Energy Economics 45 (2014) 166–171

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Estimating base temperatures in econometric models that include degree days ☆

James Woods [a], Cody Fuller [b]

[a] Department of Economics, Portland State University, P.O. Box 751, Portland, OR 97207-0751, United States
[b] Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, PO Box 751, Portland, OR 97207-0751, United States

ARTICLE INFO

Article history:
Received 8 November 2013
Accepted 14 June 2014
Available online 3 July 2014

JEL classification:
Q47
C29
C51
L94
L95
D19

Keywords:
Nonlinear models
Evaluation
Errors-in-variables
Weather statistics
Billing analysis
Demand forecasting

ABSTRACT

Heating and cooling degree days are in common use in conditional demand models, billing analysis, and large-scale energy forecasts. The implications of either choosing an ex-ante base temperature, or scanning over the base temperatures, as suggested in Fels (1986) and recommended by some evaluation protocols, are infrequently considered. These procedures result in biased estimates of weather-driven loads because of correlated errors-in-variables, and impart a downward bias to the variance of those estimates by a factor of two. A non-linear estimation procedure that corrects for these biases and an ex-ante correction factor for evaluating prior evaluations are offered.

1. Introduction

Since the publication of “PRISM: An introduction” Fels (1986), the usual procedure for including weather in PRISM-style models, which includes conditional demand analysis (CDA), billing analysis, and large-scale energy forecasts, is to explain kWh with heating and cooling degree days (HDD and CDD), often in conjunction with other non-weather variables. The base temperature of these weather summaries, often thought of as the set-points for the heating or cooling equipment, is either assumed and set at a standard level, most often 65 °F, or estimated through the scanning procedure described in Fells.

There have been previous attempts to alter the basic structure of how HDD and CDD are calculated and used in a regression context. Some fields have moved to other approximations that use the same weather data. For example, ecological fields frequently use sine-wave based approximations Allen (1976). There have been many attempts to shift from the average temperature, defined as the average of the daily high and low, or account for rounding error and the differences between the hourly and daily average measurements. In almost all cases, the base temperature is treated as being ex-ante known or estimated with ad-hoc, out-of-model, methods.

None of these approaches deal with the fundamental problem of choosing what amounts to a regression parameter, the CDD and HDD base temperatures, outside a regression and then ignoring the consequences of this non-standard estimation procedure.

This paper will take the interpretation of the base temperature in CDD and HDD as a thermostat setting as a given, and show that there are significant errors-in-variables problems when using CDD and HDD in a regression context — the errors in degree day estimates are negatively correlated. The consequences of negatively correlated errors in variables will be discussed but the theoretical results are limited. The consequences must be demonstrated empirically.

Section 4 will introduce a formulation of a PRISM style model that allows for within-model estimation of heating and cooling degree base temperatures. The results of this model and the standard method of sweeping through base temperatures for the highest R^2 will be

☆ This paper was prepared as the result of work sponsored by the California Energy Commission. It does not necessarily represent the views of the Energy Commission, its employees or the State of California. The Energy Commission, the State of California, its employees, contractors and subcontractors make no warrant, express or implied, and assume no legal liability for the information in this report; nor does any party represent that the uses of this information will not infringe upon privately owned rights. This paper has not been approved or disapproved by the California Energy Commission, nor has the California Energy Commission passed upon the accuracy or adequacy of the information in this report.

E-mail addresses: woodsj@pdx.edu (J. Woods), fullercd@pdx.edu (C. Fuller).

http://dx.doi.org/10.1016/j.eneco.2014.06.006
0140-9883/

compared in Section 4.1. In the sample of houses used, the parameters associated with CDD and HDD are biased, but more importantly the variance of those parameters is understated by a factor of two. Section 5 summarizes the insights and gives a simple procedure for re-assessing the significance of weather driven loads with special implications for past program evaluations that depend on billing analysis.

2. The scale of degree day errors

HDD is defined as $\sum max(0, b - A_t)$ and CDD is defined as $\sum max(0, A_t - b)$. In both cases A_t is the average daily temperature, the average of the daily high and low, and b is the base temperature, commonly interpreted as the average cooling and heating set-point of the household.

Increases in the base temperature reduce CDD proportionally to the number of days in the period that has average temperatures above the base. Similarly, when the base temperature is decreased HDD falls proportionally to the number of days average temperatures below the base. When heating and cooling degree days share a common base, the sum is at a minimum where there are an equal number of days with temperatures above and below the base.

The usual relationship between heating and cooling degree days as the common base temperature ranges from 50 °F to 90 °F is shown in Fig. 1.

However, there is a very clear inverse relationship between CDD and HDD when the base temperature is allowed to vary. This inverse relationship is also apparent when the base temperatures are not constrained to be the same since it is commonly assumed that the CDD base is higher than the HDD base.

Measured heating and cooling degree days can change drastically with changes in the base temperature. Fig. 2 shows how CDD changes for a single month as the base temperature changes.

The figure shows that CDD is 442 with a base of 50 °F, and it falls to zero as the base rises to 70 °F. Monthly CDD changes by about 20 CDD for every one degree change in base temperature. A small amount of uncertainty in the base temperature can result in drastic changes in CDD equivalent to the difference between a June in San Francisco and a June in San Diego.

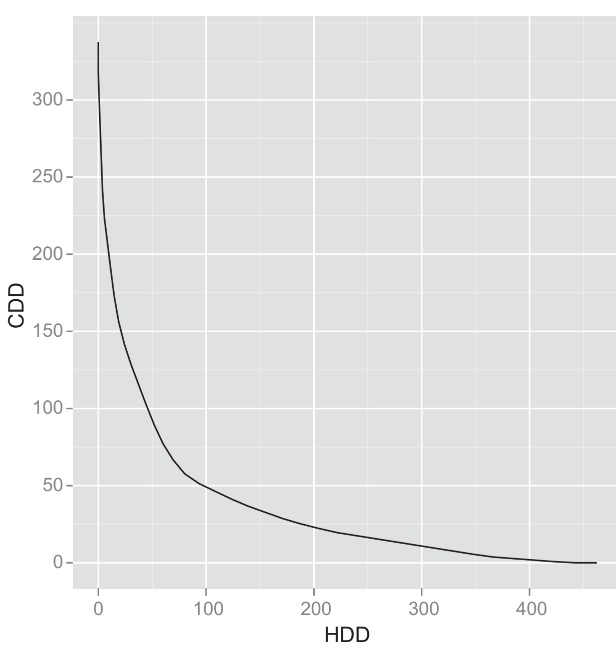

Fig. 1. Relationship between HDD and CDD for simulated weather and base temperatures from 50 °F to 90°.

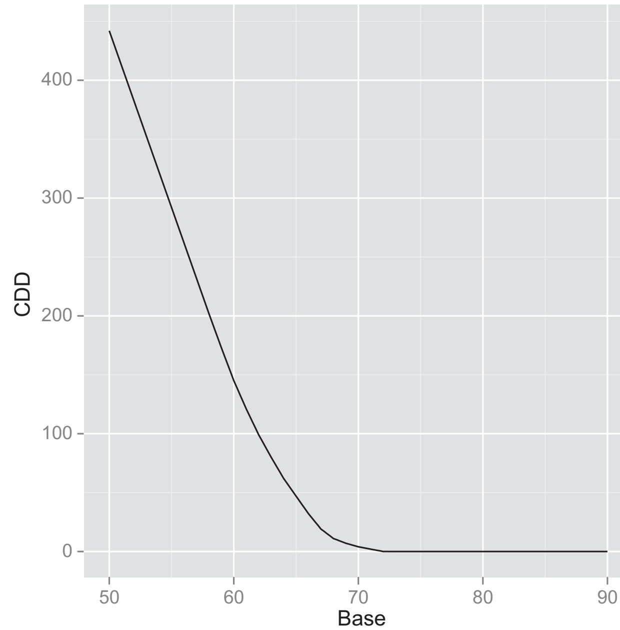

Fig. 2. CDD by base temperature for June 2000 in San Luis Obispo.

Uncertainty about the base and the resulting uncertainty in degree days are not ignorable. Simply choosing a default temperature a priori is a significant source of errors in explanatory variables.

The problem of errors in explanatory variables also exists in the scanning procedure suggested in the California Evaluation Protocols California Public Utilities Commission (2006).

2.1. How scanning induces errors in degree days

At the heart of billing analysis or conditional demand analysis is a simple linear model that attempts to explain kWh with cooling and heating degree days, an intercept and other explanatory variables. This simple linear model takes the following form:

$$kWh_t = \alpha + \beta_1 HDD_t(b) + \beta_2 CDD_t(b) + \epsilon_t \quad (1)$$

where b is the base of the heating or cooling degrees in that month:

$$HDD(b) = \sum max(0, b - AvgTemp). \quad (2)$$

In the traditional scanning technique, the optimal base temperature is found by estimating the regression equation and then finding the base that maximizes some goodness-of-fit statistic such as R^2. Examples of this procedure can be seen in dozens, if not hundreds, of program evaluation reports and journal articles. The procedure of scanning over the base temperatures has even been expanded to allow for different base temperatures for heating and cooling balance points and has been enshrined in the California Energy Efficiency Evaluation Protocols California Public Utilities Commission (2006).

I can find no field outside of energy which uses this econometric procedure, nor does it have more than a passing similarity to other model selection and estimation procedures. The flaw in the standard scanning procedure can be clearly demonstrated with a simple thought experiment.

Suppose an investigator has in mind an empirical model:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2. \quad (3)$$