An End-Use Intensity Study of the Residential Sector

Mark Rebman and Min Yu, BC Hydro Power Smart

ABSTRACT

Forecasting and managing residential electricity demand usually entails the use of consumption data for specific end-uses like space heating, water heating, lighting and other appliances with significant impacts on load. Conventional methods use some form of *Conditional Demand Analysis* (CDA) or engineering model. Both approaches assume linear relationships between consumption and the number of appliances in use and cannot use straight-forward statistical modeling to disaggregate the space and water heating components of consumption. Technical estimates of end-use values also depend on very general suppositions about differences between inside and outside temperature, the rate of solar gain and heat loss. Such assumptions and dependencies can be shown to introduce unwanted and problematic sources of error.

To address some of these issues, an end-use intensity analysis was conducted using end-use survey and billing data from 3,621 residential households across the province of British Columbia. The main goals were to estimate annual energy consumption for typical homes and disaggregate space and water heating consumption for different types of homes without recourse to technical data on particular end-uses. This entailed a general linear model of annual residential energy consumption with region, building type, heating type and water heating type as categorical variables and the linear or polynomial terms of end-use holdings as covariates reflecting appliance consumption.

Socio-economic variables such as the size and age of the building structure and persons per household were also included to analyze the consumption effects of social strata. Least square means – the predicted marginal means over a population – were used to estimate different factor levels in the final model. Space and water heating consumption was then decomposed for different type of homes based on simple contrasts of the least square means.

As this model can interpret electricity consumption by disaggregating space heating and water heating for different types of families, it can be a practical alternative to more traditional approaches.

Introduction

Forecasting residential electrical demand usually requires specific consumption data for end-uses like space heating, water heating, lighting and other appliances. This information allows the isolation and identification of electrical products that have the most significant impact on load. The two standard methods of estimating end-use demand are (1) direct metering for individual end uses and (2) Conditional Demand Analysis (CDA). Direct metering of individual appliances generally provides more precise estimates, but is costly. The cheaper, more traditional CDA uses linear regression to disaggregate billing data into particular end use values in terms of average unit electricity consumption (UEC).

CDA was first introduced by Parti and Parti in 1980. Bartels and Fiebig (1990, 2000), Bauwens et al. (1994), and Hsiao et al. (1995) conducted further studies to improve the method. Yet despite its low cost, the reliability of CDA depends on some questionable assumptions. First, the consumption level of each customer is assumed to be linearly associated with actual counts of appliances in use. This is not true for some end-uses. For example, loads associated with refrigerators, stoves, dishwashers, clothes-washers and personal computers vary non-linearly with the count of such appliances by household. Secondly, engineering formulae that estimate space heating and water heating are based on very general suppositions concerning the differential between inside and outside temperature as well as the rate of solar gain and heat loss. Thirdly, appliance ownership saturation rates can also be problematic: if a rate is very high (close to 100%), then the UEC of the associated product cannot be isolated; if a rate is very low, then the estimate is frequently unreliable. Finally, the traditional CDA does not always consider important socio-economic variables such as persons per household or the size and age of the building structure.

These assumptions make the estimation of error difficult and can render results unreliable by introducing compound error. Significant numbers of involved engineering formulae populate the functional forms of CDA models without any real accounting in the model itself for the possible extraneous error associated with them[1].

Variables based on combinations of values and variables can resist direct interpretation. For example, suppose that the surface area of a residential structure is modeled as a function of total floor area to the power of 0.5 - an assumed elasticity of surface area with respect to the amount of square footage. This estimate is therefore exogenous to the data[2]. Further suppose that one of the CDA parameters to be estimated is associated with a variable equal to the surface area times heating degree days times income times a dummy variable for electric space heat. Such variable specification introduces additional assumptions and calculations that can increase uncertainty and error.

Any of these problems can adversely affect the accuracy of a CDA model. The purpose of this study is to avoid these pitfalls by fitting a linear model that directly estimates and predicts the annual electricity consumption of average residential households. The model is based on categorical variables such as geographic region, space heating fuel, water heating fuel, existing building structure plus the presence of other electrical appliances such as refrigerators, stoves, dishwashers, clothes-washers, clothes-dryers, light bulbs, televisions, and personal computers.

The unbalanced nature of the design is addressed by using a technique known as least square (or adjusted) means. A brief explanation of the approach is offered in the Energy Decomposition Section of the paper.

Objectives

The formal objectives of this study are to (1) estimate annual energy consumption for a typical home given its region, building type, space heating type, water heating type, number of residents, home size (floor space), age of building and quantity of other specific electric appliances, and (2) disaggregate consumption for space heating and water heating for average families living in different types of buildings and regions.

---

[1] An example of how such formulae are applied in CDA may be found in Tiedemann and Kelly (2006). In their very thorough paper, the authors estimate at number of parameters based on combinations of terms like Heating Degree Days, Minutes of Sunshine, Area of Household or Income. Since some of these terms – such as minutes of sunlight or heating degree days – are highly general values, they can introduce additional sources of error.

[2] This example is taken from Kelly and Tiedemann (2006).

Data Description and Exploration

To investigate typical home characteristics and features, a disproportionate stratified random sample was drawn in 2006 from the Lower Mainland, Vancouver Island, Southern Interior and the Northern regions of the BC Hydro service area. 4,338 residential households[3] were successfully surveyed by housing type, age of home, size of home, space heating, water heating, and other electric appliances. Space heating fuel was further categorized as either *primary* (principal source) or *secondary* (backup source). The number of appliances and persons per household as well as the age category of buildings are ordinal counts with building age limited to 5 levels – although to simplify modeling, these variables are treated as continuous covariates.

Traditional CDA assumes that consumption for individual households is linearly associated with the count of electrical appliances. The data demonstrates that associations between energy consumption and appliance counts are actually more complex with consumption patterns for some appliances taking polynomial form. Modeling the non-linear relationships improves fit and suggests that (1) two refrigerator households consume more than those with three or more, (2) two stove households consume more than those with three or more and (3) one dishwasher or clothes washer households consume more than those with two or more. End-use consumption values for personal computers also vary non-linearly with the count of units per household although the relationship is complex and may result from sampling issues or household characteristics not captured in the survey[4].

The data also establishes that electricity consumption climbs with the age of the building; older homes consume more on average and consumption takes quadratic form when associated with the count of persons per household. A variable indicating family income level was eliminated from the model as it had no statistically significant relationship to energy consumption – a reasonable expectation given that electricity is affordable to even low income families in British Columbia[5]. The final model thus considers the square footage of the home, the count of persons per household and the age category of the home.

All modeling and data manipulation was performed using the SAS statistical package.

Model Description

To compare the effects of different regions, type of space heating, water heating and building on total electricity consumption, four categorical variables were coded with a number of model covariates. A full model is then fitted to annual energy consumption by building type, heating type, water heating type, interactions between these four factors, linear and polynomial terms of appliance holdings, the number of people per home and the building age levels. Following t-tests of each parameter estimate, the full model is reduced to:

---

[3] Missing values reduced the usable sample to 3,621.

[4] The fact that some of these results appear counter-intuitive should be sufficient grounds for caution in assuming any linear relationships.

[5] The number of persons per household, building square footage and location may be interpreted as proxy variables for income. Since the number of persons and square feet of living space per home are already included, the absence of a specific income variable is not a serious drawback to the model.

Energy = Intercept + region + building + heating + water heating + building heating+ building*water heating +* β_0 *fridge +* β_1 *fridge*fridge+* β_2 *stove+* β_3 *stove*stove +* β_4 *dishwasher +* β_5 *dishwasher*dishwasher +* β_6 *washer +* β_7 *washer*washer +* β_8 *dryer +* β_9 *bulb +* β_{10}*TV +* β_{11}*PC +* β_{12} *PC*PC +* β_{13} *square footage +* β_{14}*people+* β_{15} *people*people +* β_{16} *(age category of building)*

Region, building type, heating type and water heating type are the *main factor* effects. Each factor level is estimated by comparing energy consumption with *baseline* (the last level of each factor). For example, as the last level of the region effect, the Northern region is baseline and dummy-coded as zero. The effects of other regions are then individually estimated as the mean of energy consumption in the region minus the mean for the Northern Interior.

With this model, annual electricity consumption is fairly simple to estimate by specific home type. Although too extensive to present here, the analysis of variance (ANOVA) results suggest the following interactions[6]: (1) building with heating effect (implies that heating varies with buildings type) and (2) building with water heating effect (implies that water heating varies with building type)[7].

Note that parameter estimates associated with the main effects may not reflect realistic consumption values associated with appliances and other variables. In the approach used for this model we are primarily interested in statistically significant *interactions* – those special cases where one variable changes in tandem with another. Main effects are not easily interpretable. Many are only included for the dual purpose of reducing overall error and ensuring model stability.

In general, if the p-value of a specific term in the model is greater than 0.05, it is deemed not to contribute significantly to consumption and is removed from further consideration unless contained in a statistically significant interaction term. In the latter case, all main effects factors associated with the interaction remain in the reduced model regardless of their own particular p-values[8]. Some statistically insignificant terms are also retained to ensure model stability and the statistical significance of interactions of interest[9].

In summary, parameters associated with the number of dishwashers and of clothes dryers do not significantly affect total electricity consumption whereas space heating, water heating and all the other electric appliances do.

Model Validation

In traditional CDA models, total electricity consumption is normally estimated by a linear model with appliance holdings and a few household characteristics as explanatory variables. Theoretically, the results of such a model are reliable if the following assumptions apply: (1) Each observation is independent, and (2) dependent variables and their residuals are normally distributed with constant variance. If either assumption is violated – whether by serial

---

[6] Interactions show possible relationships between two or more variables in the model. Since they can show how socio-economic variables change in association with (for example) space and water heating type, they hold the greatest research interest in this instance.

[7] The ANOVA table may be obtained by requesting a copy from the authors. It is contained in an earlier, longer version of the paper.

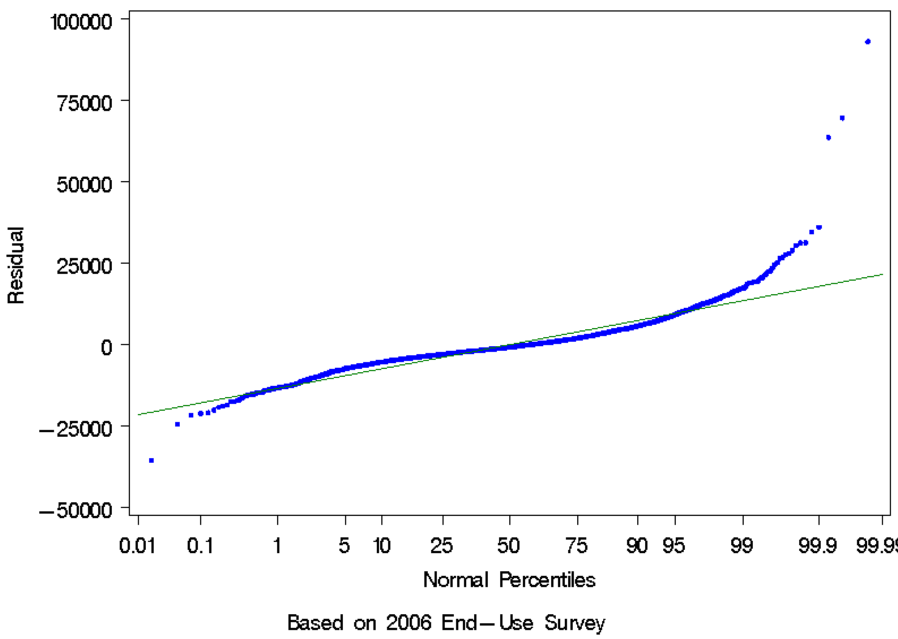
[8] Retention of all main effects terms associated with interactions is required to ensure stability of the model; they are not necessarily directly interpretable as unit consumption values.

[9] Removal of statistically weak terms sometimes affects measurement of other parameter of interest.

correlation, heteroskedasticity or non-normality – the forecasts, confidence intervals, and economic insights yielded by the regression model may be (at best) inefficient or (at worst) biased[10].

A number of tests were therefore applied to validate model assumptions. Actual annual consumption was compared with predicted values to test the model's performance. But as the model estimates mean household consumption, extreme values may be considerably understated; model validity was thus restricted to estimating the consumption of average (or near average) households. The normal Quantile-Quantile (QQ) plot was then used to test the normality of data and showed that residuals of the fitted model approximate a normal distribution except for a few extreme values (see Figure 1 below). These extremes were retained in the model as most cannot be confirmed as "true" outliers. Those that seemed like good candidates for removal such as households with swimming pools and hot tubs[11] had virtually no impact on mean values.

Figure 1: Normality (QQ Plot) of Model Residuals



The plot of residuals versus predicted energy consumption showed that variance expands as the predicted values increase. Although this suggests that an appropriate transformation of a dependent variable might be needed to ensure constant variance and model fitness, a variable transformation complicates the interpretation of the model. Since no transformation could be found to improve the model, for practical reasons the problem was set aside[12].

---

[10] For example, if consumption data consists of repeated measures in time series, serial correlation could be a problem. Since this analysis uses *annual* energy consumption, the issue does not arise.

[11] For example, this class of customer represents less than 0.15% of the total sample.

[12] A number of transformations were attempted. Since even the best transformation of the dependent variable (square root) could not improve overall predictive power or reduce variance inflation, the original model was retained for the sake of simplicity and interpretation.

Energy Decomposition Method

In this section, space heating and water heating consumption are decomposed for typical households. Generally, single family detached houses have more living space, electric appliances and occupants than other types of homes. On average, single family detached houses date to 1985, duplexes and townhouses to 1991-92, and mobile homes and apartments to 1988-89. Households with average appliance saturation rates were used to decompose space heating and water heating by the various groupings.

To decompose space and water heating consumption, adjustments are made for other variable impacts (regions, buildings, appliance holdings, square footage, age of building, and the number of people per household). This is accomplished with a technique known as *least square means*[13]. Least square means (sometimes called adjusted means) are predicted values from a multiple regression equation that contains both categorical predictors (factors) and numerical predictors (covariates). They are estimated by applying the mean value of a model covariate to estimate the mean response for all combinations of factors in the model and taking simple means of these estimates over factor levels. This estimation approach avoids any confounding of categories through sample imbalances[14]. To perform the analysis, estimates of annual consumption are expressed in the following functional form:

Electricity = ƒ (region, building, space heating, water heating, refrigerator, stove, dishwasher, washer, dryer, bulb, TV, PC, heating degree days, square footage, number of people, age of building) .

By way of example, let lsmeans[15] (REE/SFD, LM) denote the least squares mean of residential electric space heating (REE) for single family detached homes (SFD) in the Lower Mainland (LM) with average end-use saturation rates. This is the marginal mean of REE for SFD in the Lower Mainland (LM) for two models (electric and non-electric water heating).

Since the space heating effect interacts with building, and water heating interacts with the building effect, the heating effect must be estimated for homes with different building and water heating characteristics. Similarly, water heating effects must be estimated for different types of space heating and building structure. This requires the following least square means calculations:

1. Lsmeans{REE/[building(1-5)[16], water heating(electric/non-electric)]}
2. Lsmeans{REN[17]/[building(1-5), water heating(electric/non-electric)]}
3. Lsmeans{electric water heating/[building(1-5), space heating(1-3)[18]]}, and
4. Lsmeans {non-electric water heating/ [building (1-5), space heating (1-3)]}.

---

[13] The term derives from *least squares* regression.

[14] For example, if the sample contains imbalances between building type and end-use, the marginal (sample) means may produce incorrect values. This is because *marginal* means are weighted on the basis of sample size while *least square* means result from applying the mean value of some covariate to estimate the mean response for all combinations of the factors and taking simple means of these estimates over factor levels.

[15] *Lsmeans* is a **SAS** reserve word used in calculating least square means.

[16] Building categories 1 to 5 represent Single Detached Houses, Duplexes, Row/Townhouses, Mobile Homes and Apartments.

[17] REN designates non-electric space heating; REE stands for electric space heating.

[18] Space heating categories 1 to 3 are Primary, Secondary and Non-electric.

The effect of electrical space heating can be estimated by the comparison of the least squares mean of electric space heating and non-electric space heating (REE/SFD, electric water heating) which equals the least square means of (REE/SFD, electric water heating) minus least square means of (REN/SFD, electric water heating). Since the model does not distinguish between homes using electricity as a main or secondary fuel, some type of home have quite small estimates. Of building types considered, mobile homes were found to consume the most electricity.

Electric water heating consumption is lowest in the Lower Mainland and of building types considered, single family detached houses consume the most electricity for water heating. The base (energy consumption of electric end-uses plus miscellaneous consumption) can be estimated by the linear model in section 4.2 for typical homes using estimates of base, primary, secondary space heating (Table 1 and Table 2), and water heating (Table 3).

Note that there are some differences between raw and aggregation results – the latter of which are found in Tables 1, 2 and 3 below. This is explained by the fact that aggregation estimates are approximations and cannot be perfectly compared to raw data results since the latter describe consumption means based on the unadjusted (unbalanced) sample while the former describe means for typical homes based on least square means adjustments.

Table 1: Primary Electric Space Heating Consumption (kWh/year)

Building Type	Water Heating Type	Lower Mainland	Vancouver Island	Southern Interior	North	Average
Single detached house	Non-electric	4731	5373	5540	5393	5259
	Electric	6320	6963	7130	6982	6849
Duplex	Non-electric	5271	5913	6080	5932	5799
	Electric	6860	7502	7669	7522	7388
Row/townhouse	Non-electric	3513	4156	4323	4175	4042
	Electric	5103	5745	5912	5764	5631
Apartment	Non-electric	1046	1688	1855	1707	1574
	Electric	2635	3277	3444	3297	3163
Mobile Home	Non-electric	3727	4370	4537	4389	4256
	Electric	5317	5959	6126	5979	5845
Average	Non-electric	3658	4300	4467	4319	4186
	Electric	5247	5889	6056	5909	5775

Table 2: Secondary Electric Space Heating Consumption (kWh/year)

Building Type	Water Heating Type	Lower Mainland	Vancouver Island	Southern Interior	North	Average
Single detached house	Non-electric	309	951	1118	971	837
	Electric	634	1276	1443	1296	1162
Duplex	Non-electric	645	1287	1454	1306	1173
	Electric	970	1612	1779	1632	1498
Row/townhouse	Non-electric	130	244	411	264	262
	Electric	455	569	736	589	587
Apartment	Non-electric	32	146	313	165	164
	Electric	357	471	638	491	489
Mobile Home	Non-electric	812	1454	1621	1474	1340
	Electric	1137	1779	1946	1799	1665
Average	Non-electric	385	817	984	846	755
	Electric	711	1142	1309	1161	1081

Conclusions

Although some functions are quadratic and problematic to interpret, the simple model estimates some important residential consumption. Main results show that space heating and water heating consumption vary across building types and regions and that the electrical consumption of typical homes can be estimated, predicted and subsequently employed to forecast, and manage residential electrical load. This is accomplished without recourse to suppositions concerning hours of sunlight, heating degree days and other physical values. Nor is there any need to create additional variables based on combinations of values and variables that may resist direct interpretation.

The approach used here is also less time consuming than typical engineering models and reduces the need for highly detailed information on household behavior and appliance technology obtained from costly end-use metering – at least in the cases of space and hot water heating. With improved survey design, parameter estimation should also produce more significant results for specific end-uses.

Table 3: Electric Water Heating Consumption (kWh/year)

Building Type	Water Heating Type	Lower Mainland	Vancouver Island	Southern Interior	North	Average
Single detached house	Primary	4888	5530	5697	5550	5416
	Secondary	3624	4266	4433	4286	4152
	Non-electric	3299	3941	4108	3961	3827
Duplex	Primary	2486	3128	3295	3148	3014
	Secondary	1222	1864	2031	1884	1750
	Non-electric	897	1539	1706	1558	1425
Row/townhouse	Primary	2754	3396	3563	3416	3282
	Secondary	1490	2132	2299	2152	2018
	Non-electric	1164	1807	1974	1826	1693
Apartment	Primary	1148	1790	1957	1810	1676
	Secondary	823	1465	1632	1485	1351
	Non-electric	1148	1790	1957	1810	1676
Mobile Home	Primary	2538	3181	3348	3200	3067
	Secondary	2213	2856	3023	2875	2741
	Non-electric	2412	3054	3221	3074	2940
Average	Primary	3268	3911	4078	3930	3797
	Secondary	2004	2647	2814	2666	2533
	Non-electric	1679	2322	2489	2341	2208

References

Bartels, R., and Fiebig, D. G. (1990). Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads. Energy Journal 11(4), 79-97.

Bartels, R., and Fiebig, D. G. (2000). Residential End-Use Electricity Demand: Results from a Designed Experiment. Energy Journal. 21(2), 51-81.

Bauwens, L., Fiebig, D. G., and Steel, M. F .J. (1994). Estimating End-use Demand: a Bayesian Approach. Journal of Business and Economic Statistics 12(2), 221-231.

Hsiao, C., Mountain, D. C., and Illman, K.H. (1995). A Bayesian Integration of End-Use Metering and Conditional-Demand Analysis. Journal of Business and Economic Statistics 13(3), 315-326.

Kelly, J., and Tiedemann, K. (2006). Conditional Demand Analysis of Residential Energy Consumption. Vancouver: BC Hydro.

Larsen, B. M. and Nesbakken, R. (2004). Household electricity end-use consumption: results from econometric and engineering models. Energy Economics 26, 179-200.

Parti, M., and Parti, C. (1980). The Total and Appliance-Specific Conditional Demand for Electricity in the Household Sector. The Bell Journal of Economics 11(1), 309-321.Use “References” style here)