

Guidance document on “CSE DIMPSA data”

1 Introduction

This document accompanies the ‘CSE DIMPSA data’ available to download from the Centre for Sustainable Energy website (www.cse.org.uk/domestic-energy-consumption-data).

The data and guidance document have been compiled and made publicly and freely available by the Centre for Sustainable Energy as part of its Open Data project funded by the Esmée Fairbairn Foundation.

This document fulfils two key functions:

- (i) To provide brief commentary explaining the methodology behind the dataset and how the figures were derived; and
- (ii) To outline some potential uses of the data.

2 The Dataset and ‘DIMPSA’

The Centre for Sustainable Energy (CSE) has developed an analytical model - DIMPSA (‘Distributional Impacts Model for Policy Scenario Analysis’) – which is now used under license by the Department of Energy and Climate Change for the Government’s own assessments of the distributional impacts of policies. DIMPSA enables the assessment of the impact of both the costs and the benefits of policies for domestic energy consumers.

Underlying this model is a comprehensive dataset based on the socio-demographically representative sample of UK households surveyed in the ONS Living Costs and Food Survey (LCF¹). Data from six LCF surveys has been combined (financial years 2004/5 to 2009), generating a sample size of over 36,000 cases.

2.1 Deriving energy consumption estimates from energy expenditure

The LCF survey uses interview and diary-based methods to collect information on a range of personal and household expenditure, including on household fuels (see Table 1). CSE has developed a methodology to derive estimates of annual household fuel consumption (in kWh) for each case (household) in the dataset using LCF household-level survey data on expenditure on all household fuels, combined with survey data on method of payment (for electricity and mains gas), survey month and year. A series of look-up tables containing time (month and year) and location (region) specific fuel price data are created for each fuel and, where applicable, method of payment. Electricity and mains gas prices are taken from the Department of Energy and Climate Change (DECC) data on average annual electricity and gas unit costs for selected towns and cities in the UK²

¹ In April 2001 the *Family Expenditure Survey* (FES) and *National Food Survey* (NFS) were combined to form the *Expenditure and Food Survey* (EFS), which completely replaced both series. From January 2008, the EFS became known as the *Living Costs and Food* (LCF) survey module of the *Integrated Household Survey* (IHS).

² “Average annual domestic electricity bills for selected towns and cities in the UK and average unit costs (QEP 2.2.3)” and “Average annual domestic gas bills for selected towns and cities in the UK and average unit costs (QEP 2.3.3)”, available at: www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics

and non-metered fuel prices are derived from Sutherland tables³. These prices are then adjusted to provide a monthly figure to reflect fuel price fluctuations over the year. Using the look-up tables, annual energy consumption can then be derived from the survey expenditure value for each fuel through a simple calculation:

$$\left\{ \begin{array}{l} \text{LCF survey household fuel expenditure data [£]} \\ \text{Unit price of fuel [£/kWh]} \end{array} \right\} * 52.14$$

Table 1: LCF survey expenditure variables used for deriving estimates of household fuel consumption

Variable name	Label
B175	Electricity amount paid in last account
C45114t	Electricity slot meter payment
C45112t	Second dwelling: electricity account payment
B222	Elec - Amt of payment (budget scheme)
B170	Gas amount paid in last account
C45214t	Gas slot meter payment
C45212t	Second dwelling: gas account payment
B221	Gas - Amt of payment (budget scheme)
B018	Bottled gas for central heating
C45222t	Bottled gas - other
C45411t	Coal and coke
B017	Oil for central heating - last quarter
C45312t	Paraffin
C45412t	Wood and peat
C45511t	Hot water, steam and ice

2.2 Accounting for different energy tariffs

The LCF survey does not provide information on energy tariff; hence the average unit cost of each fuel type has to be applied for the region and method of payment. However, the survey does provide details of central heating fuel. Where a household has electric central heating, the unit price of electricity has been adjusted, to allow for the likely lower unit rate (most households using electricity for heating will be on special tariffs, e.g. Economy 7). Without applying this reduction factor, we would risk underestimating electricity consumption in electrically-heated households.

2.3 Modelling household energy consumption from survey derived values

The LCF survey is continuous, with interviews being spread evenly over the year to ensure that seasonal effects are accounted for. In addition, the survey is weighted to reduce the effect of non-response bias and produce population totals and means. However, as noted above, the LCF survey collects detailed expenditure information through both the diary (which respondents fill in daily for a fortnight) and a household questionnaire (interview). These approaches to data collection combined with the rolling (monthly) nature of the LCF survey have important implications for how household

³ www.sutherlandtables.co.uk

fuel expenditure data is used. For example, expenditure on utilities is collected via different methods, depending on the payment method (electricity and mains gas expenditure by prepayment meter is recorded via the expenditure diary, whilst expenditure by other payment methods (including non-metered fuels) is collected through the interview⁴). As a result, we cannot assume that an annual estimate of consumption derived from the weekly expenditure value recorded in the LCF survey for any household fuels is in fact representative of how much that household consumes over the year.

Values at the individual case (household) level in the dataset are therefore not used in DIMPSA. Instead, CSE undertakes additional modelling, using CHAID⁵, to derive estimates of annual consumption for every individual case (household) in the dataset. CHAID is a classification method which seeks to identify optimal splits in categorical 'predictor' variables with respect to their influence upon a single dependent variable - in this case household energy consumption. CHAID results in clusters - or 'nodes' - of cases with similar defining characteristics and to which a predicted value for the dependent variable is assigned (the predicted value is equivalent to the mean of the dependent variable of all cases in the node). This modelling results in a more compressed distribution, but gives a reliable estimate of baseline energy consumption for every household in the dataset, whilst maintaining the original mean value for the dataset as a whole.

The resulting DIMPSA dataset, which has been used for the analysis and on which the data provided in the accompanying spreadsheet (*'CSE DIMPSA Data'*) is based, therefore includes extensive household level socio-demographic information (as collected by the LCF survey), including household income, and modelled annual household energy consumption derived from actual stated expenditure on all household fuels.

2.4 Data health warning

As described above, the data provided in the accompanying spreadsheet is based on CSE's DIMPSA dataset of household energy consumption, derived from ONS Living Costs and Food survey data on household fuel expenditure. Whilst CSE has not sought to reconcile these survey-derived and modelled UK totals for fuel consumption with those published in the Digest of UK Energy Statistics (DUKES), the survey modelled values for total demand have been compared with these, and with mean values from DECC's National Energy Efficiency Data frameworks (NEED) database. DIMPSA modelled values are within +/- 5% of NEED with a maximum range of -9% and +6% by income band.

3 Understanding the data

The spreadsheet of 'CSE DIMPSA data' that accompanies this guidance document provides data on:

Mean & Median Consumption Values and Income Data for the whole dataset and by disposable household income quintile and decile (Tab: "Summary statistics")

⁴ For example, see: Expenditure and Food Survey, 2007. Volume B, The Household Questionnaire. pp.80 – 89.

⁵ CHAID ('Chi-square Automatic Interaction Detection') is a popular analytic technique for performing classification or segmentation analysis. It is an exploratory data analysis method used to study the relationship between a dependent variable and a set of predictor variables. CHAID modelling selects a set of predictors and their interactions that optimally predict the variability in the dependent measure. The resulting CHAID model is a classification tree that shows how major 'types' formed from the independent variables differentially predict a criterion or dependent variable. CHAID analysis has the advantage that it enables more detailed scrutiny of the socio-demographics of households in each category, whilst maintaining a sufficient number of cases to give reliable estimates of scalar values.

Electricity consumption values and counts of households for Electricity Consumption Quintiles/Deciles and Income Quintiles/Deciles (Tab: "1_Electricity quintiles")

Gas consumption values and counts of households for Gas Consumption Quintiles/Deciles and Income Quintiles/Deciles for ALL HOUSEHOLDS (including non-mains gas users) (Tab: "2_Gas quintiles_All HHs")

Gas consumption values and counts of households for Gas Consumption Quintiles/Deciles and Income Quintiles/Deciles for GAS HEATED HOUSEHOLDS only (Tab: "3_Gas quintiles_Gas HHs only")

Combined (Electricity + Gas) consumption values and counts of households for Combined Consumption Quintiles/Deciles and Income Quintiles/Deciles (Tab: "4_Combined (Elec+Gas) quintiles")

There are several key points to be aware of in order to understand and interpret this data.

3.1 Sample sizes

The accompanying spreadsheet shows two different values for number of households: the 'unweighted count' and the 'count'. The former is a count of the actual number of cases in the sample (i.e. the number of households actually surveyed in the LCF survey (around 6,000 – 8,000 individual households every year, although recall we have combined data from six LCF survey years to increase the sample size).

The latter ('count') is the weighted count. This shows the number of households that are represented by the individual surveyed households. The LCF survey is provided by ONS with an annual weight to be representative of the population in the survey year.

The (weighted) DIMPSA dataset has been divided into fifths and tenths to create quintiles and deciles of the population respectively. This has been done for disposable household income data (creating income quintiles and deciles) and electricity, gas and combined (electricity + gas) consumption data (creating consumption (kWh) quintiles and deciles).

Cross-tabulating energy consumption deciles and income deciles results in relatively small sample sizes (i.e. 100^{ths} or 'percentiles' of the dataset). Percentiles containing less than 100 cases (unweighted count) are highlighted in the spreadsheet and should be treated with some caution.

3.2 Gas consumption

Some 17% of households in the DIMPSA dataset do not have a gas supply and/or do not use mains gas and therefore have zero gas consumption. As a result, it is not possible to create equal sized deciles of gas consumption. Instead, the first "decile" contains 17% of households, all of which have zero consumption, and the second "decile" contains only 3% of households. It is therefore advisable to analyse gas consumption for only those households that use mains gas (some 20.3 million households (weighted count) in the DIMPSA dataset). This data is provided in Tab 3 of the spreadsheet.

4 Potential uses of the data

The summary data presents an overview of the distribution of energy consumption for households in Great Britain. This provides a useful starting point for analysis of the implications of current energy and policy costs for domestic consumers. For example, a unit cost of energy or specific tariff structure could be applied to the annual kWh estimates provided in the spreadsheet to calculate a total energy cost for that group of consumers.

Some further studies and sources of fuel price data and potential applications of the data provided in the accompanying spreadsheet are detailed below.

- CSE and ACE, (2012). *Impact of future energy policy on consumer bills*. Report to Consumer Focus. The Annex to this report provides a detailed breakdown of the component costs of a unit of energy between 2010 and 2020. See: www.consumerfocus.org.uk/files/2012/12/Impact-of-future-energy-policy-on-consumer-bills-ACE-CSE.pdf
- DECC publishes annual estimates of gas and electricity bills, along with statistics on household expenditure on fuel. These estimates include unit prices by region and payment method. However, the individual proportions of each element of a consumers' bill is not detailed by DECC. See: www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics
- Ofgem publishes more up to date (in April for each year) figures on total energy costs for dual fuel, electricity and gas customers. However, this does not include an assumed unit cost. See www.ofgem.gov.uk/Markets/RetMkts/rmr/smr/Pages/indicators.aspx
- Consumer switching sites. There are now numerous consumer switching sites which allow you to compare tariffs and energy bills. By entering the consumption data from DIMPSA with complimentary assumptions (e.g. region, payment method etc.), you will be able to obtain some real tariff data to calculate a final energy cost.

It is important to note that any analysis of energy costs and policy impacts looking into the future would need to allow for the impact of Government policy, which assumes that consumption in 2020, for example, will be lower than it is today. The above summarises some useful sources of data on energy and policy costs. DECC do provide long-term energy use⁶ projections which provide an indication of the Government's assumed impact of policy; however, more detailed modelling is required to examine these changes across the income distribution.

⁶ www.gov.uk/government/publications/2012-energy-and-emissions-projections