FORECASTING FOR DIRECT LOAD CONTROL IN ENERGY MARKETS

SHIRA HOROWITZ[a], BRANDON MAUCH[a], AND FALLAW SOWELL[b*]

Abstract. A doubly censored Tobit model is used to forecast hourly air-conditioner usage for individual households. The model is appropriate for a range of temperatures so it is possible to accurately forecast the electricity load to help balance the electricity grid in new settings, eg. solar and wind generation. Individual models are simulated and summed to obtain aggregate forecasts and confidence intervals. The model allows for correlation between the individual shocks that occur in a region. This approach gives substantially more accurate results than the moving average method typically used for forecasting and measuring direct load control.

1. Introduction

Demand response is the reduction of electric loads via price signals or remote access. Electric grid operators and electrical utilities implement demand response to maintain grid reliability or provide electrical service at lower cost. One type of demand response is direct load control (DLC) where electrical appliances are remotely powered off. From a grid perspective, load reduction (decrease in demand) is similar to generation increase (increase in supply). Unlike generation where the supply is deterministic (barring events that lead to a forced outage), the DLC resource is uncertain and must be forecasted. While generators are paid according to the quantity of energy supplied, DLC participants are paid based on the amount of load reduction[1]. Load reductions cannot be directly measured; they are estimated by subtracting actual load during a DLC event

Date: Compiled on February 20, 2013 at 19:48.

* Correspondence to: Fallaw Sowell, Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: fs0v@andrew.cmu.edu.

[a] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA.

[b] Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA.

[1]Load aggregators and large customers usually receive the market prices. However, residential customers usually receive a flat rate for participation from a load aggregator.

from a customer's forecasted load conditional on the DLC event not occurring. In this paper we propose a new method for forecasting and measuring DLC of residential ACs using a Tobit model with upper and lower censoring.

1.1. **Direct Load Control.** Effective DLC is widely used to reduce peak load, which delays the need to build power plants or transmission lines. However, in recent years it is also used as reserve capacity for contingencies in the grid. PJM, a northeastern grid in the US, provides 20% of its contingency reserves with DLC resources (PJM, 2012). DLC can also be used to adjust load as a means of balancing variability of wind and solar resources (Callaway, 2009; Koch et al., 2010; Newell and Felder, 2007). The Department of Energy stated that increased reliance on electricity generation from wind and solar power is one factor that will drive demand response programs (DOE, 2008). Increased use of DLC will require more accurate load forecasting techniques that are easy to implement, like the method we develop in this work. Model accuracy is needed over a range of temperatures since DLC can be called for peak load reductions at high temperatures as well as contingency reserves at lower temperatures.

ACs are well suited for DLC since they can be powered off for short periods of time without much customer discomfort. A California utility surveyed customers during a pilot study and found the majority did not notice DLC events lasting 15 minutes or less (Sullivan et al., 2012). AC's also comprise a large portion of residential loads (roughly 20% of residential electricity consumption) (EIA, 2012b).

Advanced electric meters (i.e. smart meters) allow finer control over electric loads and provide more load data which will enable greater use of DLC in electric grids (Strbac, 2008; Hamilton and Gulhar, 2010). As of 2011, 13.4% of all electricity customers had smart meters (FERC, 2011a), but the Department of Energy is providing funds to quickly increase this level (DOE, 2012). There is a need for models to efficiently take advantage of this new data. This Tobit model captures the realistic situation that AC loads are bounded below by zero and above by the maximum energy consumption for a particular AC.

Recent changes in wholesale electric markets will also increase the use of DLC. In 2011, FERC issued order number 745 which directs wholesale energy market operators to compensate demand side resources (eg. DLC) the full energy market price as long as dispatching the DR resource is cost-effective (FERC, 2011). Each market operator sets a threshold price based on historical data which is used as the minimum price at which DR resources are compensated.

1.2. **Load Forecasting.** Accurate load forecasts are essential for efficient DLC. DR resources are paid the energy market locational marginal price for the reduced load based upon the customer baseline (CBL). This is an estimate for a counterfactual event, i.e. the expected load conditional on the DLC event not happening. Inaccuracies in the CBL lead to incorrect and unfair payments. Underpayments for DR resources discourage participation while overpayments lead to excessive charges on load serving entities who pay for the reductions. System planners need to accurately know how much load reduction to expect during a DLC event. Reducing uncertainty in the load forecasts will become more important as DLC resources provide more ancillary services to help balance the smart electricity grid.

Default CBLs differ across markets (Grimm, 2008; Kema, 2011). Most are simple moving averages. In the PJM RTO, the default CBL is the average hourly load profile from the 4 highest load days of the previous 5 similar day types (weekdays, Saturdays, Sundays/holidays) (PJM, 2012). The California ISO, the New York ISO and the New England's ISO calculate CBLs by averaging loads from the previous 10 similar days (CASIO, 2012; NYISO, 2010; ISONE, 2012). The Electric Reliability Council of Texas publishes 3 different default CBL calculations: a linear regression of energy consumption on covariates representing weather conditions, daylight hours, season and day of the week; a moving average of 8 of the previous 10 similar days; or a model that averages days with load profiles similar to the event day (ERCOT, 2012). All ISO/RTOs accept alternative methods for CBL determination as long as it is approved. This paper presents such an alternative.

Broadly speaking, air conditioner load forecast models can be classified into two distinct categories: engineering models and statistical models. Both model types attempt to forecast AC load as a function of several variables, primarily: temperature and time of day. The most common are engineering models of a house that consist of a system of differential equations that capture the evolution of indoor temperature and the on/off cycles of the air conditioner compressor given weather variables such as temperature, solar radiation, etc. These models requires measuring the thermal characteristics and thermostat settings of each house for use as parameters (Bargiotas and Birdwell, 1988; Molina et al., 2003; Gustafson et al., 1993). Another approach is to use maximum likelihood to estimate these parameters from historical data (Pahwa and Brice, 1985; El-Ferik et al., 2006; Kamoun and Malhamé, 1992). The latter method still requires knowledge of the thermostat setpoints. Unfortunately, these models are sensitive to changes in the physical properties of the residence such as home improvements.

There is comparatively less work on statistical models applied to AC load forecasts, especially residential. Statistical models do not directly model the dynamics of energy flows. Instead they capture trends in historical AC load data to predict future loads.

Parametric models of AC duty cycles have been used to estimate load reductions by comparing controlled and non-controlled AC data (Ryan et al., 1989). Autoregressive models have been used in AC forecasts for non-residential buildings (Penya et al., 2011), but not the highly variable residential data. Machine learning type models have been proposed to forecast building energy consumption using support vector regression (Xuemei et al., 2010) and artificial neural networks (Beccali et al., 2004). These types of models capture the non-linearities in energy demand, but are data intensive for each household. A recent proposal to forecast load reduction from AC DLC relies on fitting a model to load measurements at a feeder circuit level (Eto et al., 2012). This method cannot forecast load for individual households and requires a large fraction of ACs on each feeder participate in DLC so that it can distinguish the signal from the noise. This is a concern for forecasts with lower temperatures.

This paper considers a doubly censored Tobit model to forecast hourly individual AC loads. This accounts for the non-linearities inherent in AC energy consumption while not requiring extreme amounts of data. The model uses ambient temperature and time of day as covariates. The individual forecasts are aggregated via simulation to create day-ahead hourly aggregate load forecast and confidence intervals.

The remainder of this paper is organized as follows: section 2 describes the dataset. Section 3 describes the Tobit model and the theoretical framework of the model. The results are in section 4 and section 5 covers the policy implications.

2. Data

Under a confidentiality agreement, a dataset was obtained from Pepco Holdings, Inc. The dataset contains AC energy consumption data, weather data and individual characteristics for the 536 residential ACs from July - October 2010. Due to data quality issues 69 units were discarded with the analysis performed on the remaining 467 units (details on data quality and cleaning protocol are in appendix A). Data loggers were installed on the ACs during the month of July so the initial date of data collection varies. The data loggers recorded current measurements for the compressor circuits. During installation, technicians took spot measurements of voltage and power which were used to convert the current measurements to power measurements. The raw data were instantaneous power values recorded at three minute intervals. A constant power level is assumed during each three minute period to estimate energy consumption at the hourly time scale, i.e. the compressor was assumed to be on or off for the entire

3 minute period. Adding the three minute periods, the hourly estimates take on 21 discrete values. This discreteness does not have a significant effect on the final results which are aggregated over all units.

The ACs were serviced by one of three utilities: Potomac Electric Power Company (PEPCO), Delmarva Power and Light (Delmarva) and Atlantic City Electric (ACE). Hourly temperature data were collected from weather stations located near each utility's territory and were assumed to be uniform throughout each region; see table 1.

[Table 1 about here.]

There were three missing temperature measurements which were interpolated from the adjacent hours.

All the customers in the data sets had signed up to participate in DLC programs. Some customers had only recently enrolled and their DLC was not active. Data on when individual customer's DLC became active is unavailable. For each AC unit, dummy variables are used to model the impact of DLC events.

In order to participate in the DLC program, customers agreed to have either switches capable of remote operation installed on the AC compressor circuit or smart thermostats that could be adjusted remotely. Customers received notice 24 hours prior to a DLC event. During the time period covered in the data, 8 DLC events occurred ranging in duration from one to four hours. Customers had the option of overriding the signal if they wanted. However, this only occurred with one customer in the dataset. Summary statistics for the dataset are in table 2.

[Table 2 about here.]

3. Framework

3.1. **Tobit Model.** Preference for AC usage is positively related to temperature. At higher temperatures, consumers want more cooling, even if their AC has reached is maximum capacity, while at cooler temperatures, consumers want less AC, and if it is cool enough, they may even want a negative amount of cooling (i.e. heat). Actual AC electricity load is constrained by zero and its maximum capacity. These structural restrictions on observed AC energy consumption are modeled with a doubly censored Tobit model (Tobin, 1958).

A model is needed to capture the AC electricity load over a range of temperatures. Most current residential DLC programs are concerned with peak shaving, i.e. reducing demand only when temperature are extremely high. The need to balance the grid during contingencies or to accommodate alternative energy sources can occur at almost any temperature, and, hence, a model is needed to accurately forecast the AC electricity load (possible savings) when the temperature is say 79 not just at 97.

The desired electricity load for an individaul AC i at time t (incremented hourly) is modeled as a latent variable $y^*_{i,t}$:

$$y^*_{i,t} = \sum_{h=1}^{24} (\beta_{D_h,i} D_{h,t} + \beta_{TD_h,i} D_{h,t} T_t) + \beta_{T^2,i} T_t^2 + \beta_{T1,i} T_{t-1} + \beta_{E,i} E_t + \beta_{P,i} P_t + \epsilon_{i,t} \quad (1)$$

where T_t is $\max(0, R_t - 65)$, R_t is the temperature in degrees Fahrenheit during hour t, $D_{h,t}$ is an indicator variable for hour h, E_t, an indicator variable for a DLC event during hour t, P_t is an indicator for the three hours immediately after an event and $\beta_{\chi,i}$ is a parameter for AC i for covariate χ. The error $\epsilon_{i,t} \sim N(0, \sigma_i^2)$ accounts for an unobservable shock.

The shifted temperature term and temperature squared account for non-linearities of the observed temperature range. The terms in the parentheses are hourly intercepts and linear responses to temperature to account for consumers' diurnal activity cycle. The lagged temperature term is included to account for the home's thermal inertia. A lag of 1 was included because the partial-autocorrelation function was insignificant for longer lags. DLC events in the dataset are accounted for with E_t. A DLC event may increase a customer's preference for AC immediately after the event, P_t is included to account for this. Additional variables were considered such as lagged AC load, eg.

humidity, however these variables were either strongly collinear with other covariates or did not improve the model's fit.

Certain ACs had one or more hours of the day that had very few uncensored $y_{i,t}$ values (eg. all the values at 3 a.m. for a particular AC were 0). This lack of observed variability prevents both identification and convergence of the optimization routine. Any hour with 3 or fewer uncensored values was combined with the previous or next hour (whichever had fewer uncensored values) until each bin contained greater than 3 uncensored values.

Combine the covariates into $\boldsymbol{X_t}$ and all $\beta_{\chi,i}$ into $\boldsymbol{\beta_i}$ to write (1) as

$$y^*_{i,t} = \boldsymbol{X_t}'\boldsymbol{\beta_i} + \epsilon_{i,t}.$$

The observed data censor $y^*_{i,t}$ between 0 and λ_i, the capacity of the AC, to obtain $y_{i,t}$, the actual energy consumption of AC i during hour t:

$$y_{i,t} = \begin{cases} 0 & y^*_{i,t} \leq 0 \\ y^*_{i,t} & 0 < y^*_{i,t} < \lambda_i \\ \lambda_i & \lambda_i \leq y^*_{i,t}. \end{cases}$$

This model is estimated for each AC by maximum likelihood. The likelihood function is in appendix B. Robust standard errors allow for heteroscedasticity and autocorrelation and use Newey-West weights.

3.2. **Forecasting and Confidence Intervals.** We simulate a load aggregator bidding DLC into the forward energy market the day before a DLC event occurs. For example, a load aggregator would place a bid in the forward market on August 14 for a DLC event that is to occur on August 15. The aggregator would have data up to and including August 13 to forecast load for the August 15 event. The forecasts reported in this paper are computed the same way. The August 15 forecast uses parameter estimates constructed with data up to and including August 13. The August 16 forecast uses data up to and including August 14. Each AC, i therefore has different parameters $\hat{\boldsymbol{\beta}}_{\boldsymbol{i,d}}$ for each day d. The estimates, $\hat{\boldsymbol{\beta}}_{\boldsymbol{i,d}}$ and $\hat{\boldsymbol{\sigma}}_{\boldsymbol{i,d}}$ use all available $X_{i,t}$ where $t < d - 2$. To ensure at least two weeks of data, the forecasts were made from August 15 to October 1.

The starting values used for the nonlinear optimization routine for August 15 were the ordinary least squares estimates using data up to August 14. The starting values for each successive day $d > 1$ are the ML estimates from the previous day: $\hat{\boldsymbol{\beta}}_{\boldsymbol{i,d-1}}$ and $\hat{\boldsymbol{\sigma}}_{\boldsymbol{i,d-1}}$. Results were not sensitive to changes in the starting values.

The forecasted latent variable for AC i at time t on day d is:

$$\hat{y}^*_{i,t} = \boldsymbol{X_t}'\hat{\boldsymbol{\beta}}_{\boldsymbol{i,d}} \; \forall \; t \in d.$$

Simulation is used to compute aggregate forecasts and confidence intervals. A simulated random error is added to the latent estimate and then censored to obtain M observed load estimates $\upsilon_{i,t,m}$ for each AC i at each time $t \in d$, $\upsilon^*_{i,t,m} = \hat{y}^*_{i,t} + e_{i,t,m}$

$$\upsilon_{i,t,m} = \begin{cases} 0 & \upsilon^*_{i,t,m} \leq 0 \\ \upsilon^*_{i,t,m} & 0 < \upsilon^*_{i,t,m} < \lambda_i \\ \lambda_i & \lambda_i \leq \upsilon^*_{i,t,m}. \end{cases}$$

Sum across ACs to obtain M aggregated loads at each time, t

$$Y_{t,m} = \sum_{i=1}^{N} \upsilon_{i,t,m} \text{ for } m = 1, \ldots, M. \tag{2}$$

ACs in the same electric utility are in the same geographic region. This raises the concern that their shock can be correlated. If ignored, the confidence intervals for the aggregate load will be too optimistic. The Gaussian errors are allowed to have correlation ρ across ACs for each t in each utility. The m^{th} simulation draws a random error, $e_{i,t,m}$ from $N(0, \boldsymbol{\Sigma_d})$. For N ACs in a single utility the covariance matrix is

$$\boldsymbol{\Sigma_d} = \begin{bmatrix} \boldsymbol{\hat{\sigma}^2_{1,d}} & \boldsymbol{\rho\hat{\sigma}_{1,d}\hat{\sigma}_{2,d}} & \cdots & \boldsymbol{\rho\hat{\sigma}_{1,d}\hat{\sigma}_{N,d}} \\ \boldsymbol{\rho\hat{\sigma}_{2,d}\hat{\sigma}_{1,d}} & \boldsymbol{\hat{\sigma}^2_{2,d}} & \cdots & \boldsymbol{\rho\hat{\sigma}_{2,d}\hat{\sigma}_{N,d}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\rho\hat{\sigma}_{N,d}\hat{\sigma}_{1,d}} & \boldsymbol{\rho\hat{\sigma}_{N,d}\hat{\sigma}_{2,d}} & \cdots & \boldsymbol{\hat{\sigma}^2_{N,d}} \end{bmatrix}.$$

The correlation ρ is selected to have the highest agreement between the observed confidence intervals and the theoretical confidence intervals for the aggregate load. This can be thought of as a GMM estimation problem. See Appendix C for details.

4. Results

The Tobit model described in section 3 is estimated for each AC in the sample. Figures 1, 2 and 3 shows the median and upper and lower quartiles of the t-statistics for the $\hat{\beta}_{\chi,i}$ for all individual models. The estimated models are used to forecast AC load each hour from August 15, 2010 to October 1, 2010. The following steps produce hourly aggregate AC load forecasts for each day during the simulation period.

(1) Estimate the Tobit models for data collected up to day $d-2$.
(2) Calculate an hourly forecast for each AC for day d.
(3) Simulate individual forecasts and aggregate to get expected load and confidence intervals.
(4) Repeat 1 - 3 for $d+1, d+2$....

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

Aggregate forecasts were produced using the simulation method described in section 3. For each electric utility a correlation coefficient ρ was estimated. Table 3 shows the correlation coefficients for each electric utility.

[Table 3 about here.]

[Figure 4 about here.]

The Tobit based forecasts are compared to the default CBL used in the PJM RTO since PEPCO is in PJM territory. The default CBL in PJM is the average hourly load profile from the 4 highest load days of the previous 5 similar day types (weekdays, Saturdays, Sundays/holidays) (PJM, 2012). Table 4 shows the mean squared error (MSE) for the default CBL and Tobit models. The Tobit model has a MSE that is an order of magnitude lower than the default CBL. Representative examples of the Tobit based forecasts, currently used default CBL and actual values are given in Figures 4 and 5. Figure 4 shows the actual load, Tobit based forecast and 50% confidence interval for the PEPCO utility. Figure 5 compares the default CBL forecast to the Tobit based forecast.

[Figure 5 about here.]

Since the Tobit model presented here uses hourly temperatures, it is important to see how well it performs over a range of temperatures. Figure 6 shows the forecast errors plotted against the ambient temperature. At high temperatures there is a tendency to over-forecast. One possible explanation is that vacations occur more frequently at the end of summer. This is the time period where the forecasts were made, based on data collected earlier in the summer. A full summer of training data may improve the forecasts by allowing monthly indicators in the model.

[Figure 6 about here.]

[Table 4 about here.]

5. Policy Implications and Discussion

Demand response is increasing in the US as a way to make the electric grid more reliable and provide services at lower cost. Forecasting, measurement and verification of direct load control are becoming increasingly important, as penetration levels of demand response increase. Forecasting is important for system planning and measurement and verification are necessary to ensure that payments are fair. Forecasting, measurement and verification are difficult because of the need to measure the quantity of power that was *not* used, i.e. a counterfactual.

This paper introduces censored regression as an improvement on current methods to forecast available direct load control resource. The aggregate forecast accounts for correlation between individual shocks. This forecast can be used to determine AC load in the counter factual where DLC is not applied. This method is more accurate than the moving averages used by most ISO's, and is simple, easy and cheap to implement. This method can be further refined and extended in future work.

6. Acknowledgments

The authors thank Jay Apt for valuable comments and discussion. Pepco Holdings, Inc. generously provided the data.

This work was supported in part by grants from the Alfred P. Sloan Foundation and EPRI to the Carnegie Mellon Electricity Industry Center; the Doris Duke Charitable Foundation, the Department of Energy National Energy Technology Laboratory, and the Heinz Endowments to the RenewElec program at Carnegie Mellon University; the U.S. National Science Foundation under Awards SES-0949710 and DGE-0750271; Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia), number SFRH/BD/33764/2009; and the Department of Energy under grants EE-OE0000300 and DE-FOA-0000058.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe

privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Appendix A. Data Cleaning Protocol

Six ACs had too few observations (fewer than 9 days; the remaining ACs all had more than 52 days) and were removed. The data from 2 ACs were collected at the wrong frequency (not three minute) and were removed. Some of the ACs showed they were rarely used (less than 3 hours during the entire summer). Data from these ACs were not analyzed since they provided no information for forecasts. Several ACs had 20A loggers even though the AC capacity was greater than 20A. If an AC logged data at 20A more than 10% of the time, we assume that it required a higher amperage logger, and discard the data. There were also several ACs that became stuck on a particular value. An AC was removed if it switched state (from off to on or vice versa) in fewer than 2% of its observations. Finally, we removed ACs that had unrealistically low readings (all observations below 3 amps). A summary of the AC data discarded is in table 5.

[Table 5 about here.]

Appendix B. Tobit Model Derivations

This is a derivation of the Tobit model. Lower and upper bounds for censoring are represented here as a and b. For this paper, $a = 0$ and $b = \lambda$. To simplify the notation, drop i, the index for AC from the derivations. The latent variable is:

$$y_t^* = \boldsymbol{X_t'}\boldsymbol{\beta} + \epsilon_t.$$

The censored variable is:

$$y_t = \begin{cases} a & y_t^* \leq a \\ y_t^* & a < y_t^* < b \\ b & b \leq y_t^*. \end{cases}$$

Define the indicator variables:

$$I_t(a) = \begin{cases} 1 & y_t^* \leq a \\ 0 & a < y_t^*, \end{cases} \qquad I_t(ab) = \begin{cases} 1 & a < y_t^* < b \\ 0 & \text{otherwise} \end{cases} \qquad \text{and } I_t(b) = \begin{cases} 1 & b \leq y_t^* \\ 0 & y_t^* < b. \end{cases}$$

Assume the latent variable, y_t^* has distribution $N(\mu, \sigma^2)$. The entire probability density of the lower censored region is applied at a, and the same for the upper censored region at b. The probability density function for the censored variable is:

$$f(y_t) \quad = \begin{cases} \Phi\left(\frac{a-\mu}{\sigma}\right) & y_t^* \leq a \\ \frac{1}{\sigma}\phi\left(\frac{y_t-\mu}{\sigma}\right) & a < y_t^* < b \\ 1 - \Phi\left(\frac{b-\mu}{\sigma}\right) & b \leq y_t^*. \end{cases}$$

The log-likelihood function $\ell(\boldsymbol{Y_\tau}, \boldsymbol{X_\tau}, \theta)$ can be expressed in terms of the vector of parameters $\boldsymbol{\theta} = [\boldsymbol{\beta}' \; \sigma]'$ and τ, the length of the time-series

$$\begin{aligned} \ell(\boldsymbol{Y_\tau}, \boldsymbol{X_\tau}, \theta)) &= \sum_{t=1}^{\tau} \ln(f(y_t)) \\ &= \sum_{t=1}^{\tau} \left[I_t(a) \ln\left(\Phi\left(\frac{a-\mu}{\sigma}\right)\right) + I_t(ab) \ln\left(\frac{1}{\sigma}\phi\left(\frac{y_t-\mu}{\sigma}\right)\right) + I_t(b) \ln\left(\Phi\left(\frac{\mu-b}{\sigma}\right)\right)\right] \end{aligned}$$

Let $\mu = \boldsymbol{X_t'\beta}$ to obtain

$$\begin{aligned} \ell(\boldsymbol{Y_\tau}, \boldsymbol{X_\tau}, \theta) \;\; = \;\; \sum_{t=1}^{\tau} \Bigg[& I_t(a) \ln\left(\Phi\left(\frac{a - \boldsymbol{X_t'\beta}}{\sigma}\right)\right) - I_t(ab)\frac{1}{2}\ln\left(2\pi\sigma^2\right) \\ & - I_t(ab)\frac{(y_t - \boldsymbol{X_t'\beta})^2}{2\sigma^2} + I_t(b) \ln\left(\Phi\left(\frac{\boldsymbol{X_t'\beta} - b}{\sigma}\right)\right)\Bigg]. \end{aligned}$$

The gradient is $\nabla\ell(\boldsymbol{Y_\tau}, \boldsymbol{X_\tau}, \theta) = \begin{bmatrix} \frac{\partial \ell}{\partial \beta'} & \frac{\partial \ell}{\partial \sigma'} \end{bmatrix}'$ where

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= \sum_{t=1}^{\tau} \left[I_t(a) \frac{\phi\left(\frac{a-\boldsymbol{X_t'\beta}}{\sigma}\right)}{\Phi\left(\frac{a-\boldsymbol{X_t'\beta}}{\sigma}\right)} \left(-\frac{\boldsymbol{X_t}}{\sigma}\right) + I_t(ab)\frac{(y_t - \boldsymbol{X_t'\beta})\boldsymbol{X_t}}{\sigma^2} + I_t(b)\frac{\phi\left(\frac{\boldsymbol{X_t'\beta}-b}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{X_t'\beta}-b}{\sigma}\right)}\left(\frac{\boldsymbol{X_t}}{\sigma}\right)\right] \\
\frac{\partial \ell}{\partial \sigma} &= \sum_{t=1}^{\tau} \left[I_t(a) \frac{\phi\left(\frac{a-\boldsymbol{X_t'\beta}}{\sigma}\right)}{\Phi\left(\frac{a-\boldsymbol{X_t'\beta}}{\sigma}\right)} \left(\frac{\boldsymbol{X_t'\beta} - a}{\sigma^2}\right) - I_t(ab)\frac{1}{\sigma} \right. \\
&\quad \left. + I_t(ab)\frac{(y_t - \boldsymbol{X_t'\beta})^2}{\sigma^3} + I_t(b)\frac{\phi\left(\frac{\boldsymbol{X_t'\beta}-b}{\sigma}\right)}{\Phi\left(\frac{\boldsymbol{X_t'\beta}-b}{\sigma}\right)}\left(\frac{b - \boldsymbol{X_t'\beta}}{\sigma^2}\right)\right].
\end{aligned}$$

The gradient contains the indeterminate form $\lim_{z\to-\infty} \frac{\phi(z)}{\Phi(z)} = \frac{0}{0}$. To evaluate the gradient we apply L'Hôpital's rule for $z < -38$ use $\frac{\phi(z)}{\Phi(z)} \approx -z$.

The HAC variance is based on Bernard and Busse (2003). The auto-covariance is

$$\gamma(\delta) = \frac{1}{\tau - \delta} \sum_{t=1}^{\tau-\delta} \nabla\ell(\boldsymbol{X_t}, \theta)\nabla\ell(\boldsymbol{X_{t+\delta}}, \theta)'.$$

The Newey-West weights are expressed as $\omega(\delta) = 1 - \frac{\delta}{\Delta+1}$ where $\Delta \leq \sqrt{\tau}$, we use $\Delta = \tau^{0.4}$. The variance of likelihood estimate with the HAC correction is expressed in terms of the auto-covariance and Newey-West weights:

$$Var\nabla\ell = \omega(0)\hat{\gamma}(0) + \sum_{\delta=1}^{\Delta} \omega(\delta)\left(\hat{\gamma}(\delta) + \hat{\gamma}(-\delta)'\right).$$

The variance with HAC correction also uses the Hessian $H = \nabla^2\ell$

$$Var(\theta) = (-H)^{-1} Var\nabla\ell(-H/\tau)^{-1}.$$

Appendix C. Confidence Interval for ρ

The confidence interval for ρ is obtained by inverting the GMM objective function, also called the DM test in Newey and McFadden (1986).

The parameter estimates for the individual Tobit models are obtained by maximum likelihood. For a single utility collect all the estimates in $\hat{\psi}$, i.e. the ML $\hat{\theta}$ estimates from each AC. The the gradient for each AC, $\nabla \ell(\boldsymbol{Y}_{\boldsymbol{\tau}}, \boldsymbol{X}_{\boldsymbol{\tau}}, \theta)$, set to zero are the moment conditions for these estimates. Together these are the scores and will be denoted $S(\psi)$. The moment conditions are satisfied at the ML estimates. The estimated models and the normality assumption on the errors (for the Tobit model) implies the forecast distributions for each AC. These can be used to obtain the forecast distribution for the utility's aggregate load for AC generation. These AC's are in the same utility and will experience correlated shocks. The common correlation of ρ is assumed between every AC in the utility.

Instead of calculating the distribution of the aggregate load forecast analytically, the correlated errors will be simulated. The implied individual forecasts will be summed to simulate an aggregate forecast. The order statistics from the simulated aggregate forecasts are a consistent estimates for the critical values for the forecast distribution. For the population parameter value ρ_0 the observed aggregate load for the utility, $Y_{d,h}$, will be in the simulated confidence interval with the correct probability. The 95%, 90% and 50% confidence intervals give moment conditions for the estimation of ρ.

$$G(\psi, \rho) = \frac{1}{\tau} \sum_{d,h} \begin{bmatrix} \chi\left(\hat{Y}(\psi, \rho)_{d,h,[.025M]} \leq Y_{d,h} \leq \hat{Y}(\psi, \rho)_{d,h,[.975M]}\right) & - & .95 \\ \chi\left(\hat{Y}(\psi, \rho)_{d,h,[.05M]} \leq Y_{d,h} \leq \hat{Y}(\psi, \rho)_{d,h,[.95M]}\right) & - & .9 \\ \chi\left(\hat{Y}(\psi, \rho)_{d,h,[.25M]} \leq Y_{d,h} \leq \hat{Y}(\psi, \rho)_{d,h,[.75M]}\right) & - & .5 \end{bmatrix}.$$

where $\hat{Y}(\psi, \rho)_{d,h,k}$ is the k^{th} order statistics of the simulated aggregate forecasts and M is the number of simulations. This paper used $M = 1000$. The results did not change for larger M.

The GMM objective function can be written

$$Q(\psi, \rho) = \begin{bmatrix} S(\psi) & G(\psi, \rho) \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & W \end{bmatrix} \begin{bmatrix} S(\psi) \\ G(\psi, \rho) \end{bmatrix}.$$

Because the Tobit parameters ψ are just identified with the scores from the MLE, the weighting matrix only influences the last three moments that estimate ρ.

Optimal two-step GMM is used with $W = I$ in the first step. Because ρ does not enter the scores, the ψ estimates do not change as ρ takes different values. The confidence intervals for ρ are obtained by

$$\tau(Q(\hat{\psi}, \hat{\rho}) - Q(\hat{\psi}, \rho_0)) \sim \chi^2_{(1)}$$

where $\hat{\psi}$ are the MLE Tobit estimates and $\hat{\rho}$ is the efficient second round GMM estimate. For this data set $\tau = (48 \text{ days}) \times (24 \text{ hours}) = 1152$.

References

Bargiotas D, Birdwell J. 1988. Residential air conditioner dynamic model for direct load control. *Power Delivery, IEEE Transactions on* **3**: 2119.

Beccali M, Cellura M, Lo Brano V, Marvuglia A. 2004. Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management* **45**: 2879–2900.

Bernard AB, Busse MR. 2003. Consistent standard errors in panel Tobit with autocorrelation. Working paper 03-25, Tuck Business School.
URL `http://ssrn.com/abstract=439061orhttp://dx.doi.org/10.2139/ssrn.439061`

CASIO. 2012. Business practice manual for metering. Manual Version 6, California Independent System Operator.
URL `https://bpm.caiso.com/bpm/bpm/version/000000000000168`

El-Ferik S, Hussain SA, Al-Sunni FM. 2006. Identification and weather sensitivity of physically based model of residential air-conditioners for direct load control: A case study. *Energy and Buildings* **38**: 997 – 1005. ISSN 0378-7788.
URL `http://www.sciencedirect.com/science/article/pii/S037877880500232X`

ERCOT. 2012. Emergency interruptible load service default baseline methodologies. Technical report, Electric Reliability Council of Texas.

Eto J, Nelson-Hoffman J, Parker E, Bernier C, Young P, Sheehan D, Kueck J, Kirby B. 2012. The demand response spinning reserve demonstration–measuring the speed and magnitude of aggregated demand response. In *System Science (HICSS), 2012 45th Hawaii International Conference on*. ISSN 1530-1605, 2012 –2019.

FERC. 2011. Demand response compensation in organized wholesale energy markets. Docket No. RM10-17-000 Order No. 745, Federal Energy Regulatory Commission.
URL `http://www.ferc.gov/whats-new/comm-meet/2011/121511/E-4.pdf`

Grimm C. 2008. Evaluating baselines for demand response programs. In *2008 AEIC Load Research Workshop*. San Antonio, TX.

Gustafson M, Baylor J, Epstein G. 1993. Estimating air conditioning load control effectiveness using an engineering model. *Power Systems, IEEE Transactions on* **8**: 972–978. ISSN 0885-8950.

ISONE. 2012. Market rule 1 section III.8. Technical report, Independent System Operator of New England.
URL http://www.iso-ne.com/regulatory/tariff/sect_3/index.html

Kamoun S, Malhamé R. 1992. Convergence characteristics of a maximum likelihood load model identification scheme. *Automatica* **28**: 885 – 896. ISSN 0005-1098.
URL http://www.sciencedirect.com/science/article/pii/0005109892901423

Kema. 2011. PJM empirical analysis of demand response baseline methods. Technical report, Kema.
URL http://www.pjm.com/~/media/committees-groups/subcommittees/drs/20110613/20110613-item-03b-cbl-analysis-report.ashx

Molina A, Gabaldon A, Fuentes J, Alvarez C. 2003. Implementation and assessment of physically based electrical load models: Application to direct load control residential programmes. *Generation, Transmission and Distribution, IEE Proceedings-* **150**: 61 – 66. ISSN 1350-2360.

Newey WK, McFadden D. 1986. Chapter 36: Large sample estimation and hypothesis testing. In Engle RF, McFadden D (eds.) *Handbook of Econometrics*, volume 4. Elsevier, 2111–2245.

NYISO. 2010. Emergency demand response program manual. , New York Independent System Operator.
URL http://www.nyiso.com/public/webdocs/products/demand_response/emergency_demand_response/edrp_mnl.pdf

Pahwa A, Brice C. 1985. Modeling and system identification of residential air conditioning load. *Power Apparatus and Systems, IEEE Transactions on* **104**: 1418 –1425. ISSN 0018-9510.

Penya Y, Borges C, Agote D, Fernandez I. 2011. Short-term load forecasting in air-conditioned non-residential buildings. In *Industrial Electronics (ISIE), 2011 IEEE International Symposium on*. ISSN Pending, 1359 –1364.

PJM. 2012. 2011 DSR annual report. Technical report, PJM.

Ryan N, Powers J, Braithwait S, Smith B. 1989. Generalizing direct load control program analysis: implementation of the duty cycle approach. *Power Systems, IEEE Transactions on* **4**: 293 –299. ISSN 0885-8950.

Tobin J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* **26**: pp. 24–36. ISSN 00129682.
URL http://www.jstor.org/stable/1907382

Xuemei L, Lixing D, Yan L, Gang X, Jibin L. 2010. Hybrid genetic algorithm and support vector regression in cooling load prediction. In *Knowledge Discovery and Data Mining, 2010. WKDD '10. Third International Conference on*. 527 –531.

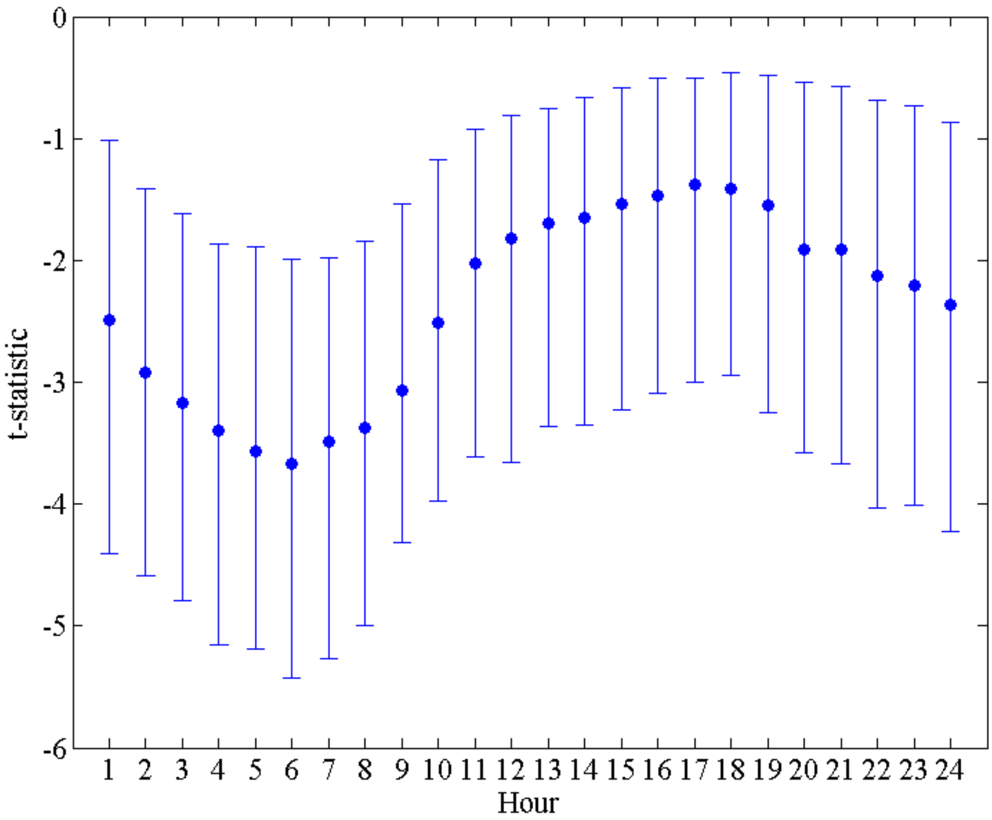

FIGURE 1. Median, upper-quartile, lower-quartile of the t-statistics for $\hat{\beta}_{D_h,i}$.

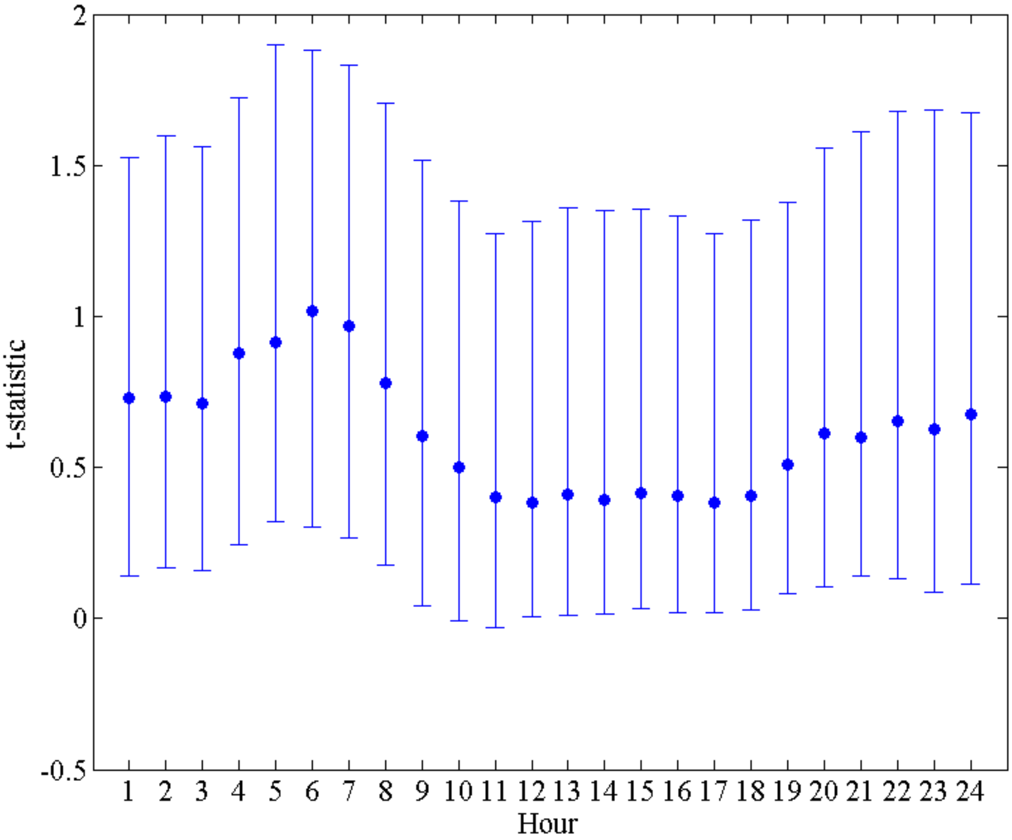

FIGURE 2. Median, upper-quartile, lower-quartile of the t-statistics for $\hat{\beta}_{TD_h,i}$.

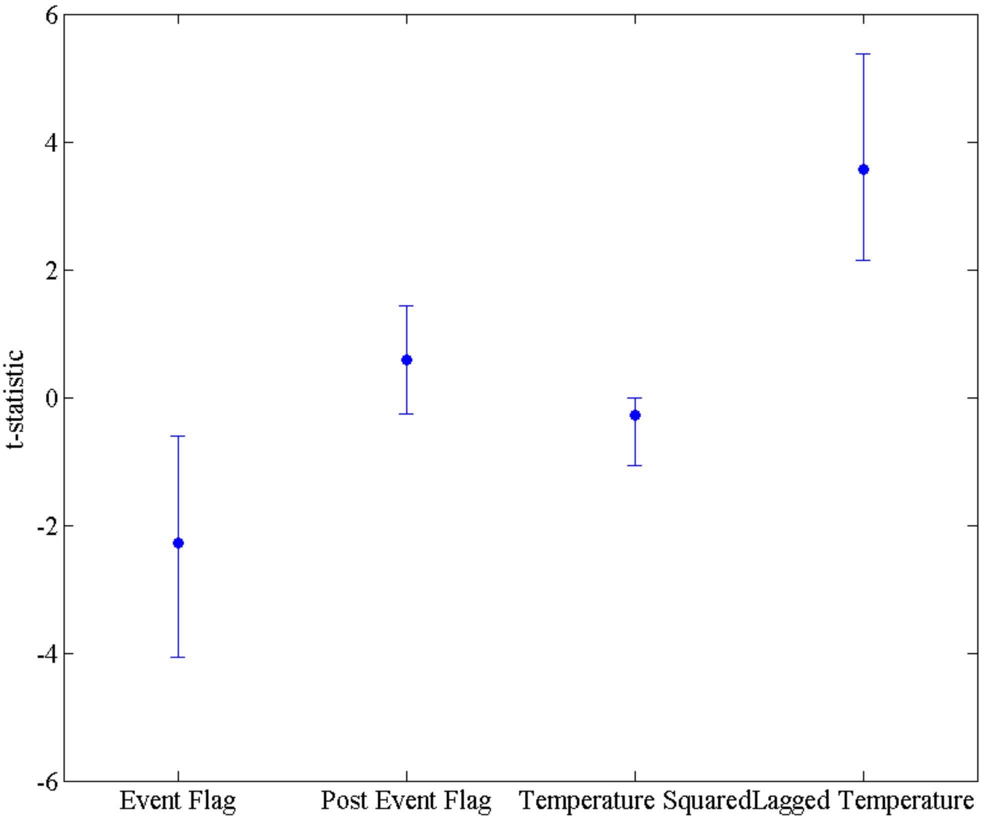

Figure 3. Median, upper-quartile, lower-quartile of the t-statistics for $\hat{\beta}_{E,i}$, $\hat{\beta}_{P,i}$, $\hat{\beta}_{T^2,i}$, $\hat{\beta}_{T1,i}$.

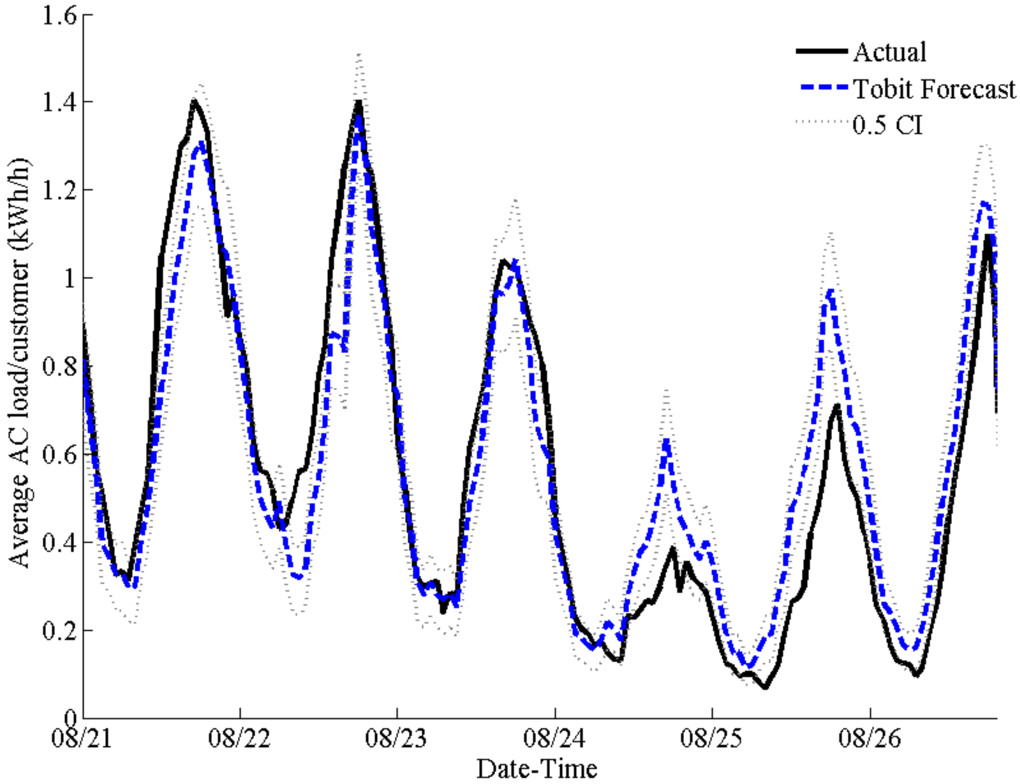

FIGURE 4. Average actual and forecasted AC usage for PEPCO with 50% confidence intervals.

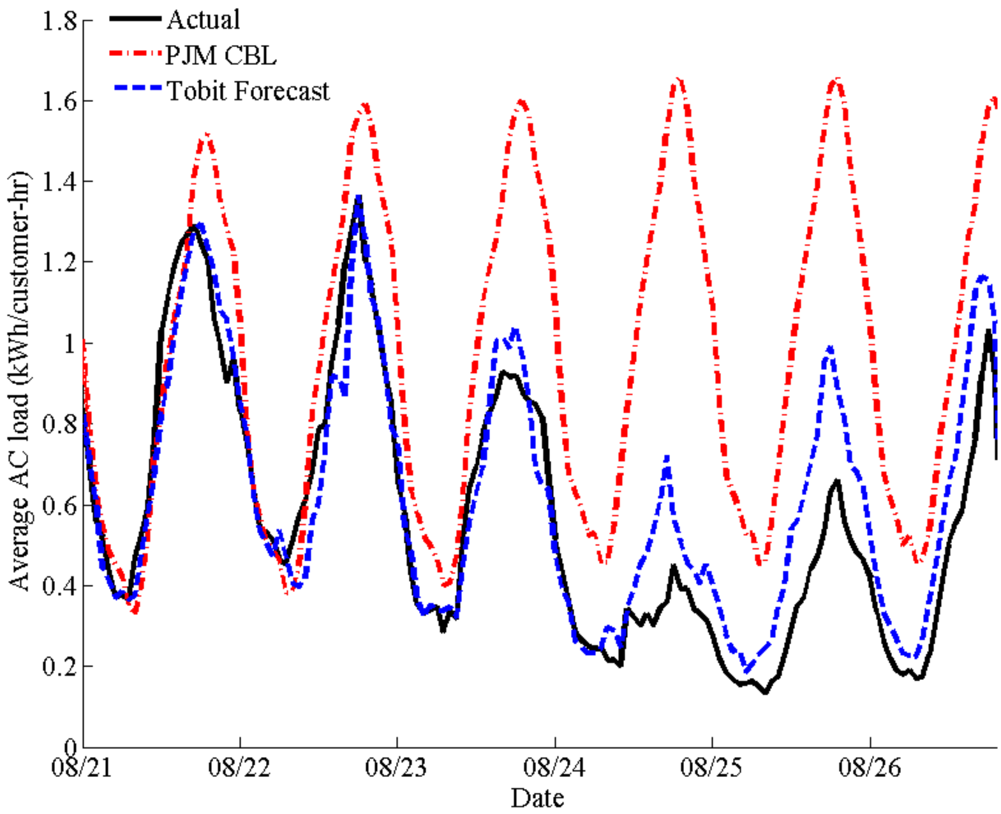

FIGURE 5. Comparison of Tobit forecast and default PJM CBL forecast for PEPCO data.

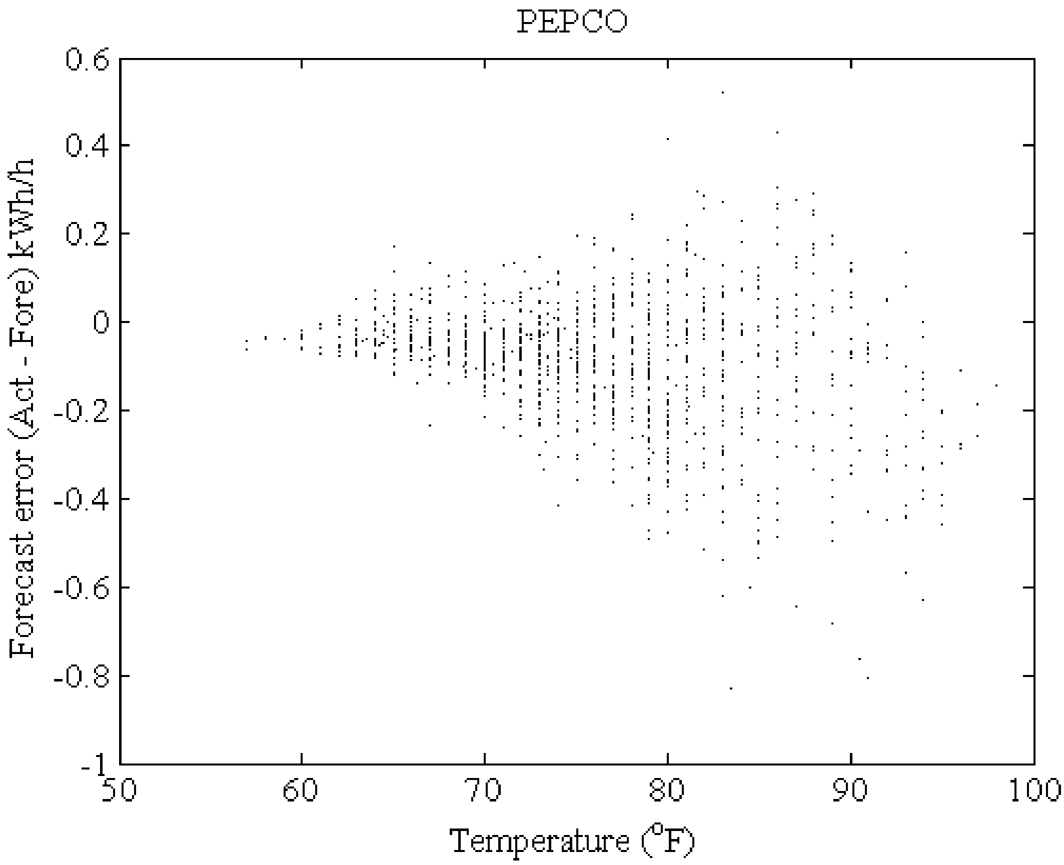

FIGURE 6. Forecast errors plotted against the ambient temperature from August 15, 2010 to October 1, 2010 in PEPCO.

Table 1. Temperature statistics during the period July - September 2010 from each region where the air conditioners are located.

	Utility		
	PEPCO	ACE	Delmarva
Minimum temperature (°F)	57	49	54
Maximum temperature	98	99	96
Mean temperature	76	73	74
Standard deviation	7	9	8

Table 2. Summary statistics from AC data set.

	Utility			
Variables	PEPCO	ACE	Delmarva	Total
Number of total air conditioners	181	72	214	467
Air conditioners cycling at 50%	58	72	88	218
Air conditioners cycling at 75%	68	0	68	136
Air conditioners cycling at 100%	55	0	58	113
Air conditioner size < 2 kW	50	10	49	109
Air conditioner size ≥ 2 and < 3 kW	83	42	110	235
Air conditioner size ≥ 3 and < 4 kW	36	17	50	103
Air conditioner size ≥ 4 kW	12	3	5	20
Average air conditioner size	2.6	2.7	2.5	2.5

TABLE 3. Estimated correlation coefficients for each electric utility.

	Utility		
	PEPCO	ACE	Delmarva
Correlation Coefficient ρ	.19	.45	.26
95% Confidence Interval for ρ	(.10, .34)	(.25, .75)	(.15, .47)

TABLE 4. Mean squared errors.

	Tobit	default CBL
PEPCO	0.034	0.260
ACE	0.041	0.347
Delmarva	0.027	0.302

TABLE 5. Number of ACs discarded from the dataset.

Data Problem	Number of ACs
Length of time less than 9 days	6
Incorrect time intervals	2
AC nearly always off	10
AC maxed out	11
Stuck values	13
Values unrealistically low	27
Total discarded	69