Energy 46 (2012) 533–540

Contents lists available at SciVerse ScienceDirect

Energy

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

Modeling short-run electricity demand with long-term growth rates and consumer price elasticity in commercial and industrial sectors

Amy Pielow [a], Ramteen Sioshansi [a,*], Matthew C. Roberts [b]

[a] Department of Integrated Systems Engineering, The Ohio State University, 1971 Neil Avenue, Columbus, OH 43210, USA
[b] Department of Agricultural, Environmental, and Development Economics, The Ohio State University, 2120 Fyffe Road, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 11 May 2012
Received in revised form
24 July 2012
Accepted 30 July 2012
Available online 30 August 2012

Keywords:
Electricity demand
Commercial
Industrial
Long-term growth
Elasticity

ABSTRACT

This paper specifies and estimates state-level models of short- and long-term electricity demand in the United States. The short-term model predicts hourly load based on weather and calendar inputs. The long-term model estimates interannual demand, and includes population, prices, and gross state product as predictors. These models are combined to incorporate the short- and long-term trends in electricity consumption when generating forecasts of diurnal patterns into the future. Finally, the authors investigate the effects of short-run price elasticities of demand. The short-term model is shown to be within 95% accuracy of actual levels in out-of-sample tests.

1. Introduction

The demand for electricity fluctuates on familiar cycles and with known influences in the short-term, and its long-term growth coincides with trends in macroeconomic indicators. Accurately forecasting the level of demand on disparate time scales is necessary for utilities to schedule generators, plan system maintenance, and devise long-term investments. Short-term forecasts are commonly made in half- or 1-h intervals 24–168 h in advance of the pertinent period. Seasonal patterns such as day of week and month of year, as well as temperature and humidity, are the most significant factors influencing demand within a year. Both types of factors work in conjunction with each other—although on a day-to-day basis the specific day of the week is of great importance, temperature also matters. Monthly seasonality primarily reflects meteorological conditions. These variables become less significant with longer time horizons; models of long-run electricity demand typically forecast aggregate monthly or annual levels. Changes in long-run demand are normally correlated with changes in economic indicators such as gross domestic product and prices of electricity and other fuels.

In the short-term, regression techniques are common for modeling the quantity of electricity demanded. Pardo et al. [1] use autoregressive least-squares regression to explore the effects of temperature and seasonality on daily load. For modeling the daily peak and monthly aggregate demand levels, Mirasgedis et al. [2] also use regressions which include seasonal and temperature variables. Both of these papers are primarily concerned with the role of weather in demand quantity fluctuations. The peak load, average load, temperature, and calendar particulars of the previous day are the basis for the bivariate model of next-day hourly peak in the work of Engle et al. [3]. Forecasts for diurnal load profiles can be made in a comparable manner; Ramanathan et al. [4] build 24 separate regression models, one for each hour of the day with unique regressors. A similar approach is used by Taylor and Buizza [5] to forecast load at various cardinal points of the day, including midday and midnight.

Similar methods are applied to long-term forecasts as well. For example, Mohamed and Bodger [6], Amarawickrama and Hunt [7] and Bianco et al. [8] use annual demand regression models that consider macroeconomic factors, such as gross domestic product and population. Efforts to combine monthly and annual aggregate forecasts from the same data set through cointegration of time series can be found in the work of Engle et al. [9]. The authors capture short-term effects in a monthly model, and then improve it by introducing a factor from a separate annual model influenced more by long-term trends. Artificial neural networks are another

* Corresponding author.
E-mail addresses: pielow.1@osu.edu (A. Pielow), sioshansi.1@osu.edu (R. Sioshansi), roberts.628@osu.edu (M.C. Roberts).

0360-5442/$ – see front matter
http://dx.doi.org/10.1016/j.energy.2012.07.059

method common in the literature for both time scales. Ringwood et al. [10] investigate the influences on demand in multiple time frames by modeling daily, weekly, and yearly quantity demanded using neural networks. More recently, Taylor et al. [11] develop an intra-hour neural network and compare it to double exponential smoothing, regression with principal component analysis, and seasonal ARMA models for two data sets of diurnal load profiles on hourly and half-hourly intervals.

A natural extension of predicting the quantity of electricity demand is investigating how this quantity changes under different pricing policies. Price elasticities are a measure of how much the quantity demanded for a good changes given a change in price. These elasticities can be estimated for changes in the price of the good itself (own) or a change in the price of a different good (cross). With respect to electricity load, the different goods can be thought of as electricity in different hours of the day. An increase (decrease) in price in an hour may decrease (increase) demand in that hour, and change the quantity demanded in adjacent hours as well. Prices can be structured in several ways, for instance a constant time-invariant price or a time-of-use tariff, with different prices applied to consumption during different blocks of time. The hourly own- and cross-price elasticities of demand used in this work are determined by Taylor et al. [12]. These elasticities are estimated from eight years of data for industrial customers across all hours of the day.

Despite the existing body of literature on short- and long-term forecasting, little work which utilizes multiple time horizons is available. In the short-run, models frequently only predict peak or aggregate daily load. We investigate how similar modeling techniques to those found in the literature can be used to predict continuous demand with one function. Furthermore, by predicting the entire diurnal load profile with a single regression model, exploring the interactions in quantity demanded between different hours is possible. Thus, effects of different pricing strategies can be captured by utilizing customers' price elasticities. We explore cases where there is a simple peak and off-peak structure and where the price increases with the consumption level. In addition, as a population and economy grow, the amount of electricity demanded in aggregate can grow as well. We develop a second model of annual aggregate electricity demand regressed against macroeconomic variables. Using the long-term growth rates, as determined by this annual aggregate model, fine-grain predictions further into the future are possible. Although predictions of peak and aggregate load as found in the literature are necessary for planning, forecasting temporally disaggregated load allows further refinement.

The remainder of this paper is organized as follows. Section 2 details the development of the two models; first, the diurnal model form and the data used to develop it are considered and then a similar treatment for the annual regression is undertaken. How these models are combined to forecast loads in the future is also discussed. In Section 3, the estimated regression coefficients are reported along with results from out-of-sample validation of the proposed short-term model and forecasted results and the application of cross- and own-price elasticities. Section 4 concludes and summarizes the work.

2. Data and methodology

2.1. Diurnal model

Our initial regression model for diurnal load profiles is of a log-linear form incorporating calendar and weather variables. Electricity demand has daily, weekly, and monthly cycles which can be described by Fourier series over the respective period. A Fourier series is a linear combination of sine and cosine functions of differing frequencies used to approximate a given function arbitrarily well. This method is advantageous over the classic approach of dummy variables to represent the particular hour, day, and month of an observation. Fewer variables are necessary to represent each time frame; a model with two frequencies at each time scale is optimal with our data, for a total of 16 predictors instead of 40. The three cycles modeled with this method are hour of the day patterns, hour of the week patterns, and month of the year patterns. The longest of these cycles is evident in Fig. 1, which shows hourly load in the state of Ohio by a subset of American Electric Power's (AEP's) commercial and industrial customers in the year 2010. Demand peaks during the summer months, but otherwise is relatively steady on a seasonal basis. This may be in part because cooling technologies used in the summer tend to be electric, while heating technologies used in winter months are not. Fig. 2 displays the first two weeks of hourly demand levels from the same consumers as in Fig. 1. The weekly and daily cycles are apparent. Also evident is the difference in demand from weekdays to weekends, as the first 48 h are over a holiday weekend, and the next 120

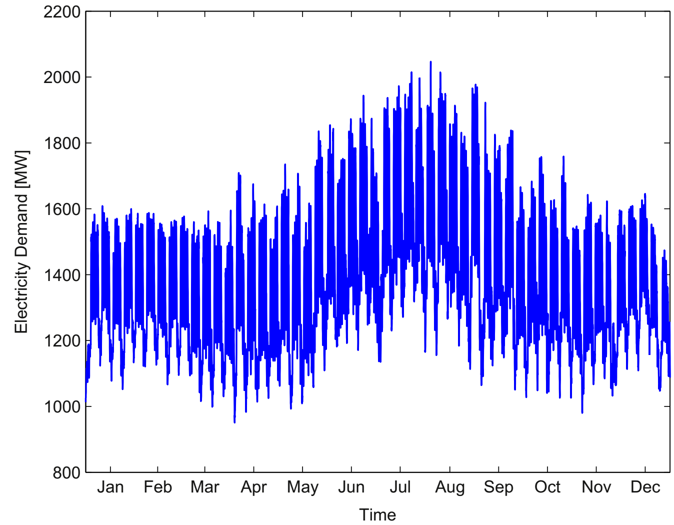

Fig. 1. Hourly electricity demand of 78,000 commercial and industrial customers in the state of Ohio in the year 2010. Data obtained from AEP.

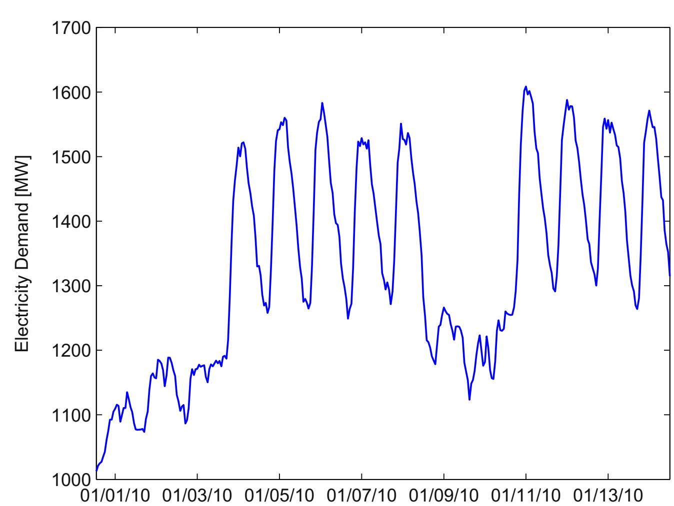

Fig. 2. Diurnal demand patterns for same subset of commercial and industrial customers shown in Fig. 1 in the first two weeks of the year. Data obtained from AEP.

represent Monday through Friday. The quantity demanded in the first 48 h is particularly low, even for Saturdays and Sundays, because of the holiday, and motivates the inclusion of variables to capture this effect. The data shown in Figs. 1 and 2 are obtained from AEP.

In addition to these seasonal terms, there are indicator variables for each year in the sample. The three final calendar regressors represent weekends and holidays. This holiday variable includes days which are not considered holidays, but immediately precede or follow a holiday and exhibit holiday-like tendencies. Indicator variables for Saturdays and Sundays also capture the behavior over the entire respective day. This model structure allows even atypical patterns such as floating holidays to be represented. Variability from weather conditions is modeled by heating degrees (HD) and cooling degrees (CD), defined as $\mathrm{HD}_t = \max\{T_{\mathrm{ref}}-T_t, 0\}$ and $\mathrm{CD}_t = \max\{T_t-T_{\mathrm{ref}}, 0\}$ where T_t is the mean daily temperature for hour t and T_{ref} is a reference temperature. Following the work of Engel et al. [3] and Pardo et al. [1], we assume $T_{\mathrm{ref}} = 65$ °F. The relationship between total quantity demanded and temperature is nonlinear due to some electric heating; hence, a single temperature variable is not sufficient. Combining these terms in a least-squares regression results in the following model:

$$\begin{aligned}\%\Delta Q_t &= \beta_0 + \beta_1 \mathrm{HD}_t + \beta_2 \mathrm{HD}_t^2 + \beta_3 \mathrm{CD}_t + \beta_4 \mathrm{CD}_t^2 + \beta_5 \mathrm{HOL}_t \\ &\quad + \beta_6 \mathrm{SUN}_t + \beta_7 \mathrm{SAT}_t + \sum_{i=1}^{12}\sum_{j=1}^{3}\sum_{m=1}^{2}\left[\gamma_i \sin(2\pi m \tau_{j,t}) \right. \\ &\quad \left. + \gamma_i \cos(2\pi m \tau_{j,t})\right] + \sum_{k=1}^{4} \delta_k Y_{k,t} + \varepsilon_t. \end{aligned} \tag{1}$$

The variables $\tau_{j,t}$ represent the time of the respective Fourier expansion for a particular observation. For instance, the first hour of the year ($t = 1$) would have $\tau_{1,1} = 1/24$, $\tau_{2,1} = 1/168$, and $\tau_{3,1} = 1/12$, for hour of the day, hour of the week, and month of the year periods, respectively. HOL_t, SUN_t, SAT_t, and $Y_{k,t}$ are binary indicator variables for holidays, Sundays, Saturdays, and the years 2006–2009, respectively, and ε_t is the residual term. The percent difference of the dependent variable is to prevent scaling and heteroscedasticity issues.

Our data for the short-term diurnal regressions consist of hourly electricity demand and daily weather information. Three data sets contain load data for a subset of approximately 105,000, 65,500, and 78,000 AEP commercial and industrial customers from the states Texas, Virginia, and Ohio, respectively, from January 1, 2006 to December 31, 2010 with 43,824 hourly observations. A fourth data set of roughly 16,000 customers in Michigan has 43,080 observations, spanning the same time frame except for December 2009. The average load for the customers in these samples are 183, 1447, 1372, and 1016 MW for Michigan, Ohio, Texas, and Virginia, respectively. Each of these data sets is used to estimate a separate model specific to the location. Of these data, 35,064 (34,320 for Michigan) observations are used to estimate the model coefficients, and the final year of data, consisting of 8760 observations, is used for out-of-sample validation. Maximum and minimum daily temperatures from three weather stations[1] in each state, as obtained from the U.S. National Climatic Data Center, are averaged to determine HD and CD in the model.

Diagnostic tests on the model determined by (1) indicate positive autocorrelation of the residuals, a common issue in time series, suggesting model adjustments are needed. The Durbin-Watson (DW) test of the residuals checks for evidence of correlation; a statistic from this test differing significantly from 2 (with a possible range of 0–4) indicates correlation. Despite the model seeming well fit, with R^2 values of approximately 0.8 for all of the data sets used, these and other statistics such as t-tests are misleading and biased if autocorrelation is present. To correct this issue, autoregressive residual terms with 1-, 2-, 3-, 144-, 145-, and 168-h lags are included in the model. These are determined by analysis of autocorrelation and partial autocorrelation functions. The modified model is given by:

$$\begin{aligned}\%\Delta Q_t &= \beta_0 + \beta_1 HD_t + \beta_2 HD_t^2 + \beta_3 CD_t + \beta_4 CD_t^2 + \beta_5 HOL_t \\ &\quad + \beta_6 SUN_t + \beta_7 SAT_t + \sum_{i=1}^{12}\sum_{j=1}^{3}\sum_{m=1}^{2}\left[\gamma_i \sin(2\pi m \tau_{j,t}) \right. \\ &\quad \left. + \gamma_i \cos(2\pi m \tau_{j,t})\right] + \sum_{k=1}^{4} \delta_k Y_{k,t} + \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} \\ &\quad + \rho_3 \varepsilon_{t-3} + \rho_4 \varepsilon_{t-144} + \rho_5 \varepsilon_{t-145} + \rho_6 \varepsilon_{t-168} + \varepsilon_t. \end{aligned} \tag{2}$$

Terms that are also found in (1) retain their meanings and ε_{t-i} represents the ith lag from time t of the residual term. The coefficients of these autoregressive terms are determined by minimizing the sum of squared errors.

Table 1
Results of ADF test on dependent variable in (3).

State	Test statistic	p-value
Michigan	−2.381	0.020
Ohio	−3.060	0.005
Texas	−2.429	0.018
Virginia	−2.968	0.005

2.2. Annual aggregate model

Although calendar and weather factors best predict short-run levels of electricity demand, macroeconomic changes predominately determine long-term trends in aggregate quantities demanded. For our model, these include the gross state product (GSP) and average annual retail price of electricity. In addition, industrial customers are those most capable of fuel switching in response to the relative prices of electricity and its alternatives; therefore, average annual retail natural gas prices are included to capture this substitutability. Finally, the level of economic activity is correlated to the population in an area. Our model takes the form:

$$\begin{aligned}\%\Delta A_y &= \alpha_0 + \alpha_1 \%\Delta \mathrm{GSP}_y + \alpha_2 \%\Delta P_y + \alpha_3 \%\Delta \mathrm{NGP}_y + \alpha_4 \%\Delta \mathrm{POP}_y \\ &\quad + \alpha_5 T_y + \varepsilon_y, \end{aligned} \tag{3}$$

where A_y is the annual aggregate commercial and industrial quantity demanded, GSP_y is the gross state product, P_y is the annual quantity-weighted mean price of electricity, NGP_y is the annual quantity-weighted mean price of natural gas, POP_y is the state population, T_y is a time trend variable, and ε_y is the error term, all in year y. The percent difference operator from $y-1$ to y is used to make the data stationary. Augmented Dickey Fuller (ADF) unit root tests of the variables find them to be nonstationary, as is typical for such time series. An ADF test rejects the null hypothesis that the percent difference data are nonstationary at a 95% confidence level, so these values are taken to be stationary. Results from these tests for the dependent variable can be found in Table 1—results for the other variables are similar.

[1] We choose weather stations based on data availability and geographic diversity. The locations, by state, of the stations are: Adrian, Chatham, and Gladwin, Michigan, Akron, Bellefontaine, and Jackson, Ohio, Camp Pickett, Clintwood, and Edinburg, Virginia, and Borger, Dell City, and Refugio, Texas.

For the long-run model, aggregate annual electricity demanded quantities of commercial and industrial sectors from 1990 to 2010 are used as the dependent variable. These data are available from the U.S. Department of Energy's Energy Information Administration (EIA), as are the electricity and natural gas price data also used. Electricity prices are state average retail prices in cents per kilowatt hour. The natural gas prices used are state average retail prices in dollars per thousand cubic feet. For both electricity and natural gas, the prices are the mean as weighted by the respective commercial and industrial sector load. GSP is publicly available from the Bureau of Economic Analysis. Lastly, population estimates are from the U.S. Census Bureau.

2.3. Simultaneous use of short- and long-run models

The two models described in Sections 2.1 and 2.2 convey informative predictions separately, but used in conjunction they allow forecasts of hourly demand level further into the future than otherwise possible. The process follows three steps. First, the short-run forecasts are made. Then the long-run growth rates are predicted and applied to the short-run forecasts. Finally, changes in the quantity demanded as a consequence of short-run inter-hourly price changes are calculated using short-run own- and cross-price elasticities.

As most of the regressors in (2) are dependent on the time being forecasted, predictions based on this model are simple to produce. Because the short-run model includes indicator variables for 2006–2009, a base year for predictions is necessary. The year variable must be within the sample period, but the other calendar variables should reflect the actual year of the forecast, to accurately represent when holidays and weekends occur. Once a given time frame is decided, pertinent weather data must be input to (2). This gives a diurnal load profile corresponding to the base year selected, which we denote y_0, with the calendar and weather data for the future year to be modeled, which is denoted $\{\widehat{Q}_t\}$.

The model (3) describes how aggregate annual electricity demand changes as a result of differences in macroeconomic variables. The interannual demand growth rates estimated by this model can be used to improve the short-run forecasts generated by (2) into the future. If we let $\tilde{y}$ denote the future year for which a diurnal load forecast is desired, the base-year load profile is scaled to the future year as:

$$\tilde{Q}_t = \widehat{Q}_t \cdot \prod_{y=y_0+1}^{\tilde{y}} (1 + \%\Delta A_y), \tag{4}$$

where the $\%\Delta A_y$'s are estimated using model (3), based on estimates of future macroeconomic variables. Equation (4) assumes that the change in annual aggregate demand applies uniformly to all hours of the day.

The load profiles given by (4) can be further refined to account for the effect of time-variant pricing schemes using interhourly demand elasticities. If we let $\eta_{t,j}$ denote the elasticity of quantity demanded in hour t relative to a price change in hour j, p_j^b denote an assumed baseline price in hour j, and p_j the actual price in hour j, the resulting change in hour-t quantity demanded is given by:

$$\%\Delta\tilde{Q}_t = \sum_{j=1}^{24} \eta_{t,j} \frac{p_j - p_j^b}{p_j^b}. \tag{5}$$

Equation (5) assumes that the change in hour-t quantity demanded depends on price changes during all 24 h of the corresponding day, although the model can be generalized to consider the effects of less or more hours. The changes in quantity demanded determined in (5) can be applied to the load profiles computed in (4) to arrive at the load profile:

$$\overline{Q}_t = \tilde{Q}_t \cdot \left(1 + \%\Delta\tilde{Q}_t\right), \tag{6}$$

which corresponds to year $\tilde{y}$ and includes interhourly demand elasticities.

For our application of the model, we assume a base year of 2009. The calendar variables are determined based on the specific year forecast. For diurnal forecasts, the 30-year norm temperatures from the same weather stations used for estimating the models are used as the future weather. Some assumptions are necessary to facilitate long-run forecasting as well; independent variables such as GSP and population in (3) must first be estimated before the regression model can be used. We assume these variables grow linearly into the future. Fig. 3 shows the GSP of Virginia for the years 1990–2009, published by the U.S. Department of Commerce's Bureau of Economic Analysis, and the linear fit used to forecast future GSP. The model has an R^2 of 0.9826. Although prices are more volatile than GSP, this approximation is representative of the fits for all the data sets and independent variables, thus a simple linear regression is justifiable. The elasticities used are values reported by Taylor et al. [12], and the baseline price is the predicted price from the linear approximation of retail price used in (3). The results from these two models, and how the price elasticities affect the quantity demand forecasted, are explained further in the Section 3.

3. Results and forecasts

3.1. Model results

Regression coefficients for short-run models (1) and (2) are given in Tables 2–5. Each table contains the results for one particular data set and associated t-values for the models with and without autoregressive residual variables. As expected, when considering the results from (1), holidays and weekends are highly significant for all the data sets. In general, the values for these variables are negative, which follows the intuition that demand is lower during these periods. The Fourier terms representing hour of

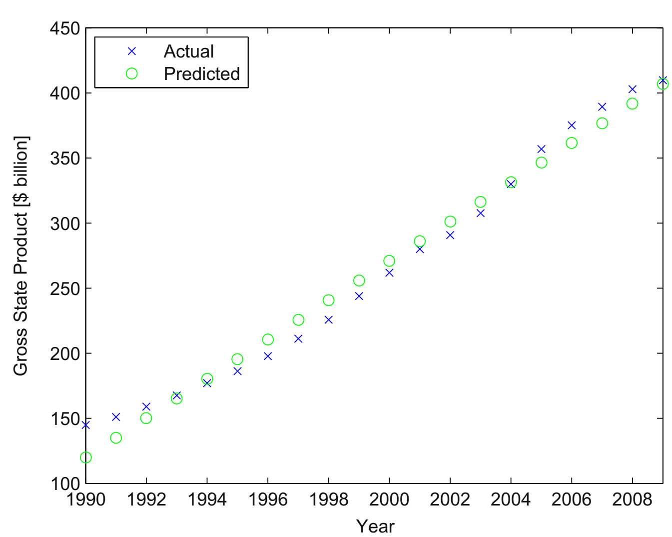

Fig. 3. Actual and predicted values of gross state product in Virginia for the years 1990–2009. $R^2 = 0.9826$.

Table 2
Regression estimates for short-term models in Equations (1) and (2) for Michigan. Regression R^2 's are 0.653 and 0.793, respectively. Model (1) has a DW statistic of 1.078.

	Equation (1)		Equation (2)	
Variable	Coefficient	t-value	Coefficient	t-value
Constant	−0.144	−1.990	−0.089	−9.990
HD	0.001	0.231	−0.003	−0.030
CD	0.025	1.170	0.027	0.002
HD^2	0.000	−0.334	0.000	0.021
CD^2	−0.003	−1.310	−0.002	−0.006
Day Hour 1	4.400	212.000	4.400	2.360
Day Hour 2	−1.380	−66.500	−1.380	−0.741
Day Hour 3	−1.210	−58.100	−1.210	−0.648
Day Hour 4	−1.980	−95.400	−1.990	−1.070
Week Hour 1	0.808	36.000	0.761	0.408
Week Hour 2	−0.497	−10.800	−0.438	−0.235
Week Hour 3	0.834	33.600	0.762	0.408
Week Hour 4	−0.470	−14.300	−0.428	−0.230
Month 1	−0.006	−0.162	−0.012	−0.006
Month 2	−0.002	−0.035	−0.008	−0.005
Month 3	0.006	0.217	−0.037	−0.020
Month 4	−0.001	−0.030	−0.002	−0.001
Y_{200}	0.037	0.874	0.009	0.004
Y_{2007}	−0.007	−0.172	0.007	0.003
Y_{2008}	0.001	0.031	−0.001	0.000
HOL	−0.184	−2.210	−0.063	−0.033
SAT	1.240	13.900	0.942	0.433
SUN	0.393	4.400	0.463	0.213
ε_{t-1}			35.700	0.000
ε_{t-2}			8.750	19.200
ε_{t-3}			1.300	4.700
ε_{t-144}			20.100	0.697
ε_{t-145}			13.700	10.800
ε_{t-168}			30.500	7.360

Table 3
Regression estimates for short-term models in Equations (1) and (2) for Ohio. Regression R^2 's are 0.389 and 0.433, respectively. Model (1) has a DW statistic of 1.848.

	Equation (1)		Equation (2)	
Variable	Coefficient	t-value	Coefficient	t-value
Constant	0.221	2.300	0.195	10.600
HD	0.002	0.290	0.003	0.013
CD	0.003	0.111	0.005	0.001
HD^2	0.000	−1.080	0.000	−0.019
CD^2	0.000	−0.229	−0.001	−0.002
Day Hour 1	4.300	128.000	4.310	1.120
Day Hour 2	−1.590	−47.300	−1.590	−0.413
Day Hour 3	−0.864	−25.700	−0.868	−0.225
Day Hour 4	−1.690	−50.100	−1.690	−0.437
Week Hour 1	0.595	16.400	0.568	0.147
Week Hour 2	−0.206	−2.770	−0.203	−0.053
Week Hour 3	0.688	17.100	0.647	0.168
Week Hour 4	−0.218	−4.100	−0.211	−0.055
Month 1	0.086	1.670	0.075	0.019
Month 2	0.070	1.060	0.053	0.014
Month 3	0.092	2.280	0.076	0.020
Month 4	−0.027	−0.790	−0.027	−0.007
Y_{2006}	−0.196	−2.880	−0.188	−0.037
Y_{2007}	−0.140	−2.060	−0.138	−0.027
Y_{2008}	−0.148	−2.200	−0.144	−0.028
HOL	−0.119	−0.772	−0.007	−0.001
SAT	0.892	4.330	0.670	0.149
SUN	−0.179	−0.888	−0.188	−0.042
ε_{t-1}			1.060	0.275
ε_{t-2}			5.480	1.420
ε_{t-3}			−0.102	−0.026
ε_{t-144}			12.800	3.330
ε_{t-145}			4.130	1.070
ε_{t-168}			15.400	4.000

Table 4
Regression estimates for short-term models in Equations (1) and (2) for Texas. Regression R^2 's are 0.563 and 0.705, respectively. Model (1) has a DW statistic of 1.256.

	Equation (1)		Equation (2)	
Variable	Coefficient	t-value	Coefficient	t-value
Constant	0.243	3.350	0.262	20.800
HD	−0.007	−0.724	−0.017	−0.033
CD	0.005	0.382	0.008	0.014
HD^2	0.000	0.506	0.000	0.015
CD^2	0.000	−0.583	−0.001	−0.020
Day Hour 1	4.630	198.000	4.790	1.820
Day Hour 2	−2.540	−108.000	−2.700	−1.030
Day Hour 3	−1.080	−46.200	−1.130	−0.430
Day Hour 4	−1.230	−52.400	−1.280	−0.485
Week Hour 1	0.457	18.100	0.427	0.162
Week Hour 2	0.016	0.306	−0.058	−0.022
Week Hour 3	0.545	19.500	0.501	0.190
Week Hour 4	−0.103	−2.790	−0.140	−0.053
Month 1	0.009	0.210	−0.005	−0.002
Month 2	0.020	0.324	0.059	0.022
Month 3	0.038	1.140	0.035	0.013
Month 4	0.001	0.051	0.072	0.027
Y_{2006}	0.010	0.216	−0.041	−0.012
Y_{2007}	−0.033	−0.707	−0.005	−0.002
Y_{2008}	−0.051	−1.110	−0.074	−0.021
HOL	−0.361	−3.350	−0.141	−0.037
SAT	0.700	4.900	0.437	0.142
SUN	−0.431	−3.070	−0.394	−0.128
ε_{t-1}			34.300	13.000
ε_{t-2}			5.340	2.030
ε_{t-3}			−0.161	−0.061
ε_{t-144}			18.900	7.150
ε_{t-145}			12.100	4.590
ε_{t-168}			18.800	7.130

Table 5
Regression estimates for short-term models in Equations (1) and (2) for Virginia. Regression R^2 's are 0.544 and 0.726, respectively. Model (1) has a DW statistic of 1.123.

	Equation (1)		Equation (2)	
Variable	Coefficient	t-value	Coefficient	t-value
Constant	0.177	3.210	0.176	20.800
HD	0.003	0.509	0.001	0.008
CD	−0.010	−0.687	−0.017	−0.012
HD^2	0.000	−1.890	0.000	−0.029
CD^2	0.001	0.519	0.001	0.009
Day Hour 1	3.290	173.000	3.300	1.860
Day Hour 2	−0.766	−40.200	−0.767	−0.433
Day Hour 3	−0.788	−41.400	−0.793	−0.447
Day Hour 4	−1.640	−86.300	−1.640	−0.928
Week Hour 1	0.508	24.700	0.475	0.268
Week Hour 2	−0.008	−0.192	0.020	0.012
Week Hour 3	0.524	23.100	0.473	0.266
Week Hour 4	−0.071	−2.360	−0.038	−0.021
Month 1	0.021	0.728	0.002	0.001
Month 2	0.030	0.793	−0.001	−0.001
Month 3	0.038	1.590	0.028	0.015
Month 4	0.014	0.742	−0.044	−0.024
Y_{2006}	−0.062	−1.600	−0.041	−0.017
Y_{2007}	−0.057	−1.470	−0.010	−0.004
Y_{2008}	−0.014	−0.359	−0.003	−0.001
HOL	−0.260	−2.970	−0.237	−0.092
SAT	0.592	5.080	0.397	0.192
SUN	−0.223	−1.950	−0.153	−0.074
ε_{t-1}			29.500	16.700
ε_{t-2}			5.520	3.110
ε_{t-3}			−1.500	−0.844
ε_{t-144}			20.300	11.400
ε_{t-145}			15.600	8.820
ε_{t-168}			30.100	17.000

Table 6
MAPE of models (1) and (2) for one year of out-of-sample data used in the model validation process.

State	Equation (1)	Equation (2)
Michigan	4.326	3.744
Ohio	4.031	3.870
Texas	3.638	3.443
Virginia	4.508	3.725

the day and hour of the week are also significant. These terms together are capable of capturing most of the trend on an hour-to-hour basis; however, variables representing the year-to-year growth are also important, which in part necessitates models for longer time scales for forecasts to be accurate. The weather-related terms—heating and cooling degrees and their squared terms—are most significant for Texas.

Relative summer temperatures in Texas are higher than in the other three states, which may explain the increased importance. Other potential predictors, such as an indicator for daylight savings time and a set of Fourier terms for seasons of the year are found to be insignificant. The remaining variables are used to maintain the same model structure for all four data sets, despite varying levels of significance.

As discussed in the previous section, a second formulation is developed, motivated by the low DW statistics indicating positive residual autocorrelation in (1). With the inclusion of autoregressive residual terms in Equation (2), some changes in the results are observed. Coefficients of some of the original variables change in magnitude and significance, but relative size and direction remain the same for most. In all cases, the constant term remains or increases in significance. The additional autoregressive terms are significant as well. The hour of day, holiday, Saturday, and Sunday terms remain large in relative magnitude. In all cases the R^2 improves from (1) to (2), especially for Virginia and Texas.

Values in Tables 2–5 are estimated using four of the five years of available data, with the remaining observations excluded to use for out-of-sample validation. To check the fitness of the models, we compare the actual observations to the forecasted values for this final year. A common measurement of this deviation is mean absolute percentage error (MAPE), which measures the difference between actual and predicted values averaged over the time horizon. MAPE is given by:

$$\frac{100}{T}\sum_{t=1}^{T}\frac{\left|Q_t - \widehat{Q}_t\right|}{Q_t}, \tag{7}$$

where Q_t represents the actual value at time t and $\widehat{Q}_t$ is the corresponding predicted value. Table 6 details these results for both formulations of the short-run model, where $T = 8760$. The inclusion of autoregressive error terms improves the forecasts significantly, and our MAPE values are comparable to values reported in the literature.

The results of the long-term model of aggregate annual demand are reported in Table 7. We only estimate the long-term model for the four states for which we have data to fit the short-term model. Nevertheless, the long-term model could be estimated for other states, since the required data are publicly available. Coefficients from the regression are the first value of each column and state, and the associated t-values are reported in parentheses. In general, changes in GSP and the price of electricity are the most significant factors in determining changes in quantity demanded. On the other hand the time trend, the final regressor, is less significant. As the data have been made stationary, it is logical that little time-varying trend remains. Our R^2 values are in line with the long-run models estimated by Engle et al. [9]. Although a model with absolute levels (as opposed to percent differences) of annual aggregate demand yields higher R^2 values, the use of such non-stationary data yield inconsistent coefficient estimates and spurious regression results [13].

3.2. Load predictions: the effects of growth and price elasticity

Fig. 4 summarizes how diurnal load patterns change over the long-term when interannual growth rates are applied. It shows a one-week load pattern in June during a base year of 2009, computed using model (2). It also shows the load pattern during the same week in 2014, with growth rates derived from model (3) applied using Equation (4). This load pattern assumes 2.6% and 3.5% annual GSP and electricity price growth, respectively, based on a linear fit to historical data as illustrated in Fig. 3. For purposes of comparative statics, the figure also shows load profiles for the same week in 2014 with different rates of macroeconomic growth. One of the load patterns considers a case in which the economy experiences high growth during the five years, with an annual GSP growth rate of 8%. The other load pattern assumes annual electricity price increases of 5%.

Because the diurnal load data used to estimate the short-term diurnal model represents only a small subset of all commercial and industrial customers in each of the four states, the diurnal load profiles generated by model (2) are scaled to represent aggregate statewide consumption. This is done by comparing the total quantity demand in 2009 in the short-run data sets to aggregate statewide consumption reported by the EIA. If we let A_y denote aggregate annual statewide consumption and Q_t the hourly consumption of the customer subset in the diurnal data, the scaling factor for the year is given by:

$$S_y = \frac{A_y}{\sum_{t=1}^{8760} Q_t}. \tag{8}$$

For instance, the hourly quantity demand by the subset of commercial and industrial customers in Ohio in 2009 is approximately 15% of the total statewide consumption, thus the outputs of model (2) are scaled by a factor of 6.667 to forecast statewide commercial and industrial consumption. This scaling factor is applied to all of the hourly loads generated by the short-term model, and is included in Fig. 4 and all subsequent figures.

Although the days shown in Fig. 4 represent different calendar days of the month in the years 2009 and 2014, they represent the same days of the first week of June. In the base case with long-run average macroeconomic growth rates, natural gas price increases

Table 7
Regression estimates for long-term model in Equation (3). t-values for each coefficient are reported in parentheses.

State	Constant	GSP	P	NGP	POP	T	R^2	DW
Michigan	0.008 (0.157)	0.736 (2.434)	−0.223 (−0.814)	0.101 (1.151)	−2.155 (−0.536)	−0.002 (−0.551)	0.556	2.28
Ohio	−0.029 (−0.599)	0.763 (1.330)	−0.698 (−1.516)	0.046 (0.957)	−0.699 (−0.161)	0.002 (0.728)	0.615	1.41
Texas	0.001 (0.027)	0.281 (1.448)	−0.014 (−0.136)	0.015 (0.447)	0.275 (0.298)	−0.000 (−0.242)	0.278	2.37
Virginia	0.105 (1.701)	−1.377 (−1.444)	−0.730 (−2.046)	0.211 (2.529)	−1.133 (−0.512)	0.002 (0.898)	0.420	2.31

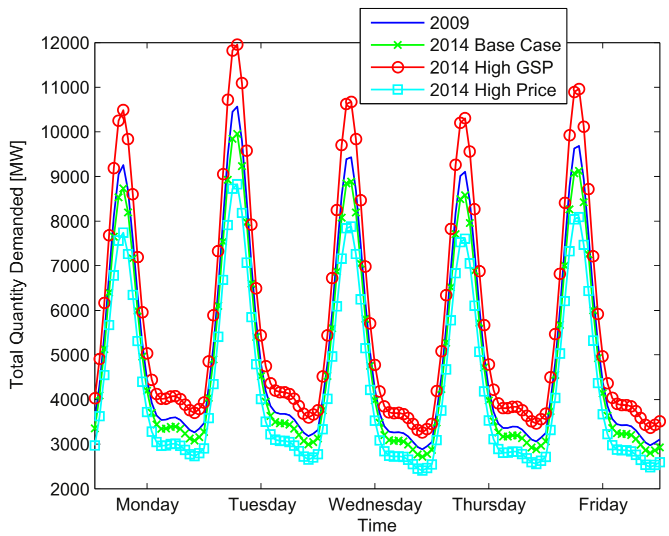

Fig. 4. Modeled diurnal demand of entire commercial and industrial sector in Ohio during the first week of June in 2009 and 2014.

and GSP growth increase electricity consumption. However, there are increases in electricity prices as well, causing a net decrease in consumption relative to 2009. Population growth in Ohio is also forecasted to decline, further reducing demand. This results in a 295 MW reduction in the maximum peak during the week between 2009 and 2014. When electricity prices increase 5% annually, this negative impact dominates the other macroeconomic changes, with further declines in consumption levels. On the other hand, high GSP growth results in increasing quantities demanded.

In addition to modeling the diurnal load profile, we also explore the effects of short-run inter-hourly price changes on demand. This is done using own- and cross-price elasticities of demand, which describe how consumers' electricity demand changes due to price changes during different hours of the day. For instance, if the price during the middle of the afternoon, which tends to be a peak demand period, is higher relative to the price during other hours of the day, an industrial consumer may move production and associated electricity demand away from this period. A utility scheduling generators can use this information to create a price structure which minimizes costs by shifting some demand away from the peak. Given our prediction of the entire diurnal load profile, we can utilize hourly elasticity estimates to capture such effects. Taylor et al. [12] report estimated own- and cross-price elasticities of demand between each pair of hour within a day. Many customers do not face complex tariff structure with hourly prices, thus our analysis uses these elasticities to estimate the effects of simpler pricing schemes.

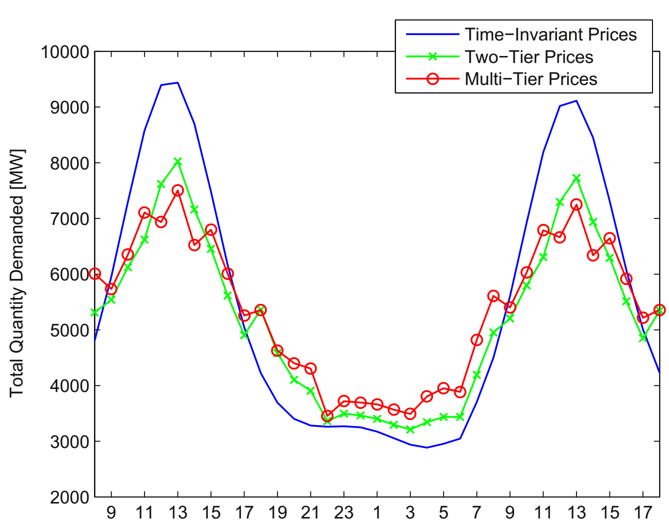

Fig. 5. Modeled diurnal demand of entire commercial and industrial sector in Ohio between 8 am of 3 February, 2009 and 6 pm of 4 February, 2009.

Fig. 5 shows the results of applying these elasticities over a 35-h period in February under three different pricing schemes. The first is a standard time-invariant tariff, with the price of energy being uniform across all hours. The loads in this case are not affected when the elasticities are applied using Equation (5), since all of the hourly price changes are zero. The second case considers a two-tier price structure in which prices between 9 am and 6 pm are 12.5% higher than the price in the time-invariant tariff and prices in the remaining hours are 12.5% lower. The third case considers a more complex multi-tiered price structure. From 8 pm to 6 am, prices are 20% lower than the baseline price in the time-invariant tariff, from 7 to 8 am prices are 12.5% lower than the baseline, from 11 am to 5 pm prices are 12.5% higher than the baseline, and in the remaining hours prices are 6.25% above the baseline. Both of these price tariffs result in shifting of demand away from high-price hours toward lower-price hours and a flattening of the load profile. The multi-tiered tariff yields a 20% reduction in the peak quantity demanded and an overall consumption decrease of 2% relative to the time-invariant tariff.

4. Conclusion

In this paper we develop a model to represent the level of electricity demand in commercial and industrial sectors as a combination of short-term and long-term trends. The short-term model, which is based on weather and calendar data, can predict diurnal load profiles within a year with 95% accuracy in out-of-sample forecast validation. We employ Fourier series of varying periodicities to capture the effects of different seasonalities. This model is of a percent difference autoregressive form. Incorporating forecasts of weather variables, instead of historical averages, and evaluating how forecasts change is of future interest, as well as investigating the impact of other weather conditions, such as humidity, wind speed, and ambient sunlight.

In addition, a second model of aggregate electricity demand levels measures interannual growth rates based on population, GSP, and price changes. In combination, these allow for fine-grain forecasts into the future. Presently, the model assumes, when shifting the diurnal predictions by this interannual growth rate, that the change affects all hours uniformly. Whether or not peak and off-peak loads do indeed grow at equal rates in the long-term is a topic worthy of further investigation. Finally, changes in quantity demanded are observed when different pricing policies are enacted based on measurements of commercial and industrial customers' hourly own- and cross-price elasticity of demand. Peak loads are found to shift up to 20% with only a modest change in relative prices.

The models are calibrated, validated, and demonstrated based on four states for which we have hourly data to fit the short-term model. Results from each of these states, in terms of model fit and predictive ability, are similar. Thus, the model seems applicable for a variety of underlying data sets. The same techniques could be used to forecast and model electricity demands in other states or countries, if similar data are available. The results of this model can be a useful input for long-term generation, transmission, and distribution capacity planning done by utilities and system operators. Utilities, system operators, and policy makers may also be interested in using this model to study the effects of different pricing tariffs on electricity demand and ancillary impacts associated with that energy use.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 1029337. The authors would like to thank Alan Graves of American Electric Power, who provided demand data, and Armin Sorooshian, who provided helpful suggestions and comments.

References

[1] Pardo A, Meneu V, Valor E. Temperature and seasonality influences on Spanish electricity load. Energy Economics 2002;24:55–70.
[2] Mirasgedis S, Sarafidis Y, Georgopoulou E, Lalas DP, Moschovits M, Karagiannis F, et al. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 2006;31:208–27.
[3] Engle RF, Mustafa C, Rice J. Modelling peak electricity demand. Journal of Forecasting 1992;11:241–51.
[4] Ramanathan R, Engle RF, Granger CWJ, Vahid-Araghi F, Brace C. Short-run forecasts of electricity loads and peaks. International Journal of Forecasting 1997;13:161–74.
[5] Taylor JW, Buizza R. Using weather ensemble predictions in electricity demand forecasting. International Journal of Forecasting 2003;19:57–70.
[6] Mohamed Z, Bodger P. Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy 2005;30:1833–43.
[7] Amarawickrama HA, Hunt LC. Electricity demand for Sri Lanka: a time series analysis. Energy 2008;33:724–39.
[8] Bianco V, Manca O, Nardini S. Electricity consumption forecasting in Italy using linear regression models. Energy 2009;34:1413–21.
[9] Engle RF, Granger CWJ, Hallman JJ. Merging short-and long-run forecasts: an application of seasonal cointegration to monthly electricity sales forecasting. Journal of Econometrics 1989;40:45–62.
[10] Ringwood JV, Bofelli D, Murray FT. Forecasting electricity demand on short, medium and long time scales using neural networks. Journal of Intelligent and Robotic Systems 2001;31:129–47.
[11] Taylor JW, de Menezes LM, McSharry PE. A comparison of univariate methods for forecasting electricity demand up to a day ahead. International Journal of Forecasting 2006;22:1–16.
[12] Taylor TN, Schwarz PM, Cochell JE. 24/7 hourly response to electricity real-time pricing with up to eight summers of experience. Journal of Regulatory Economics 2005;27:235–62.
[13] Granger CWJ, Newbold P. Spurious regressions in econometrics. Journal of Econometrics 1974;2:111–20.