.8403 *	0.9977	ind221	1.3344 *	0.6952
ind052	–0.9864 *	0.5263	ind222	1.5051 ***	0.5281
ind060	2.5136 ***	0.6234	ind223	1.2062 *	0.6391
ind070	1.0945	0.7729	ind224	1.6304 ***	0.6235
ind080	2.2971 ***	0.4958	ind229	1.6186 ***	0.5464
ind091	3.1321 ***	0.6483	ind231	0.8076	0.5402
ind099	2.0980 ***	0.6818	ind239	0.8607 *	0.5121
ind101	1.9610 ***	0.6582	ind241	0.6516	0.6100
ind109	1.9419 ***	0.5470	ind242	0.3257	0.6310
ind111	2.3510 ***	0.5213	ind243	0.9201	0.5887
ind112	1.7198 ***	0.6393	ind244	1.3962 **	0.5842
ind113	1.8497 ***	0.5075	ind245	1.5630 *	0.8011
ind114	1.7764 ***	0.5985	ind246	1.8561 ***	0.6392
ind115	1.1108	0.6814	ind249	1.5709 **	0.7489
ind116	1.8897 ***	0.6164	ind251	0.8822	0.5537
ind117	1.6603 ***	0.5380	ind259	0.9934 *	0.5420
ind118	1.7343 ***	0.5511	ind261	0.9971	0.8008
ind119	2.4657 ***	0.5566	ind263	–1.9849	1.2355
ind121	1.5422 ***	0.5151	ind264	–0.2298	0.8764
ind122	4.0646 **	2.0446	ind270	0.6323	0.7730
ind131	2.1963 ***	0.6309	ind291	–0.4035	0.9976
ind132	1.0766 *	0.6482	ind292	–0.5340	0.6815
ind133	1.4907 **	0.5886	ind301	–1.3638 ***	0.4860
ind134	1.6446 **	0.7285	ind302	–1.0438 *	0.5601
ind135	0.6183	0.5263	ind310	–0.9125 *	0.5352
ind141	3.0803 ***	0.5985	ind321	–1.0300 *	0.5843
ind149	1.5341 ***	0.5464	ind322	–0.8189	0.6486
ind151	0.8811	0.6042	ind323	–1.2503 **	0.5381
ind152	1.5423 ***	0.5724	ind324	–0.0390	0.5726
ind161	1.2510 **	0.5420	ind329	–0.3677	0.6815
ind170	1.7760 ***	0.5464	ind331	0.6042	0.8010
ind181	2.0019 **	0.9294	ind332	–1.1804 **	0.5560
ind182	1.5167 **	0.7727	ind333	–0.7967	0.6696
ind183	2.2603 ***	0.5889	ind341	–0.5550	0.8013
ind184	1.2893 **	0.5490	ind349	–1.3871 **	0.5411
ind185	1.1009 *	0.6165	ind350	–1.1548 **	0.5541
ind189	1.4289	1.0927	ind360	–0.5479	0.5409
ind191	1.6651 ***	0.5593	ind371	–0.9176	0.8010
ind192	1.6360 ***	0.5986	ind372	–1.9190 **	0.7498
ind201	1.1427 *	0.6392	ind373	–0.6004	0.5690
ind202	2.5768 ***	0.6100	ind380	–1.4640 **	0.5890
ind203	2.2470 ***	0.5464	ind391	–0.5744	0.5628
ind209	1.2940 **	0.5985	ind392	–0.6302	0.6952

4.3 Single equation model – regression results (cont.)

	Coefficient	Standard error		Coefficient	Standard error
ind400	–0.4623	0.5804	ind664	0.3133	0.8763
ind411	1.8113 ***	0.5541	ind671	0.8230 *	0.4620
ind412	0.9372	0.5843	ind672	–0.0414	0.4994
ind421	0.2113	0.7285	ind691	0.8322	0.6313
ind422	–0.6609	0.5802	ind692	–1.0456 **	0.4801
ind423	–1.1270 *	0.6815	ind693	–0.7281	0.5152
ind424	–0.6608	0.6392	ind694	–0.5878	0.5807
ind425	0.0143	0.5842	ind695	–0.9393	0.9979
ind426	–2.7556	2.0447	ind696	–1.2936 **	0.5186
ind427	0.0075	0.5724	ind697	–0.0774	0.8346
ind431	–0.7116	0.6816	ind699	0.9578	1.2352
ind432	–2.5451	2.0443	ind700	–0.8951 *	0.4988
ind440	2.1651 ***	0.4828	ind721	–1.4417 ***	0.4798
ind451	1.2576 ***	0.4871	ind722	–1.3963 **	0.5576
ind452	1.7212 ***	0.5986	ind729	–0.9264 *	0.5179
ind453	2.1182 ***	0.5317	ind731	–1.7377 ***	0.5310
ind461	2.7674 ***	0.4981	ind732	1.2008 *	0.6957
ind462	2.6558 ***	0.6103	ind771	–1.2723 **	0.5249
ind471	3.0726	0.9286	ind772	–0.8368	0.9975
ind472	1.7878	1.2353	ind801	–0.3365	0.6310
ind481	0.7565	0.8349	ind802	1.5235 ***	0.4976
ind482	2.7600 ***	0.7727	ind810	0.0801	0.5295
ind490	1.3419 **	0.6313	ind821	0.2862	0.5361
ind501	1.4637 **	0.5986	ind822	1.0709	0.9977
ind502	1.3884	1.0927	ind840	1.0897 **	0.5488
ind510	1.4938 ***	0.5381	ind851	–0.5965	0.5442
ind521	0.9606 *	0.5663	ind852	1.1480	0.9977
ind522	1.5839 **	0.7492	ind853	–0.4298	0.5420
ind529	1.6151 ***	0.4927	ind859	0.3740	0.9976
ind530	2.1978 ***	0.5723	ind860	1.5912 ***	0.4980
ind541	–0.6472	0.5842	ind871	–0.0506	0.5264
ind542	0.1656	1.4799	ind879	0.4926	0.5039
ind551	–0.4814	0.5512	ind891	0.9844 *	0.5760
ind552	0.1238	0.9975	ind892	0.5969	0.5044
ind561	–0.2428	0.6815	ind900	0.0180	0.4868
ind562	0.3663	0.8761	ind911	0.4846	0.4943
ind570	0.1822	0.7727	ind912	0.7102	0.5934
ind580	–1.5498 **	0.6482	ind913	0.9875	0.7285
ind591	–0.3699	0.7732	ind920	1.4962 **	0.6164
ind592	–0.1174	0.7110	ind941	–0.6271	0.5174
ind601	–0.4331	0.7489	ind942	–0.2387	0.5888
ind602	–3.5508 ***	1.0930	ind949	–0.3273	1.0927
ind641	–0.4768	0.4741	ind951	–0.2979	0.5728
ind642	–0.1742	0.6391	ind953	0.9256	0.5887
ind661	0.0450	0.6310	ind954	1.3964 **	0.5441
ind663	–0.2823	0.4995	ind955	–0.0556	0.4771
Adjusted R^2	0.4220				
Observations	6302				

Note: Log_bure_turnover is business turnover from BUREs transformed to natural log scale;
ind030–ind955 are industry dummies from ANZSIC 030 to ANZSIC 955;
*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

4.3.3 Model with ANZSIC (2006) two-digit industry dummies

In this model, the dummy variables are specified at the two-digit industry level, which results in a smaller model with 78 coefficients rather than 186 coefficients as in the case of the model with three-digit industry dummies.

Estimation results are presented in Appendix H. This model has an adjusted R-square of 0.4108, slightly lower than that of the model with three-digit industry dummies.

As outputs for some industries are required at the three-digit industry level, we have decided not to use this model for the energy gap estimation.

4.3.4 Model with both turnover and energy price

Recall that the single equation approach can also include energy price if there are data available. From the 2008–09 EWES data, a measure of energy price ($ per GJ) can be derived. Results for this model are reported in Appendix I.

This model has an adjusted R-square of 0.7222. The coefficient of energy price has a negative sign (–0.9092) as expected. This suggests that a one per cent increase in energy price could lead to a reduction of 0.9 per cent in business energy consumption, assuming other factors stay the same.

Due to the unavailability of energy price data in the future years, this model is not used for the energy gap estimation.

4.3.5 Model for a specific industry division

In this model, the single equation (5) is used for each industry division. This means the total sample is broken down to division level and regression is done for each of them. The division sample size ranges from 49 observations for the smallest sample (Division D – Electricity Gas Water and Waste Services) to 1,550 observations for the largest sample (Division C – Manufacturing).

Regression results for selected divisions are provided in Appendix J. Generally, there is not much gain in splitting the total sample, i.e. the R-squares and root mean square errors (RMSEs) are at the same level as in the regression results for the total sample.

4.3.6 Using sampling weights in model estimation

We have also considered using sampling weights (available in the EWES data) for model estimation. Using weights, however, does not increase the model's goodness-of-fit (R-square). Moreover, due to the data cleaning process, there may be a need to re-adjust these weights. However, as the energy data are not currently on the survey frame, re-calculating weights is difficult.

4.4 Other models

We also considered two other cases: (1) model using financial data from the EAS, and (2) model with energy expenditure as the main explanatory variable (Expenditure Model).

4.4.1 Model using EAS data

In this model, turnover data are sourced from the EAS data. Results of the single equation model are presented in Appendix K. Estimation results are similar to those of the model using the BURE data (adjusted R-square=0.4288). However, using the EAS data results in a smaller sample size after taking out missing and zero observations (5,818 observations vs 6,302 observations when using BURE). Also, the 2010–11 EAS is not yet available at the time of this study to enable energy gap estimation for this year.

4.4.2 Expenditure model

This model specifies total energy expenditure as the main explanatory variable. It can be considered as an alternative option for estimating energy consumption volume when there are energy expenditure data but no energy price data. For the 2008–09 sample, energy expenditure data are available from EWES.

The model is specified as:

$$E_i = \beta_1 + \beta_2 V_i + \sum_{j}^{n-1} \gamma_j D_j + u_i \tag{7}$$

where

E_i is total energy consumption volume of unit i (GJ);

V_i is total energy expenditure;

D_j are industry dummies (here we categorise industries using ANZSIC at three-digit level);

n is the number of the three-digit ANZSIC industries; and

u_i is a random error term, which represents the difference between observed data and model estimates.

The results are presented in Appendix L.

This model has reasonably high goodness-of-fit (Adjusted R-square=0.6367). It, however, merely shows the relationship between energy volume and value (in the absence of a price variable) and does not reflect the output-energy relationship.

5. ESTIMATION OF NGER UNDERCOVERAGE FOR 2009–10 AND 2010–11

This section discusses the use of the single equation model estimated in Section 4.3.1 (table 3) to predict the energy consumption of those units that are not required to report to the NGER.

The key estimation steps are:

1. Exclude NGER units from the BURE datasets;
2. Derive estimates of energy consumption volume using regression results from the single equation approach;
3. Derive estimated probability of units reporting zero energy volume;
4. Derive an adjustment term to correct for log transformation bias;
5. Derive final energy volume estimates at the unit level; and
6. Aggregate energy consumption volumes from the unit level to the required industry level.

The following sections describe each of these steps.

5.1 Exclusion of NGER units

Before the BURE data could be used in the energy consumption gap estimation for 2009–10 and 2010–11, units that already reported to the NGER were removed from the datasets.

This was done by matching the Australian Business Numbers (ABN) of those units in the 2009–10 and 2010–11 BURE data sets with those from the NGER database of the corresponding year.

The current match rate is about 30% (190 units from BURE 2009–10 and 245 units from BURE 2010–11) due to the profiled units not having ABNs. We overcome this issue by only adding those energy estimates less than 200 TJ in the total gap estimate.

5.2 Unadjusted estimates of energy consumption volume at unit level

The model estimation results as in table 3 (i.e. estimated coefficients of the intercept term, turnover and industry dummies) and turnover data of non-NGER units from the 2009–10 and 2010–11 BURE were used to predict total energy consumption volume (GJ) for each unit for these two years.

For those industries excluded from the 2008–09 EWES (see Section 3.1), only the coefficients of the intercept and turnover were used for their energy consumption estimation (i.e. the industry dummy variable is zero).

The unit-level estimates were then adjusted for the probability of zero-energy reporting and log transformation bias. These adjustments are explained further below.

5.3 Adjustment for probability of zero-energy reporting

Due to the presence of zero energy volume data, there is a significant proportion of units in the cleaned sample which is excluded from the model estimation step (see table 3.2 in Section 1.2.3 on data cleaning). The number of units excluded due to zero energy and turnover data is 2,874, most of which are units reporting zero-energy volume (97%).

Zero-energy reporting could be due to having no information on energy volume or non-response. There are cases when businesses report energy expenditures but not energy volumes (959 units). There are also cases when energy consumption is bundled in other costs (e.g. rent paid to lessor), in which case, it may be included in energy consumption of another industry.

In this study we used a blanket approach to deal with zero-energy data, which is to estimate the probability of zero energy volume reporting regardless of the reason. An alternative approach is to impute energy volumes of those businesses that report energy expenditures. However, we did not follow this approach as it only addresses about a third of the total zero-energy units.

To estimate the probability of zero-energy reporting, we used a logistic modelling approach:

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_1 + \beta_2 \ln Y_i + \sum_{j=1}^{n-1} \gamma_j D_j \tag{8}$$

where

$\ln$ represents the natural logarithm of the term in the brackets;

P_i is the probability for a unit NOT reporting zero-energy volume (this is equivalent to giving a value 1 to units with non-zero energy volume and zero to a zero-energy case);

$\ln Y_i$ is log of business turnover;

D_j are industry dummies; and

n is the number of the three-digit ANZSIC industries.

Regression results are presented in Appendix M. We then used the 2009–10 and 2010–11 BURE data to estimate this probability for each unit and used this estimated probability to adjust the energy volume estimates derived from step 5.2 above.

5.4 Adjustment for log transformation bias

It is widely discussed in the literature that when estimates from a log-scale model are re-transformed to obtain the raw-scale estimates, a simple anti-log retransform would result in an underestimation of the variable of interest.

In our case of modelling energy consumption (E), after obtaining an estimate of (natural) log of energy (logE), then taking anti-log to derive E (*e* raised to the power of logE)[4], this often underestimates the mean energy consumption (Duan, 1983; Newman, 1993; Chambers and Dorfman, 2003). This is referred to as the bias from transforming variables, or log transformation bias in the case of log transforming.

There are several options for correcting for this bias, for example, *naïve correction* (Newman, 1993), *smearing estimate* (Duan, 1983), *weighted smearing estimate* (Chambers and Dorfman, 2003) and other general corrections (Karlberg, 2000; Chandra and Chambers, 2006).

In this study, we utilised the original smearing method from Duan (1983).

The adjustment factor A is specified as:

$$A = \frac{\sum_{i=1}^{N} e^{u_i}}{N} \tag{9}$$

where

u_i is the regression residual; and

N is the regression sample size (6,302).

The estimated value for the adjustment factor is 4.5, which is higher than estimates from some existing studies (e.g. Newman, 1993; Chambers and Dorfman, 2003). This potentially includes not just transformation bias but also model specification bias.

4 e = 2.71828182845904

5.5 Adjusted estimates of energy consumption volume at unit level

The adjusted estimate is the product of the unadjusted estimate, estimated probability of non-zero energy volume and bias correction.

$$E_i^* = E_{0i} \cdot P_i \cdot A \tag{10}$$

where

E_i^* is the final adjusted estimate of a unit's energy consumption volume;

E_{0i} is the unadjusted estimate derived from step 5.2;

P_i is estimate of probability of non-zero energy volume; and

A is the adjustment factor estimated from step 5.4 above.

5.6 Aggregate industry estimates

The estimates of energy consumption volume at unit level were then aggregated up to the required industry level, including both two-digit and three-digit level. The results are presented in table 5.1 below.

For 2009–10, total energy consumption of those businesses which are not required to report to the NGER (i.e. energy consumption gap) is estimated at 810,228 TJ or 9.3% of total industry energy consumption. This shows the proportion of total undercoverage consumption in total business energy use, which is the sum of total undercoverage and total NGER consumption. This compares to an estimate of 470,922 TJ for 2008–09 previously provided by the Energy and Environment Statistics Section using a non-modelling method and the 2008–09 EWES data.[5] This study had a narrower scope compared to the current project (excluding ANZSIC06 subdivision 01 and 28).

The gap estimate for 2010–11 is 834,610 TJ or 8% of total industry energy consumption.

---

5 *Estimation of undercoverage of the National Greenhouse and Energy Reporting system (NGER) due to Business falling under the Threshold,* unpublished ABS working paper, December 2010.

5.1 Energy consumption gap estimates, 2009–10 and 2010–11 (TJ)

ANZSIC 06 Industry	*2009–10*	*2010–11*
Agriculture, forestry & fishing	18,127	19,201
Mining	37,416	40,867
Manufacturing		
Food, beverages, textiles (11–13)	53,821	54,035
Wood, paper, printing (14–16)	29,097	29,213
Chemicals (17–19)	24,142	23,737
Iron & steel (211,212)	6,964	7,046
Non-ferrous metals (213,214)	2,558	2,584
Other manufacturing (20, 22–25)	73,657	75,042
Total Manufacturing	190,239	191,656
Electricity, gas, water & waste	4,825	5,004
Construction	14,273	14,945
Transport		
Road (46)	194,826	199,227
Rail (47)	1,685	2,059
Water (48)	2,082	2,160
Air (49)	2,418	2,472
Other transport & support services (50–53)	36,975	39,582
Total Transport	237,987	245,500
Commercial & services		
Wholesale & retail trade (F,G)	77,759	80,094
Accommodation (H)	108,537	112,526
Communication (J–M)	51,342	53,017
Other (N–S)	69,724	71,801
Total Commercial & services	307,361	317,437
Total Energy Gap	810,228	834,610
Total NGER	7,886,866	9,561,494
Total Energy Use	8,697,093	10,396,104
% Gap in Total	9.3%	8.0%

Note: Total energy gap is the sum of total energy consumption of each division in the table (this is for businesses under the 200 TJ threshold). Total NGER is the total energy consumption volume reported to the NGER. Total NGER in 2009–10 reflects the 350TJ threshold. Total energy use is the sum of total energy gap and total NGER.

5.7 Confidence interval estimates

We also derived the 95% confidence interval for the gap estimates using the normal approximation method of bootstrap with 1,000 resamples[6] (Haukoos and Lewis, 2005). Results of the lower and upper bounds of estimates are presented in table 5.2.

6 This method generates 1000 random samples from the original sample of estimates of energy consumption for each industry derived from step 5.6. Each resample has the same sample size as the original sample. The sum of each sample is then calculated, from which the confidence interval estimates are calculated based on the z-distribution of the resamples. This method produces reliable confidence interval estimates, assuming the energy consumption estimates are good representations of the energy consumption of units consuming less than 200 TJ.

5.2 Lower and upper bounds of energy gap estimates (at 95% confidence interval)

	2009–10		2010–11	
ANZSIC 06 Industry	*Lower*	*Upper*	*Lower*	*Upper*
Agriculture, forestry & fishing	17,938	18,316	18,948	19,453
Mining	34,539	40,294	37,919	43,815
Manufacturing				
Food, beverages, textiles (11–13)	50,952	56,690	51,250	56,819
Wood, paper, printing (14–16)	27,518	30,677	27,598	30,828
Chemicals (17–19)	22,301	25,983	22,014	25,459
Iron & steel (211,212)	6,066	7,861	6,195	7,896
Non-ferrous metals (213,214)	1,869	3,247	1,886	3,282
Other manufacturing (20, 22–25)	71,440	75,875	72,732	77,352
Total Manufacturing	180,146	200,333	181,675	201,636
Electricity, gas, water & waste	3,871	5,779	4,084	5,923
Construction	14,037	14,509	14,694	15,197
Transport				
Road (46)	191,543	198,108	196,015	202,438
Rail (47)	917	2,454	1,232	2,887
Water (48)	1,697	2,466	1,736	2,584
Air (49)	1,848	2,988	1,907	3,037
Other transport & support services (50–53)	35,310	38,641	37,680	41,485
Total Transport	231,315	244,657	238,570	252,431
Commercial & services				
Wholesale & retail trade (F,G)	76,404	79,114	78,545	81,643
Accommodation (H)	106,686	110,387	110,653	114,399
Communication (J–M)	50,378	52,305	52,046	53,987
Other (N–S)	67,967	71,480	70,085	73,516
Total Commercial & services	301,435	313,286	311,329	323,545
Total Energy Gap	783,281	837,174	807,219	862,000
Total NGER	7,886,866	7,886,866	9,561,494	9,561,494
Total Energy Use	8,670,147	8,724,040	10,368,713	10,423,494
% Gap in Total	9.0%	9.6%	7.8%	8.3%

Note: Total energy gap is the sum of total energy consumption of each division in the table. Total NGER is the total energy consumption volume reported to the NGER. Total energy use is the sum of total energy gap and total NGER.

The range of the undercoverage estimate is from 9% to 9.6% in 2009–10 and from 7.8% to 8.3% in 2010–11.

6. CONCLUSION

In this report, we discuss the approaches to modelling energy consumption which provide the basis for deriving the suitable option for estimating the NGER energy consumption undercoverage. We also discuss a range of model specifications, among which the chosen model for the estimation task is a log-linear model with turnover and industry dummy variables as the main explanatory variables. The steps in applying the chosen model to NGER undercoverage estimation are also presented.

The estimated undercoverage of total industry energy consumption is 9.3% in 2009–10 and 8.0% in 2010–11. These estimates exclude energy consumption from Finance, Insurance and Superannuation Funds, Public Administration and Defence industries.

Our recommendations are:

- In the absence of energy price data, a logarithmic linear model with turnover and industry dummies as the main explanatory variables could be used for the estimation of the NGER undercoverage;
- In the future, when energy price data are available, both single and system equation approach could be considered;
- Energy expenditure data could also be used for energy volume estimation, although this approach does not directly model output-energy relationship.

Our investigation also suggests that the 2008–09 EWES data can be utilised for a pure energy demand analysis, which takes into account the impact of output, prices and other inputs on energy use. From this dataset, a measure of energy price ($ per GJ) can be derived which will facilitate the use of the system of (cost) equation approach. This can be an area of future research (see Appendix N for examples of future research ideas).

It may also be useful to re-run the model with the new waves of energy consumption data from future EWES surveys (the next survey is for the 2011–12 year). New data could be pooled with existing data or used in separate samples. Re-running existing models with new data helps to examine the impact from energy efficiency improvement or other structural changes.

REFERENCES

Australian Bureau of Statistics (2010a) *Australian Industry, 2008-09*, cat. no. 8155.0, ABS, Canberra.

—— (2010b) *Energy, Water and Environment Management, 2008-09*, cat. no. 4660.0, ABS, Canberra.

Berndt, E.R. and Wood, D.O. (1975) "Technology, Prices and the Derived Demand for Energy", *Review of Economics and Statistics*, 57(3), pp. 259–268.

Chambers, R. and Dorfman, A. (2003) "Transformed Variables in Survey Sampling", 2003 Joint Statistical Meetings – Section on Survey Research Methods.

Chandra, H. and Chambers, R. (2006) *Small Area Estimation with Skewed Data*, Southampton Statistical Sciences Research Institute Methodology Working Paper M06/05. (last viewed 20 February 2013) < http://eprints.soton.ac.uk/38417/ >

Christensen, L.; Jorgenson, D. and Lawrence, L. (1973) "Transcendental Logarithmic Production Frontiers", *Review of Economics and Statistics*, 55, pp. 28–45.

Coelli, T.; Rao, D.S.P.; O'Donnell, C.J. and Battese G.E. (2005) *An Introduction to Efficiency and Productivity Analysis*, Springer, New York.

Duan, N. (1983) "Smearing Estimate: A Nonparametric Retransformation Method", *Journal of the American Statistical Association*, 78(383).

Griffin, J.M. (1991) "Methodological Advances in Energy Modelling", *Energy Journal*, 14(1), pp. 111–123.

Haukoos, J.S. and Lewis, R.J. (2005) "Advanced Statistics: Bootstrapping Confidence Intervals for Statistics with "Difficult" Distributions", *Academic Emergency Medicine*, 12(4), pp. 360–365.

Karlberg, F. (2000) "Survey Estimation for Highly Skewed Population in the Presence of Zeroes", *Journal of Official Statistics*, 16, pp. 229–241.

Newman, M.C. (1993) "Regression Analysis of Log-Transformed Data: Statistical Bias and its Correction", *Environmental Toxicology and Chemistry*, 12, pp. 1129–1133.

Ryan, D.L. and Plourde, A. (2009) "Empirical Modelling of Energy Demand", in J. Evans and L.C. Hunt (eds), *International Handbook on the Economics of Energy*, Edward Elgar, Cheltenham.

APPENDIXES

A. SUMMARY STATISTICS OF ENERGY CONSUMPTION (TJ) FROM 2008–09 EWES (BEFORE DATA CLEANING)

Industry Division	*Mean*	*SD*	*Observations*
Agriculture, Forestry and Fishing	1.3	7.1	457
Mining	500.2	2,822.0	626
Manufacturing	377.5	3,945.9	2,721
Electricity, Gas, Water and Waste Services	10,230.4	38,740.2	206
Construction	7.8	75.5	1,017
Wholesale Trade	9.7	41.8	748
Retail Trade	39.4	365.4	647
Accommodation and Food Services	25.4	117.7	504
Transport, Postal and Warehousing	255.3	2,223.5	832
Information Media and Telecommunications	24.1	233.3	409
Financial and Insurance Services	5.7	65.2	679
Rental, Hiring and Real Estate Services	10.1	73.3	1,128
Professional, Scientific and Technical Services	4.6	24.0	975
Administrative and Support Services	2.0	10.0	728
Public Administration and Safety	0.9	4.2	138
Education and Training	4.0	7.6	419
Health Care and Social Assistance	10.5	32.1	835
Arts and Recreation Services	5.5	24.6	702
Other Services	3.7	28.5	633
Total			14,404

B. SUMMARY STATISTICS OF ENERGY CONSUMPTION (TJ) FROM 2008–09 EWES (AFTER DATA CLEANING)

Industry Division	*Mean*	*SD*	*Observations*
Agriculture, Forestry and Fishing	1.3	5.1	171
Mining	16.7	32.4	206
Manufacturing	7.0	19.5	1,550
Electricity, Gas, Water and Waste Services	1.5	3.5	49
Construction	1.9	9.7	325
Wholesale Trade	2.7	9.9	273
Retail Trade	3.1	7.5	293
Accommodation and Food Services	7.6	11.9	294
Transport, Postal and Warehousing	11.0	27.5	390
Information Media and Telecommunications	1.2	3.3	161
Financial and Insurance Services	1.4	6.2	174
Rental, Hiring and Real Estate Services	7.3	24.1	460
Professional, Scientific and Technical Services	1.6	8.0	390
Administrative and Support Services	1.9	11.1	286
Public Administration and Safety	0.3	0.5	57
Education and Training	3.1	5.0	202
Health Care and Social Assistance	3.4	7.7	338
Arts and Recreation Services	2.9	12.6	355
Other Services	1.7	9.4	328
Total			6,302

Note: These data exclude ABS profile units, part period operators, units with zero energy and/or turnover and units having total energy consumption greater than 200 TJ.

C. SUMMARY STATISTICS OF TURNOVER FROM 2008–09 BURE ($'000) (AFTER DATA CLEANING)

Industry Division	*Mean*	*SD*	*Observations*
Agriculture, Forestry and Fishing	5,962	14,524	171
Mining	23,698	48,343	206
Manufacturing	15,569	29,845	1,550
Electricity, Gas, Water and Waste Services	43,246	112,746	49
Construction	50,205	102,087	325
Wholesale Trade	103,793	180,374	273
Retail Trade	48,021	143,567	293
Accommodation and Food Services	13,997	18,302	294
Transport, Postal and Warehousing	26,591	70,746	390
Information Media and Telecommunications	17,531	34,586	161
Financial and Insurance Services	20,818	51,009	174
Rental, Hiring and Real Estate Services	14,452	34,964	460
Professional, Scientific and Technical Services	30,211	55,208	390
Administrative and Support Services	30,769	40,883	286
Public Administration and Safety	8,334	11,103	57
Education and Training	15,716	17,067	202
Health Care and Social Assistance	16,388	22,638	338
Arts and Recreation Services	9,984	32,771	355
Other Services	8,779	15,295	328
Total			6,302

Note: These data exclude ABS profile units, part period operators, units with zero energy and/or turnover and units having total energy consumption greater than 200 TJ.

D. CONVERSION FACTORS APPLIED TO THE EWES DATASET

D.1 Conversion factors used to convert EWES units to GJ

Fuel type	*EWES unit*	*Conversion factor*
Electricity	kWh	$\times 3.6\times10^{-3}$
Natural gas	MJ	/ 1000
Liquefied petroleum gas (LPG)	kL	× 26.2
Diesel	kL	× 38.6
Petrol	kL	× 34.2
Coke	t	× 27.0
Brown coal (a)	t	× 10.2
Coal by-products	t	× 37.5
Brown coal briquettes	t	× 22.1
Liquid biofuel	kL	× 23.4
Biogas	kL	$\times 37.7\times10^{-3}$
Bagasse	t	× 9.6
Wood and wood waste	t	× 16.2
Aviation turbine fuel	kL	× 36.8

Source: *Estimation of Undercoverage of the National Greenhouse and Energy Reporting system (NGER) due to Business falling under the Threshold*, unpublished ABS working paper, December 2010.

(a) For brown coal used in South Australian electricity generation, the conversion factor is 15.2 GJ/t.

For black coal, the energy content varies by type (black coal for electricity, black thermal coal, black metallurgical coal) and by source location:

D.2 Energy content of solid fuels (GJ/t)

	Energy content (GJ/t)		*Energy content (GJ/t)*
Black coal		Black coal	
New South Wales		Western Australia	
Exports		Thermal coal	19.7
– Metallurgical coal	29.0	Tasmania	
– Thermal coal	27.0	Thermal coal	22.8
Electricity generation	23.4		
Steelworks	30.0	Lignite	
Washed thermal coal	27.0	Victoria	9.8
Unwashed thermal coal	23.9	Briquettes	22.1
Queensland		South Australia	15.2
Exports			
– Metallurgical coal	30.0	Other	
– Thermal coal	27.0	Coke	27.0
Electricity generation	23.4	Wood (dry)	16.2
Other	23.0	Bagasse	9.6

Source: *Energy in Australia 2011*, BREE.

Programming codes to derive conversion factors for coal are based on the following steps:

1. Regardless of ANZSIC, if the unit is Western Australian use 19.7 GJ/t and Tasmanian use 22.8 (it all comes from the same place).

2. Otherwise look at the ANZSIC:

   - If ANZSIC is group 211 or 212 (ferrous metals production), use 30GJ as it is almost all metallurgical coal.

   - If the ANZSIC is subdivision 26 (electricity supply), use 23.4 except for the South Australian generator units, use 15.2.

   - For all other ANZSICs use 27.

E. SINGLE EQUATION MODEL (RAW SCALE MODEL)

Dependent variable: Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	2,994.8	2643.7
bure_turnover	9.24×10^{-7} **	3.86•10–7
ind020 (=reference)		
ind030	–1,179.7	3,612.0
ind041	–2,685.1	2,838.7
ind042	–2,975.3	4,204.5
ind051	–2,929.3	5,016.0
ind052	–2,582.8	2,990.5
ind060	16,026.3 ***	3,703.9
ind070	9,096.0 **	4,204.5
ind080	6,135.9 **	2,903.7
ind091	9,807.3 **	3,858.7
ind099	28,694.1 ***	4,283.3
…		
ind953	–673.7	3320.7
ind954	–1,702.8	3190.1
ind955	–1,464.6	2813.6
Adjusted R^2	0.0696	
Observations	9,176	

*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

F. MODEL WITH INTERACTION DUMMY VARIABLES (LOG SCALE)

Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	–3.5304 ***	
Log_bure_turnover	0.5932 ***	0.0334
LogTurnover × ind020 (=reference)		
LogTurnover × ind030	–0.0645	0.0434
LogTurnover × ind041	–0.0266	0.0352
LogTurnover × ind042	–0.0282	0.0677
LogTurnover × ind051	–0.1086 *	0.0641
LogTurnover × ind052	–0.0689 *	0.0362
LogTurnover × ind060	0.1559 ***	0.0418
LogTurnover × ind070	0.0741	0.0501
LogTurnover × ind080	0.1583 ***	0.0342
LogTurnover × ind091	0.2030 ***	0.0431
LogTurnover × ind099	0.1261 ***	0.0440
…		
LogTurnover × ind953	0.0729 *	0.0417
LogTurnover × ind954	0.0908 **	0.0375
LogTurnover × ind955	0.0007	0.0325
Adjusted R^2	0.4187	
Observations	6,302	

Note: LogTurnover × ind030 – LogTurnover × ind955 are the interaction terms between business turnover in log scale and industry dummy.
*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

G. MODEL WITH BOTH INTERCEPT AND INTERACTION DUMMY VARIABLES (LOG SCALE)

Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	–2.808	3.374
Log_bure_turnover	0.5440 **	0.2314
ind020 (=reference)		
ind030	–5.9519	6.9449
ind041	–4.2045	3.9501
ind042	–1.1274	8.0150
ind051	–14.9409	15.2054
ind052	3.1998	4.4979
ind060	5.7531	4.2144
ind070	–1.3347	5.5755
ind080	1.1495	3.6099
ind091	3.0592	5.5581
ind099	5.1629	4.7737
…		
ind953	–8.1209 *	4.8723
ind954	3.8890	3.9433
ind955	–1.9615	3.9509
LogTurnover × ind020 (=reference)		
LogTurnover × ind030	0.3144	0.4484
LogTurnover × ind041	0.2718	0.2748
LogTurnover × ind042	0.0553	0.6611
LogTurnover × ind051	0.8339	0.9614
LogTurnover × ind052	–0.2868	0.3081
LogTurnover × ind060	–0.2091	0.2836
LogTurnover × ind070	0.1612	0.3614
LogTurnover × ind080	0.0792	0.2480
LogTurnover × ind091	0.0106	0.3661
LogTurnover × ind099	–0.1791	0.3096
…		
logTurnover × ind953	0.6608 *	0.3462
LogTurnover × ind954	–0.1755	0.2711
LogTurnover × ind955	0.1287	0.2660
Adjusted R^2	0.4351	
Observations	6,302	

Note: *, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

H. MODEL WITH ANZSIC (2006) TWO-DIGIT INDUSTRY DUMMIES

Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	–4.1527 ***	0.4846
Log_bure_turnover	0.6371 ***	0.0124
ind02 (=reference)		
ind03	–1.1187 *	0.6647
ind04	–0.3780	0.5050
ind05	–1.0636 **	0.5247
ind06	2.5106 ***	0.6294
ind07	1.0888	0.7803
ind08	2.2982 ***	0.5006
ind09	2.6563 ***	0.5711
ind10	1.9405 ***	0.5239
…		
ind92	1.4951 **	0.6223
ind94	–0.4958	0.4985
ind95	0.2717	0.4689
Adjusted R^2	0.4108	
Observations	6,302	

Note: ind03–ind95 are the two-digit ANZSIC (2006) industry dummies.
*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

I. MODEL WITH TURNOVER AND ENERGY PRICE

Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	–4.6095 ***	0.3346
Log_bure_turnover	0.5894 ***	0.0088
Log_energy_price	–0.9092 ***	0.0112
ind020 (=reference)		
ind030	1.3110 ***	0.4574
ind041	0.5658	0.3510
ind042	0.0166	0.5792
ind051	0.3054	0.6922
ind052	–0.0631	0.3651
ind060	0.5476	0.4329
ind070	–0.6194	0.5363
ind080	0.6523 *	0.3448
ind091	1.2744 ***	0.4501
ind099	0.3383	0.4732
...		
ind953	–0.7594 *	0.4087
ind954	–0.0752	0.3777
ind955	–0.7778 **	0.3309
Adjusted R^2	0.7222	
Observations	6,301	

Note: Log_energy_price is the log scale of implicit price for energy, which is the ratio of total energy expenditure to total energy consumption volume.

*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

J. EXAMPLES OF MODELS FOR A SPECIFIC INDUSTRY DIVISION

J.1 Division C
Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	–3.6261 ***	0.3207
Log_bure_turnover	0.7031 ***	0.0211
R^2	0.4173	
Observations	1,550	

*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

J.2 Division D
Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	1.0238	1.5297
Log_bure_turnover	0.2999 ***	0.0974
R^2	0.1501	
Observations	49	

*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

K. EAS MODEL

Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*
Intercept	0.3259	0.4560
Log_EAS_turnover	0.6320 ***	0.0130
ind020 (=reference)		
ind030	–0.8694	0.6684
ind041	–0.3711	0.5077
ind042	–0.1483	0.8759
ind051	–1.8190 *	0.9963
ind052	–0.9902 *	0.5303
ind060	2.3540 ***	0.7102
ind070	1.4114 *	0.7479
ind080	2.1132 ***	0.6092
ind091	3.1184 ***	0.6474
ind099	2.3612 ***	0.6945
…		
ind953	0.9523	0.5879
ind954	1.5533 ***	0.5616
ind955	0.1909	0.4787
Adjusted R^2	0.4288	
Observations	5,818	

Note: *, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

L. EXPENDITURE MODEL

Dependent variable: Log of Total Energy Consumption (GJ)

	Coefficient	*Standard error*		*Coefficient*	*Standard error*
Intercept	1.3756 ***	0.3573			
Log_expenditure	0.9696 ***	0.0110			
ind020 (=reference)			ind211	2.1357 ***	0.4346
ind030	–2.5337 ***	0.5241	ind212	2.2012 ***	0.5018
ind041	–1.0844 ***	0.4023	ind213	1.7508 ***	0.4852
ind042	–0.4571	0.6640	ind214	1.9439 ***	0.5653
ind051	–2.2519 ***	0.7935	ind221	2.0407 ***	0.5530
ind052	–1.0074 **	0.4187	ind222	2.0882 ***	0.4198
ind060	2.2211 ***	0.4959	ind223	1.9417 ***	0.5084
ind070	1.9144 ***	0.5957	ind224	1.9954 ***	0.4958
ind080	1.8048 ***	0.3904	ind229	2.0340 ***	0.4346
ind091	2.1442 ***	0.5159	ind231	1.7315 ***	0.4295
ind099	2.0574 ***	0.5422	ind239	1.9560 ***	0.4066
ind101	2.0118 ***	0.5084	ind241	1.5385 ***	0.4853
ind109	2.1684 ***	0.4347	ind242	1.6716 ***	0.5018
ind111	1.9711 ***	0.4136	ind243	1.8222 ***	0.4681
ind112	1.8986 ***	0.5018	ind244	1.9631 ***	0.4646
ind113	1.9619 ***	0.4036	ind245	1.9924 ***	0.6370
ind114	2.0681 ***	0.4760	ind246	2.0533 ***	0.5084
ind115	1.6729 ***	0.5420	ind249	2.3204 ***	0.5957
ind116	1.9371 ***	0.4903	ind251	1.5082 ***	0.4405
ind117	1.8388 ***	0.4279	ind259	1.8277 ***	0.4312
ind118	2.2892 ***	0.4384	ind261	1.4200 **	0.6370
ind119	2.2927 ***	0.4427	ind263	0.4310	0.9825
ind121	1.8337 ***	0.4097	ind264	1.1473 *	0.6969
ind122	3.3307 **	1.6260	ind270	0.5451	0.6147
ind131	2.1904 ***	0.4958	ind291	–1.5376 *	0.7936
ind132	1.9762 ***	0.5084	ind292	–0.7127	0.5323
ind133	2.2035 ***	0.4682	ind301	–0.3211	0.3855
ind134	2.1998 ***	0.5795	ind302	0.0531	0.4448
ind135	1.8342 ***	0.4189	ind310	–0.8824 **	0.4253
ind141	2.2266 ***	0.4762	ind321	–1.8264 ***	0.4649
ind149	1.9300 ***	0.4346	ind322	–0.2255	0.5156
ind151	1.3497 ***	0.4804	ind323	–0.7243 *	0.4264
ind152	2.1340 ***	0.4551	ind324	–0.0156	0.4525
ind161	1.8780 ***	0.4311	ind329	–0.2090	0.5323
ind170	2.1498 ***	0.4328	ind331	1.2617 **	0.6369
ind181	2.4236 ***	0.7388	ind332	0.0137	0.4406
ind182	2.2473 ***	0.6146	ind333	0.1183	0.5322
ind183	2.3845 ***	0.4682	ind341	0.6244	0.6369
ind184	1.8965 ***	0.4364	ind349	0.3055	0.4295
ind185	1.8047 ***	0.4903	ind350	–0.0834	0.4404
ind189	2.2573 ***	0.8692	ind360	0.9102 **	0.4279
ind191	1.9467 ***	0.4448	ind371	0.7243	0.6370
ind192	2.0210 ***	0.4760	ind372	–0.5042	0.5957
ind201	2.0028 ***	0.5084	ind373	0.6726	0.4523
ind202	2.4539 ***	0.4852	ind380	–0.2030	0.4681
ind203	2.2211 ***	0.4346	ind391	0.4993	0.4448
ind209	1.8463 ***	0.4761	ind392	0.4336	0.5530

Expenditure model (cont.)

	Coefficient	Standard error		Coefficient	Standard error
ind400	0.9347 **	0.4612	ind664	1.7044 ***	0.6369
ind411	2.0679 ***	0.4405	ind671	1.3948 ***	0.3674
ind412	1.4669 ***	0.4646	ind672	1.0299 ***	0.3974
ind421	0.6443	0.5794	ind691	1.8592 ***	0.5018
ind422	0.8174 *	0.4613	ind692	–0.0713	0.3815
ind423	0.1722	0.5421	ind693	0.8234 **	0.4099
ind424	0.9541 *	0.5085	ind694	1.5088 ***	0.4581
ind425	1.2988 ***	0.4613	ind695	0.5034	0.7934
ind426	–3.0298 *	1.6262	ind696	0.1213	0.4116
ind427	0.9370 **	0.4551	ind697	0.8792	0.6639
ind431	0.5605	0.5420	ind699	1.3153	0.9824
ind432	–0.8289	1.6262	ind700	0.6763 *	0.3950
ind440	1.7252 ***	0.3833	ind721	0.2950	0.3812
ind451	1.5818 ***	0.3872	ind722	1.3155 ***	0.4426
ind452	1.8242 ***	0.4761	ind729	0.4573	0.4115
ind453	1.7486 ***	0.4228	ind731	–0.8129 *	0.4224
ind461	2.1430 ***	0.3948	ind732	1.3502 **	0.5531
ind462	2.1120 ***	0.4852	ind771	–0.7948 *	0.4175
ind471	2.1134 ***	0.7388	ind772	0.3989	0.7387
ind472	2.2213 **	0.9825	ind801	1.0251 **	0.5023
ind481	1.4729 **	0.6369	ind802	1.5956 ***	0.3955
ind482	1.9260 ***	0.6146	ind810	1.0008 **	0.4211
ind490	1.3112 ***	0.4959	ind821	0.9697 **	0.4237
ind501	1.7510 ***	0.4761	ind822	1.8921 **	0.7934
ind502	1.2999	0.8691	ind840	1.7464 ***	0.4346
ind510	2.1406 ***	0.4267	ind851	0.5484	0.4296
ind521	2.0289 ***	0.4472	ind852	1.8536 **	0.7934
ind522	1.8530 ***	0.5957	ind853	0.8652 **	0.4297
ind529	2.1439 ***	0.3915	ind859	1.5432 *	0.7934
ind530	2.2904 ***	0.4551	ind860	1.7451 ***	0.3959
ind541	0.6443	0.4646	ind871	1.0606 **	0.4166
ind542	2.1952 *	1.1768	ind879	0.7682 *	0.3994
ind551	0.9607 **	0.4385	ind891	1.4415 ***	0.4582
ind552	0.8621	0.7934	ind892	0.4955	0.4000
ind561	0.4009	0.5421	ind900	1.0565 ***	0.3872
ind562	0.7389	0.6969	ind911	0.9704 **	0.3919
ind570	1.0657 *	0.6146	ind912	1.0814 **	0.4720
ind580	–0.5040	0.5156	ind913	0.9225	0.5654
ind591	1.5140 **	0.6146	ind920	1.7045 ***	0.4903
ind592	1.4053 **	0.5530	ind941	–0.0841	0.4099
ind601	0.5463	0.5795	ind942	0.1592	0.4682
ind602	–0.9340	0.8693	ind949	1.0507	0.8692
ind641	1.1177 ***	0.3765	ind951	0.8096 *	0.4559
ind642	1.1769 **	0.5085	ind953	1.7649 ***	0.4649
ind661	–0.2804	0.5019	ind954	1.4448 ***	0.4295
ind663	–0.1264	0.3962	ind955	0.8248 **	0.3788
Adjusted R^2	0.6367				
Observations	6,389				

Note: *, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

M. ESTIMATION OF PROBABILITY OF NON-ZERO ENERGY CONSUMPTION

Dependent variable: Probability of non-zero energy consumption

	Coefficient		*Standard Error*
Intercept	–2.6088	***	0.4618
Log_bure_turnover	0.2469	***	0.0123
ind020 (=reference)			
ind030	–0.5412		0.5961
ind041	–0.9598	**	0.4573
ind042	–0.4035		0.6531
ind051	–0.7151		0.8115
ind052	–0.6109		0.4781
ind060	0.3348		0.6474
ind070	–0.5594		0.6952
ind080	0.4006		0.4865
ind091	0.4568		0.7103
ind099	1.4733		1.1320
…			
ind953	–0.0947		0.5435
ind954	0.4237		0.5459
ind955	–0.0421		0.4610
Adjusted R^2	0.3117		
Per cent Concordance	79.9%		
Observations	8,928		

*, ** and *** implies the coefficient is statistically significant at 10%, 5% and 1% level, respectively.

N. POSSIBLE RESEARCH EXTENSIONS

The analysis presented in this paper provides useful preliminary work to advance other related analyses, which may be relevant to many internal and external stakeholders given the current interests in carbon pricing, climate change, etc.

Economic analysis of industry energy demand using unit record data from the ABS Energy survey

For an analytical work aiming to provide further information on the impacts of prices and other production input factors on energy demand, a system of cost functions can be used to analyse demand elasticities. Measures of labour and energy prices can easily be derived from the current EWES linked to BURE dataset.

The constraint in the availability of capital and non-capital material prices will require some modification of the usual production economic framework. This study will showcase the modelling options and their impacts on results.

The outputs from this work such as price elasticities and elasticities of substitution between energy and other production inputs (capital, labour, materials) will be useful for examining the impacts of price changes on energy consumption or could be used in benchmarking the parameters used in other energy forecast models.

Energy demand through time: an economic analysis using time-series data from the ABS Energy Accounts and BREE Energy Statistics

Time-series energy consumption data at the industry level are available either through ABS Energy Accounts or BREE Energy Statistics.

Using time series data requires a different modelling approach compared to using cross-sectional unit record data. ABS's recent investigation into the economic literature on energy demand modelling has suggested several modelling methods for this type of data (e.g. Error Correction Model, Structural Time Series Model, asymmetric price model). This study will consider the application of these available methods to the available Australian data.

Analysis using time-series data will help to address the impact of improvement in energy efficiency on energy consumption. Possible outputs from this analysis are the impacts of technological change on energy consumption and long-run impacts of energy prices and other variables on energy demand. This will also help to provide a robust model for estimating energy consumption at the aggregate industry level, taking into account energy efficiency changes and long-run price impacts.

Measurement of energy efficiency using unit record data from ABS Energy survey

There is a vast array of data in EWES dataset that have not been utilised. Under the NGER project, ABS has merged EWES with BURE data to generate a bigger dataset of both energy/water consumption and financial data. Apart from utilising the same dataset for water consumption modelling purpose, the data can also be used to analyse other issues such as energy/water/environment management. More importantly, it may also be possible to derive measures of industry energy efficiency. Several approaches for measuring energy efficiency have been unveiled through the ABS's investigation of energy economic literature. The possible outputs are analyses of aspects of energy and environmental management, and measurements of energy use efficiency across industries (for example, are some industries doing better than others and why?).

As a next step from this project, which has only focused on a model for all industries and all fuel types, in this study energy demand or other analyses can also be done for each specific industry or a specific fuel type.

FOR MORE INFORMATION . . .

INTERNET	**www.abs.gov.au** The ABS website is the best place for data from our publications and information about the ABS.
LIBRARY	A range of ABS publications are available from public and tertiary libraries Australia wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our website for a list of libraries.

INFORMATION AND REFERRAL SERVICE

Our consultants can help you access the full range of information published by the ABS that is available free of charge from our website, or purchase a hard copy publication. Information tailored to your needs can also be requested as a 'user pays' service. Specialists are on hand to help you with analytical or methodological advice.

PHONE	1300 135 070
EMAIL	client.services@abs.gov.au
FAX	1300 135 211
POST	Client Services, ABS, GPO Box 796, Sydney NSW 2001

FREE ACCESS TO STATISTICS

All statistics on the ABS website can be downloaded free of charge.

WEB ADDRESS	www.abs.gov.au