Quantile Regression for Peak Demand Forecasting*

Charles E. Gibbons Ahmad Faruqui
The Brattle Group

July 31, 2014

Abstract

We demonstrate that annual peak demand days are characterized by both extreme values of predictors (such as weather) and large unpredictable "shocks" to demand. OLS approaches incorporate the former feature, but not the latter, leading OLS to produce downwardly-biased estimates of the annual peak. We develop a new estimation procedure, optimal forecast quantile regression (OFQR), that uses quantile regression to estimate a model of daily peak demand, then uses a loss function framework to estimate a quantile to predict the annual peak. We compare the results of the OLS and OFQR estimation approaches for 32 utility zones. While the OFQR approach is unbiased, OLS under-forecasts by nearly 5% on average. Further, OFQR reduces the average absolute percent error by 43%. A bootstrapping procedure generates forecast intervals with accurate 95% coverage in sample and 87% coverage out of sample.

1 Introduction

The electricity business is very capital intensive. Electricity generation, transmission, and distribution require investment in power plants, transmission lines and towers, distribution substations, transformers, switches, circuits, feeders, and a variety of other equipment. These investments have long lead times and are expensive; planning and execution must begin well before the additional capacity is needed.

A key indicator of the need for new generation and transmission capacity is the forecast of peak load. Though distribution capacity is driven by local considerations, it is ultimately related to

*The authors would particularly like to thank the various utilities and ISOs that provided data for this project. They also thank Neil Lessem, Sanem Sergici, seminar participants at The Brattle Group and the Center for Research in Regulated Industries Eastern and Western conferences, and webinar participants for their helpful comments.

the system peak as well. Peak load forecasts are used to derive reserve margins and perform resource adequacy computations. These calculations are then used for resource planning. Consequently, the effectiveness and efficiency of subsequent investments depend upon the accuracy of the peak load forecast.

It is problematic both to over-forecast and to under-forecast peak demand. In the former case, excess capacity is built and the costs are eventually recovered from retail customers in the form of higher prices. In the latter case, capacity is underbuilt, creating a threat to reliability with the attendant possibility of blackouts. These conditions are economically costly. Furthermore, when a utility is part of an independent system operator (ISO), peak load forecasts are made by the ISO and provide a backdrop against which energy and capacity markets function.[1] Inaccurate forecasts of peak load can disrupt markets and electricity supplies, lead to inefficient investments, and inflict substantial economic costs.

The most widely used method for forecasting peak loads is ordinary least squares (OLS). These models typically use weather conditions, economic activity, and electricity prices as drivers of peak load, though sometimes the latter two variables are replaced by energy sales. OLS models predict the average level of demand expected for given values of the predictors included in the model. While OLS forecasts can reflect that annual peak days tend to experience "extreme" values for the predictors, such as high temperatures, annual peak days are decidedly *not* average, even conditional on these observable factors.

Instead, we show that large, unobservable "shocks" to demand distinguish annual peak loads from other, merely high loads. Indeed, the median estimated shock on an annual peak day in our sample is equal to the 93rd percentile of estimated non-annual peak day shocks. These shocks must be accounted for in forecasts of annual peak demand, otherwise the predictions will be downwardly biased. Because OLS cannot incorporate positive, unobserved shocks, an approach entirely reliant on this methodology will be inaccurate.

This paper introduces a new technique for forecasting annual peak demand: optimal forecast quantile regression (OFQR). Standard quantile regression (QR) has been used in a number of applications in the electricity industry, but, as far as we know, it has not been used to forecast

[1]Sometimes the ISOs are referred to as Regional Transmission Organizations (RTOs). There are minor technical differences between the concepts, but, for all practical purposes, they are the same.

peak demand. While the regression part of QR can have the same linear form as OLS, the choice of quantile for estimation enables this approach to model an arbitrary conditional percentile, rather than the conditional mean, of daily peak demand.

The major methodological contribution of this paper is to establish a loss function framework that uses only annual peak days to estimate the optimal quantile for the model, while all days are used to estimate the coefficients of the regression itself. The effectiveness of this approach in general and in comparison to the standard OLS approach is demonstrated by estimating OLS and OFQR regressions of the same functional form for 32 utility zones. Using out-of-sample metrics, the OFQR approach is found to outperform OLS in 75% of the zones, reducing the median zone mean absolute percent error by 25%. Across zones, OFQR is unbiased while OLS under-forecasts by 4.8% on average and OFQR has 43% smaller mean absolute percent errors.

In Section 2, we discuss applications of QR in the electricity sector and in forecasting more generally. Our contribution begins in Section 3 by describing the data used in our meta-study. Section 4 provides an intuitive exposition of the role of unobservable shocks in determining the annual peak and proposes two key features distinguishing the annual peak demand day. These features are confirmed using the data from our meta-study. Our estimation procedure is formally presented in Section 5, as is our bootstrapping procedure for generating forecast intervals. In Section 6, we present the results of our meta-study. We discuss the implications of these results in Section 7 and offer guidance for the practitioner.

2 Relevant literature

Quantile regression was first proposed by Koenker and Bassett (1978). Its primary use has been to investigate whether the effects of treatment or other predictors vary across conditional quantiles of an outcome of interest (for examples, see Koenker and Hallock, 2001).[2] An early application of QR in the electricity industry by Hendricks and Koenker (1992) follows this approach. In their paper, the authors examine the impact of demographic factors, appliance ownership, and weather on quantiles of hourly household energy usage for residential customers in Chicago. Kaza (2010) performs a similar analysis using survey data from the U.S. Energy Information Administration.

[2]Because QR minimizes a weighted sum of absolute, rather than squared, residuals, it can be more robust to outliers than OLS; Worthington and Higgs (2013) use QR in their study of electricity spot prices for this reason.

While these studies aim to understand determinants of individual-level load, our goal is to forecast system load.

In terms of forecasting, QR methods have been used to generate probabilistic forecasts of electricity prices (Bunn et al., 2013; Jónsson et al., 2013) and wind farm output (Bremnes, 2004). Cabrera and Schulz (2014) forecast percentiles of quarter-hourly system load. While this type of model is important for short-term dispatching, we are interested in annual peaks, which inform medium- and long-term planning decisions.

Furthermore, the use of QR to produce probabilistic forecasts is qualitatively different from our application. Probabilistic forecasts using QR estimate deterministic percentiles of demand, say, to quantify the risk surrounding a primary forecast, which is based upon an average or median value. We use QR to create the primary forecast itself, choosing the quantile optimally to forecast the annual peak using a model that includes observations from all days.[3] We characterize the uncertainty in our forecasts using bootstrapping methods.

Perhaps closest to our work is Soyiri, Reidpath and Sarran (2013). In their paper, the authors estimate a QR model of daily asthma-related emergency room admissions in London. They use 16 months of data for estimation and evaluate the model using 8 months of out-of-sample data. The authors define a peak event as being a day with admissions exceeding the 90th *unconditional* percentile of admissions and estimate a (conditional) QR model at that percentile. They find that this model is able to predict days of peak admissions well in- and out-of-sample.

3 Data

To judge the broad applicability of the OFQR method in forecasting peak demand, we perform a meta-study covering an array of utilities from the United States. A number of ISOs/RTOs and utilities provided data for our research, all under the condition that their contributions remain anonymous. Hence, all results in this paper have been anonymized. The participants are divided into 32 "zones," a term that we use to mean either a utility proper, an entire ISO/RTO, or a region within an ISO/RTO.[4] The zones span eight of the nine U.S. Census divisions.[5]

[3]A general discussion of peak load forecasting methodology is beyond the scope of our paper and we refer interested readers to Jebaraj and Iniyan (2006) and Suganthi and Samuel (2012).

[4]We either consider an ISO/RTO in its entirety or broken into regions, not both.

[5]Only the East South Central (Alabama, Kentucky, Mississippi, and Tennessee) is not represented in our data.

Though our goal is to forecast annual peak demand, our models are estimated using daily data. These data represent aggregate zone sales and demand for all customer classes. Ultimately, we forecast the summer peak, focusing our attention to observations from the months May, June, July, August, and September. As discussed further in Section 6.1, our estimation period for most zones begins in 1999 and our evaluation period ends in 2013.[6] In the tables that follow, our data are limited to the summer months in these years.

Most of our data are provided by the zones themselves. The outcome of interest is daily peak demand and predictors used in the models of Section 6.1 include:

- Monthly electricity sales
- Daily temperatures. Many zones provide these data to us directly. Twenty zones do not provide any weather data and we obtain daily minimum and maximum temperatures for the largest city in the zone's service area from the National Weather Service (Menne et al., N.d.).[7]
- Daily humidity[8]
- Monthly heating and cooling degree days[9]
- Quarterly economic activity index.[10]

Table 1 summarizes the variation across zones and Table 2 gives the means of four key variables by zone.[11] The former table summarizes the latter table; the first four columns of Table 1 give the mean, minimum, maximum, and standard deviation of their respective columns in Table 2. The standard deviation in the fourth column is a measure of between-zone variation. The standard deviation in the final column of Table 1 is the average intra-zone standard deviation, a measure of within-zone variation. The annual peak ratio is the ratio of the year's annual peak demand divided by the year's average daily peak demand multiplied by 100.

[6] Data series from 22 zones begin in 1999 or earlier, 2 begin in 2000, and 8 begin in 2003. Table 2 gives the number of observations by zone.

[7] For all but Zone 14, we have daily minimum and maximum temperatures. For Zone 14, we have mean daily temperature. Some temperature series obtained from the National Weather Service contain missing values; the numbers of observations listed in Table 2 count the observations with temperature data.

[8] We have these data for Zones 4, 5, 7, 12, 13, 14, 20, 21, 26.

[9] We have these data for Zones 14 and 19.

[10] We have these data for Zones 1, 2, 3, 6, 8, 9, 10, 11, 14, 17, 18, 22, 23, 24, 25, 27, 29, 30, 31, 32.

[11] For Zone 14, we do not have daily minimum and maximum temperatures. See the data description above.

There is substantial variation in the size of the zones in our data, ranging from average monthly sales of 804.5 GWh to 11,308.5 GWh. Average daily peak demand also varies from 1.4 GW to 18.0 GW. The ratio of the annual to average peak varies more across zones in our sample than within a zone from one year to the next. In the case of temperatures, we find less variation across zones than within a given zone.

Table 1: Summary of key variables across zones

	Mean	Min	Max	SD (B)	SD (W)
Annual peak ratio	135.5	116.6	161.0	10.9	5.0
Daily peak (GW)	6.7	1.4	18.0	5.6	1.1
Monthly sales (GWh)	3,998.0	804.5	11,308.0	3,346.0	448.6
Temp., min.	64.2	57.0	79.3	6.3	7.6
Temp., max.	84.1	77.5	102.8	5.9	8.0

Notes: Standard deviations (SD) are measured between (B) and averaged within (W) zones. The annual peak ratio is the annual peak divided by the year's average daily peak times 100.

4 The determinants of annual peak demand

Daily peak demand can be decomposed into a component that is predictable, based upon factors like weather, and an unpredictable "shock" to demand. The key insight of our paper is that annual peak demand occurs when both the predictable and unpredictable components of daily peak demand are high. This result provides the motivation for using QR to model the annual peak.

In this section, we begin with a simple simulated example illustrating our claim. We show that, even when *daily* peak demand follows a linear model and is well predicted using OLS, *annual* peak demand is not well predicted by OLS due to the presence of large unpredictable shocks on certain days. We offer evidence of this phenomenon in our zonal data. We conclude by discussing how other approaches to forecasting peak demand fail to incorporate both the predictable and unpredictable components of annual peak demand and thus provide biased forecasts.

4.1 A simple example

Figure 1 plots simulated peak demand data. The data are generated such that daily peak demand is a linear function of temperature plus a shock; in other words, a standard linear model. These

Table 2: Means of key variables by zone

Zone	Obs.	Peak ratio	Daily peak	Monthly sales	Temp., min.	Temp., max.
1	2,265	139.8	4.7	2,831.0	66.1	83.2
2	2,265	144.4	2.7	1,667.0	60.3	82.4
3	2,119	138.1	2.5	1,551.0	61.9	82.6
4	1,622	129.4	18.0	10,295.0	71.7	91.5
5	1,622	128.4	2.0	1,160.0	70.8	90.3
6	2,265	135.1	4.0	2,430.0	59.6	82.5
7	1,622	125.5	1.4	804.5	69.1	90.9
8	1,944	128.5	2.2	1,437.0	60.4	80.5
9	2,259	145.3	5.8	3,590.0	62.2	81.4
10	2,265	134.7	13.4	7,999.0	64.2	84.6
11	2,202	161.0	3.7	2,164.0	58.6	81.2
12	1,622	117.0	4.0	2,345.0	75.6	92.1
13	1,622	120.8	14.7	8,739.0	75.0	89.6
14	2,295	141.8	18.0	11,157.0		
15	2,142	137.7	6.3	3,961.0	58.6	77.5
16	1,683	141.5	3.1	1,692.0	61.7	78.4
17	2,250	137.8	2.1	1,286.0	60.4	79.6
18	2,101	132.9	1.4	813.4	60.0	81.9
19	2,295	142.1	1.6	884.0	60.6	82.0
20	1,622	123.0	8.9	5,120.0	71.9	92.1
21	1,622	117.4	1.6	1,027.0	68.0	92.9
22	2,265	139.9	9.3	5,886.0	57.1	77.6
23	2,265	129.4	17.9	11,308.0	60.7	80.7
24	2,177	137.1	2.0	1,281.0	57.8	79.4
25	2,259	130.7	6.4	4,020.0	57.0	79.8
26	2,142	127.1	5.2	2,688.0	79.3	102.8
27	2,265	134.6	5.3	3,352.0	57.8	79.6
28	2,295	116.6	17.7	9,317.0	75.9	87.5
29	2,265	143.2	4.9	2,956.0	62.0	82.2
30	2,265	150.6	6.9	4,204.0	63.5	80.9
31	2,265	153.4	1.8	1,056.0	60.9	80.6
32	2,265	152.2	14.4	8,912.0	61.4	79.0

Notes: Daily peak is measured in GW and monthly sales is measured in GWh. The peak ratio is the annual peak divided by the year's average daily peak times 100.

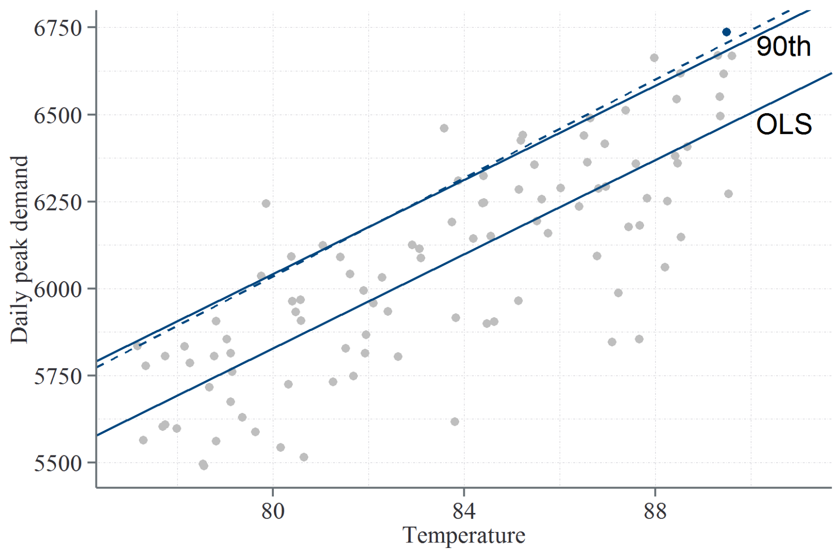

Figure 1: Simulated year of daily peak demand with three models

data points are meant to represent a single year of data. All but one point is gray; only the annual peak is colored blue.

The goal of our model is to best predict this single point. With this goal in mind, it becomes obvious that OLS will not provide an adequate prediction. Indeed, the OLS line shown in Figure 1 goes through the middle of the cloud of points and is relatively distant from the annual peak. While OLS is the correct model of *daily* peak demand for these simulated data, OLS does not capture the features of *annual* peak demand.

Proposition (Key features of annual peak demand)**.** *Annual peak demand exhibits two key features:*

- *It occurs on a hot day (though not necessarily the hottest day of the year).*
- *It is higher than would be expected even on a hot day due to the presence of a large shock (though not necessarily the largest shock of the year).*

Notice that neither of these conditions is sufficient alone. For instance, there are two days in the simulated data with temperatures at least as severe as those witnessed on the annual peak day. Yet, peak demands on those days are below that of the annual peak. Similarly, there are days with larger unpredictable shocks to demand, but, due to lower temperatures, they experience lower daily peak demand. The annual peak is driven by both of these features and a modeling approach must account for both to be accurate.

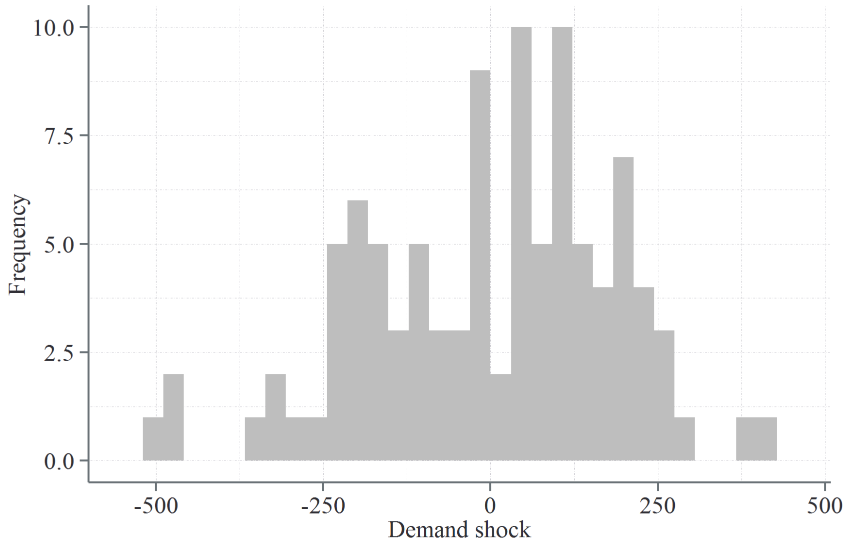

Figure 2: Simulated year of daily peak demand shocks (*i.e.*, OLS residuals)

Figure 2 plots the distribution of the residuals from the OLS model. Because OLS predicts the expected daily peak demand on a given day, it implicitly assumes a shock of size 0. Suppose that we find the 90th percentile of the OLS residuals (representing the shock terms) and add this value to the OLS predictions, effectively shifting the OLS line in a parallel fashion. This line is depicted in solid blue and labeled "90th" in Figure 1. This line comes much closer to the annual peak demand point because it considers both a hot temperature (*i.e.*, the actual temperature on the annual peak day) and a large shock (here, the 90th percentile OLS residual).

QR does not simply shift the OLS predictions by a fixed amount, however. The estimation procedure is more complicated and changes the slope of the line in addition to its intercept (see Section 5.1). The actual 90th quantile QR line is the dashed line labeled "90th" in Figure 1. Though we only use the parallel shift of an OLS line as an expository device, the similarity of this line with the actual QR model reveals that it offers accurate intuition.

4.2 Evidence in the data

The previous section uses simulated data to provide an intuitive illustration of the importance of both the predictable and unpredictable components in determining annual peak demand. In this section, we provide evidence of the two features given in our Proposition using our zone data.

To the first feature, Figure 3 shows the distribution of zone-demeaned daily minimum and

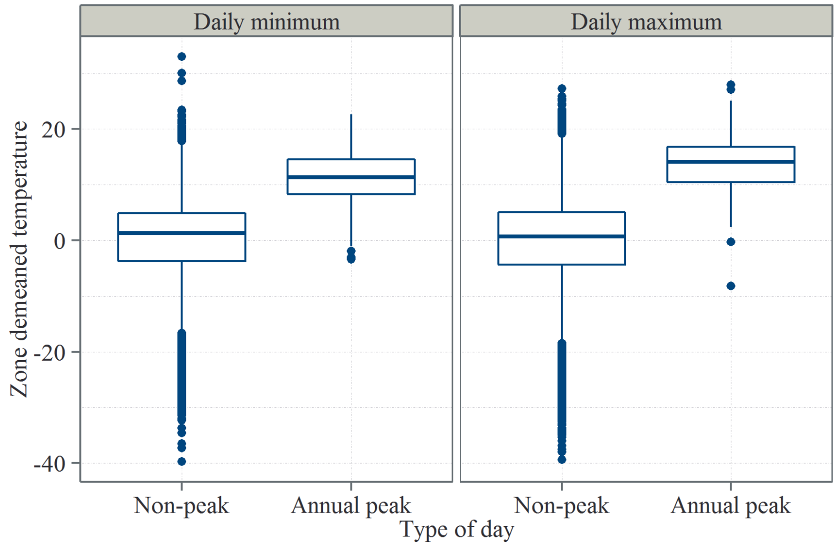

Figure 3: Distribution of temperatures on annual peak days relative to non-peak days

Notes: All boxplots in the paper use the 25th and 75th percentiles as boundaries for the box with the median indicated inside. The "whiskers" extend to the highest value within 1.5 times the interquartile range. Further values are indicated as points.

maximum temperatures for days that correspond to annual peak demand and those that do not. Peak days do not necessarily have the very highest temperatures, but they are generally much higher than on non-peak days. For a zone, minimum temperatures on peak days are on average 11.2 degrees higher than those on non-peak days and maximum temperatures are on average 13.7 degrees higher. The median demeaned temperatures for annual peak days exceed the 95th percentile of demeaned temperatures on non-peak days.

The story is similar for the shock term. Figure 4 plots the in-sample residuals from the OLS models considered in Section 6, which have been standardized by zone to achieve consistent scaling. The median residual among peak days exceeds more than 90% of residuals on non-peak days.[12]

Our findings in the data confirm the intuition developed using the simplified example of Section 4.1 and presented in our Proposition: both weather and shocks take on relatively extreme values on annual peak demand days.

[12]The median annual peak standardized residual is 1.45, which corresponds to the 93rd percentile of a Normal distribution.

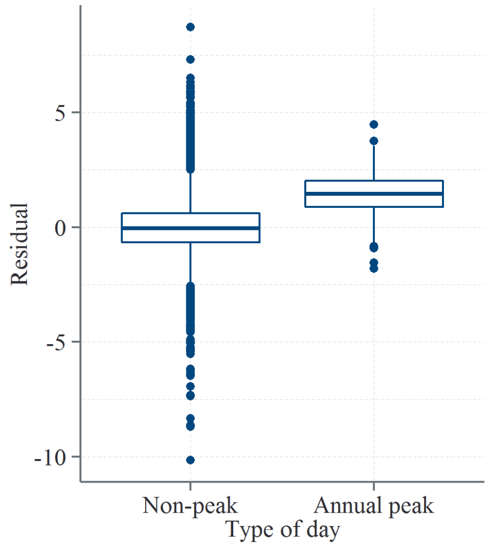

Figure 4: Distribution of OLS residuals on annual peak days relative to non-peak days

4.3 Approaches to prediction

It is common for utilities to use standard OLS models to forecast peak demand (see, as examples ISO New England, 2014; Resource Adequacy Planning Department, 2014). The simple example in Section 4.1 shows that this approach is inadequate, however. The OLS model of Figure 1 can account for extreme temperatures, but it cannot account for large shocks. As the figure makes clear, the OLS prediction of the annual peak will be downwardly biased. The modeling results presented in Section 6.3 confirm this conclusion.

Some utilities recognize that peak demand days are extreme and estimate OLS models using only days with extreme weather conditions. This would be equivalent to erasing the leftmost points in Figure 1 and running OLS on this limited sample. Still, OLS passes through the middle of the remaining points and away from the annual peak. While this approach focuses a utility's attention on severe weather days, it does nothing to account for the shock term.

To account for the shock term through sample selection, the utility would need to eliminate observations with "small" (less than 90th percentile, say) shocks. Combining extreme versions of these two selection procedures, a utility might run OLS using only annual peak demand days, which would account for both features of the annual peak, but would limit the sample to a mere handful of observations.

Instead, as we have already seen, QR can account for both the shock component and weather's contribution to peak demand without arbitrarily discarding observations from the sample or focusing on annual peak days alone.

5 Estimation and inference

QR is often used to estimate the impacts of one or more predictors of interest across a range of quantiles. Rarely is QR used for prediction or forecasting. Furthermore, the quantile of interest is usually determined *ex ante*; we are not aware of any other work that considers the quantile itself a parameter to be estimated. In this section, we develop an approach that estimates a QR model at an optimal quantile, as chosen by a loss function. We call this procedure *optimal forecast quantile regression*. We also provide a bootstrapping method to produce forecast intervals.

5.1 Estimation of a QR model for a given quantile

For a given quantile τ, we follow the standard approach to estimate the QR model. QR minimizes the tilted absolute value or "check" function $\rho_\tau(\cdot)$, which asymmetrically weights the residuals from the model to a degree that depends upon τ (Koenker and Bassett, 1978):

$$\rho_\tau(e) = \begin{cases} (\tau - 1)e & \text{if } e < 0 \\ \tau e & \text{if } e \geq 0, \end{cases} \qquad 0 < \tau < 1.$$

Consider T observations, with outcome y_t (*e.g.*, daily peak demand) and predictors x_t (*e.g.*, daily maximum and minimum temperatures). A linear QR model is estimated by solving:

$$\min_{\hat{\beta}_\tau} \sum_{t=1}^{T} \rho_\tau \left(y_t - x_t' \hat{\beta}_\tau \right). \tag{1}$$

There is no closed-form solution for $\hat{\beta}_\tau$; instead, these parameters can be found using linear programming. Estimation routines are available in conventional statistical software, including `R`, Stata, and SAS. It should also be noted that $\hat{\beta}_\tau$ is indexed by τ because these parameters are estimated for a specified quantile. This notation becomes useful in the following section.

5.2 Optimal choice of τ

There is no theory that directs us to choose a particular quantile for predicting peak demand. As we see in Section 6.2, the "best" choice of quantile can vary substantially across zones. The best choice of quantile is the one that leads to the best prediction of annual peak demand; that is, it is the quantile that minimizes a loss function.

Estimation of the QR model's coefficients in Equation 1 uses daily data. Because our goal is to accurately predict the *annual* peak, the loss function used to choose the optimal quantile is with respect to annual peak demand. The loss function, then, depends upon each year's annual peak and the highest predicted daily peak for the year. The day of the actual annual peak may differ from the day giving the highest prediction of daily peak demand (hence the distinction between t ant t' in Equation 2 below).[13]

There are several obvious contenders for an appropriate measures of loss: mean prediction error, mean absolute prediction error, and mean squared prediction error, plus similar calculations using percent prediction errors. The most appropriate choice depends upon the use of the forecasts. For our analysis in Section 6, we choose mean percent prediction error as our measure of loss to obtain unbiased forecasts.

Suppose that each (day) t is a member of some (year) $a \in 1, \ldots, A$.[14] For some loss function $L(\cdot)$, the optimal quantile $\hat{\tau}$ is given by

$$\hat{\tau} = \operatorname*{argmin}_{\tau} \sum_{a=1}^{A} L\left(\max_{t \in a}\{y_t\}; \max_{t' \in a}\left\{x'_{t'}\hat{\beta}_\tau\right\}\right). \tag{2}$$

This minimization problem is most clearly, though not necessarily most efficiently, solved using a grid search.[15] We restrict our attention to integer values of τ, which are most easily explained to the regulator of a utility, a typical consumer of these forecasts. Furthermore, a narrower grid will produce flat spots of the minimand and an (arbitrary) choice may be required.

Method 1 (Optimal forecast quantile regression (OFQR) procedure)**.** *Estimation of a model for annual peak demand occurs in two steps:*

[13] If the loss function compared the annual peak to the prediction on that same day, overfitting tends to result.

[14] If a utility was interested in monthly peaks, $1, \ldots, A$ could instead be month-year pairs.

[15] Because regressions at various quantiles for a particular data set can, and almost always do, cross, the minimand is not a monotonic function of τ.

1. *For each $\tau \in [\tau_0, \tau_1]$, estimate $\hat{\beta}_\tau$ according to Equation 1.*

2. *Find the $\hat{\tau} \in [\tau_0, \tau_1]$ that solves Equation 2 given the set $\left\{\hat{\beta}_\tau\right\}$.*

The lower bound τ_0 can be set such that it is not binding. In the general QR literature, researchers have been wary of estimating models with $\tau > 95\%$ and we recommend that forecasters respect this wisdom.

5.3 Using the bootstrap to produce forecast intervals

We use a bootstrapping procedure to calculate intervals for our forecasts. In the exposition and results here, we focus on two sources of uncertainty:

- Coefficient and optimal quantile values are estimated rather than known

- Prediction errors that would arise even from a model with known coefficients.

We do not consider uncertainty in the predictors because, in our analysis, we focus on "oracle" forecasts (see Section 6.3) where the predictors are assumed to be known. If the forecasting procedure for the predictors provides a distribution of possible values, then incorporating this source of uncertainty would be straightforward.[16]

The each bootstrap iteration begins by drawing a random sample of the data with replacement with the same number of observations as in the actual data. To account for parameter estimation uncertainty, the model is re-estimated for each bootstrap sample. To incorporate prediction errors, bootstrap prediction errors are chosen from the observed prediction errors with replacement and are added to the bootstrap model's predictions. Each bootstrap iteration provides a set of annual peak demand forecasts. Forecast intervals at the $100(1-\alpha)\%$ level are calculated by finding, for each year, the $100 \times \alpha/2$ and $100 \times (1-\alpha/2)$ percentiles of the bootstrap forecasts.

One additional consideration is serial correlation. There should be no direct serial correlation between the prediction errors for annual peaks, though there likely is serial correlation between daily peak demand error terms. Because we forecast annual peaks, accounting for serial correlation among daily peaks is less important, but doing so is straightforward using a block bootstrapping

[16]For an example of a comprehensive forecasting procedure that includes bootstrapping, see Hyndman and Fan (2010).

approach (Davison and Hinkley, 1997). Rather than randomly drawing separate days, the block bootstrap draws blocks of days. The length of the block should reflect the degree of dependence in the data. Block lengths that are too long reduce power, rendering forecast intervals unnecessarily wide, with the opposite result arising if blocks are too short (assuming positive serial correlation).

To summarize:

Method 2 (Bootstrapping $100(1-\alpha)$ forecast intervals)**.** *The bootstrapping method follows these steps:*

1. *Draw R bootstrap samples from the in-sample data, possibly in blocks, with replacement to produce data sets with the same number of observations as the original data.*[17]

2. *For each bootstrap sample,*

   (a) *Re-estimate the model and calculate predictions for the annual peaks.*

   (b) *For each annual prediction, sample from the observed prediction errors and add this error to the predicted annual peak.*

3. *Find the* $100 \times \alpha/2$ *and* $100 \times (1-\alpha/2)$ *percentiles of the bootstrap forecasts for each year.*

6 Results

Now, we turn to the results of our meta-study. We begin by describing the functional form for the models that we use for all zones. Though we are not particularly interested in the coefficients of the predictors in these models, the estimated optimal quantiles provide evidence of how different the OFQR and OLS forecasts are expected to be; we provide these estimates for each zone. Next, we describe how we evaluate the forecasting performance of the two approaches and present those comparisons. Lastly, we document the performance of our bootstrapping procedure for generating forecast intervals.

[17]When the block bootstrap is used, each bootstrapped data set has the same number of blocks as the original data. Due to the presence of partial weeks at the beginning or end of the May-September period each year, the number of days may not be the same.

6.1 Model functional form

In this section, we develop models for the annual summer peak using daily peak demand. While OLS and OFQR are two different estimation procedures, the underlying model can be the same for both. Indeed, for all zones and for both approaches, we use the same model. We do not endeavor to find the best functional form for each (or any) zone's data and model selection criteria are beyond the scope of our work. Each model of daily peak demand includes the following predictors:

- Month fixed effects
- Day-of-week fixed effects
- Indicator that the day is a weekday holiday
- Monthly energy sales
- Daily temperature (contemporaneous and lagged).[18]

When the data are available (see Section 3), we also include:

- Daily humidity
- Monthly heating and cooling degree days
- Economic activity index.

This specification is typical of that used by utilities and ISOs (see, as examples ISO New England, 2014; Resource Adequacy Planning Department, 2014).

The models are estimated using ten years of daily observations from the summers of 1999 through 2008. Recall that some zones do not begin reporting data in 1999 (see Section 3). In these cases, we begin the estimation period as early as possible and end the estimation period in 2008.

6.2 Estimation results

Though particular coefficients are not of interest to us, we are interested in the optimal quantiles selected by our procedure. A QR model for the 50th quantile relates to an OLS model in the same

[18] For all but Zone 14, the models include minimum and maximum daily temperatures plus lags of each. The model for Zone 14 includes mean daily temperature and its lag.

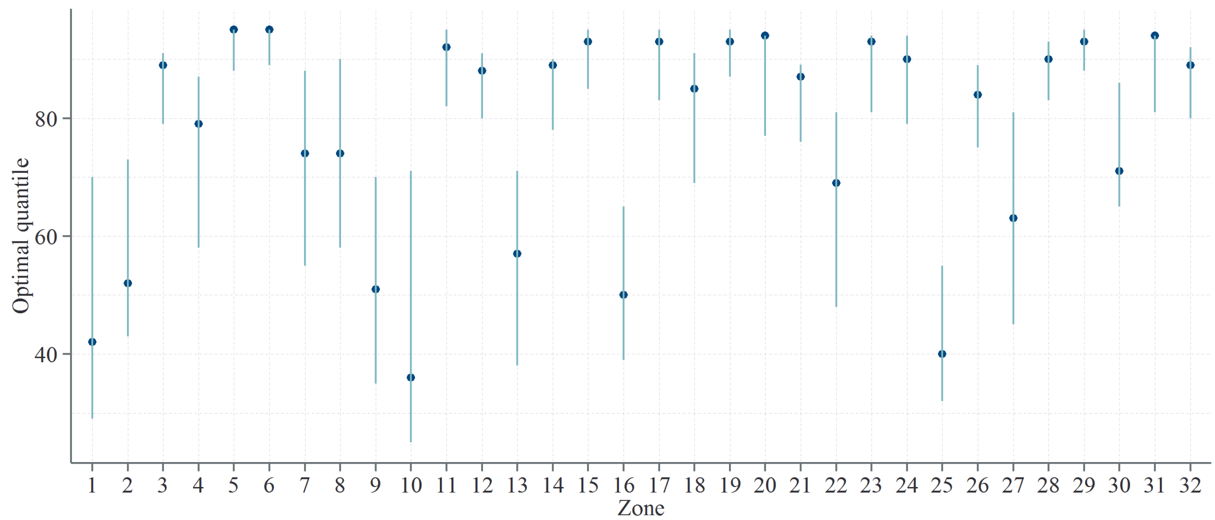

Figure 5: 95% confidence intervals for optimal quantiles by zone

way as the median of a variable relates to its mean. Hence, the further the estimated optimal quantile is from the 50th, the larger the difference we expect between OFQR and OLS predictions.

Because the optimal quantiles are estimated, we can create confidence intervals for these paramters like any other. Figure 5 plots 95% confidence intervals derived from the bootstrapping procedure. There is a great deal of spread in these values, ranging from the 36th to the 95th. More than half, however, are above the 85th quantile, suggesting that there is generally a strong divergence between the QR and OLS approaches. These intervals can be quite wide, especially for lower values of the optimal quantiles.[19]

To understand the spread in optimal quantile values, recall that the advantage of the OFQR approach over OLS is that QR can incorporate the impact of large shock terms on annual peak demand. If "large" shocks are nonetheless small compared to the predictable component of demand, then OFQR should provide predictions similar to that of OLS. Shocks are relatively small if the R^2 of the OLS model is relatively high. Indeed, Figure 6 reveals that the optimal quantiles are largest for models with lower R^2 values (below 0.87) and are more varied for models with higher R^2 values. It appears that a necessary, though not sufficient, condition for the estimated quantile being low is for the R^2 of the OLS model to be high.

There is nothing inherently wrong with having lower quantiles used for OFQR estimation

[19] Recall that we set the upper bound for the estimated quantile to be the 95th. Hence, this is the upper bound for the optimal quantiles and serves to compress optimal quantiles at the upper end of the intervals.

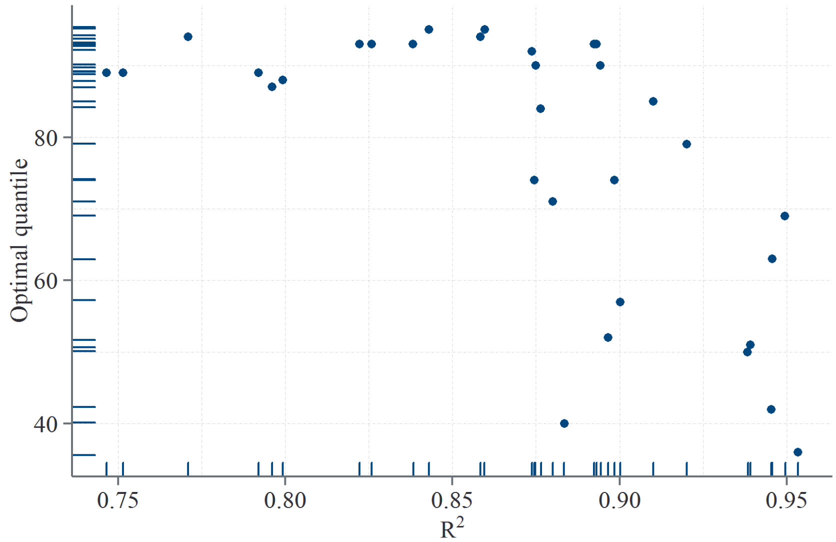

Figure 6: Comparison of OFQR optimal quantiles and OLS R^2 values

nor is it a signal of a defect in the QR procedure. The two approaches should be judged by their accuracy in predicting annual peak demand.

6.3 Performance

In this section, we provide two estimates of accuracy in predicting annual peak demand: mean percent error (MPE) and mean absolute percent error (MAPE). As with the case of the loss function in Section 5.2, the choice of accuracy measure depends upon the context of evaluation (Hyndman and Koehler, 2006). We calculate these measures for both the in-sample period from 1999 to 2008 and the out-of-sample period 2009 to 2013.

Method 3 (Evaluation approach)**.** *The models are evaluated:*

1. *In sample using ten years of data from the estimation period, 1999 through 2008*

2. *Out-of-sample using five years of data from 2009 through 2013 using the observed values of the predictors.*

Forecasts in the out-of-sample period are called "oracle forecasts" *because they are based upon the true values of the predictors.*

For out-of-sample evaluation, we generate forecasts using the true values of the predictors for two reasons. First, in assessing the performance of forecasts of the annual peak based upon forecasts

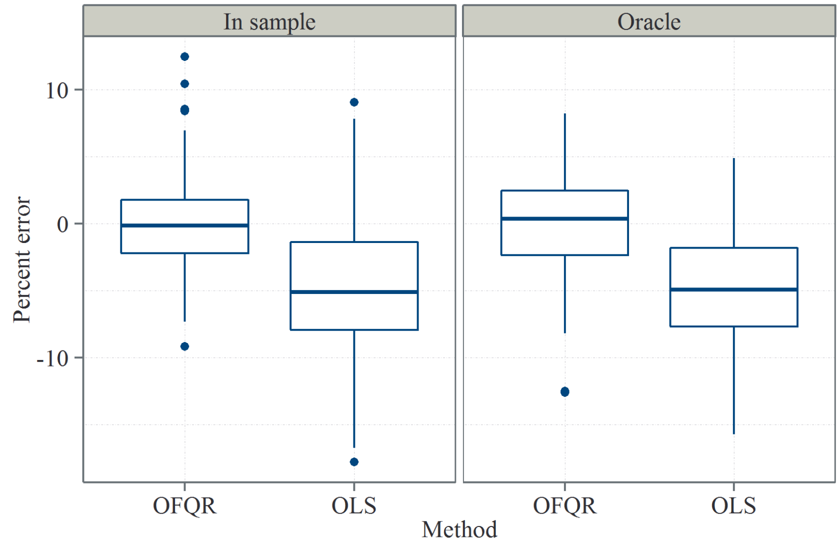

Figure 7: Distribution of percent prediction errors for each zone-year

of the predictors, it is difficult to disentangle inaccuracies arising from a poor peak forecasting procedure and those arising from poor forecasts of the predictors. Second, providing accurate methods for forecasting the predictors is beyond the scope of this paper.

Figure 7 shows the distribution of percent forecast errors for each year for each zone.[20] Because we use a mean percent error loss function for estimation, we expect the percent error for the OFQR model to be centered near 0 for the in-sample period. Because the optimal quantile is estimated using only variation in annual peak demand, we might be concerned that our approach overfits to the ten annual observations in the estimation window at the expense of out-of-sample performance. Our results indicate that this is not a concern. The distribution of QR percent errors is centered near 0 out of sample, with a mean value of -0.05 percent.

As expected from intuition developed in Section 4.1, OLS is downwardly biased, with an mean percent error of -4.74 percent in sample and -4.82 percent out of sample. As with the QR models, the similarity of the in-sample and oracle results here also dismisses the concern of overfitting, an unsurprising conclusion given that we did not perform any model selection procedure.

In addition to being unbiased, the OFQR approach also has errors that are smaller in magnitude than OLS. Figure 8 shows the absolute percent errors by zone-year. Once again, the

[20]That is, a zone with data from 1999 through 2013 contributes ten points to the "In sample" panel (annually from 1999–2008) and five to the "Oracle" panel (annually from 2009–2013).

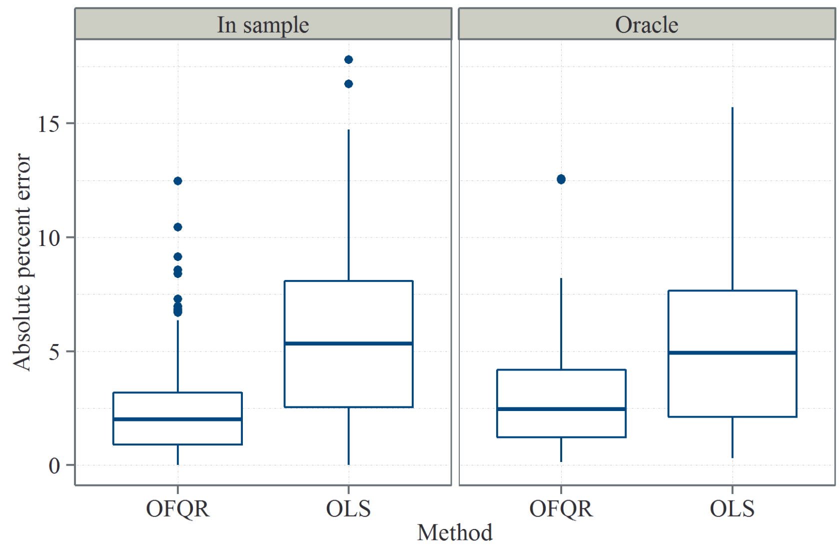

Figure 8: Distribution of absolute percent prediction errors for each zone-year

OFQR model has better performance than the OLS model in- and out-of-sample. The mean absolute prediction error for the OFQR approach is 3.05 percent out of sample as compared to 5.28 percent for OLS, a reduction of 42%.

These results generally hold across zones. Figure 9 shows the oracle forecast MPE by zone.[21] The MPE of the oracle QR approach is negative for 15 of 32 zones; about half underforecast on average and half overforecast on average out of sample. In comparison, 30 zones have OLS oracle MPEs below 0, confirming the general downward bias of the OLS approach. Of the 32 zones, 22 have a QR MPE that is smaller in magnitude than the OLS MPE out of sample. Similar results hold for the oracle MAPE, as seen in Figure 10. There are 24 zones with a QR MAPE that is smaller than the OLS MAPE out of sample; the median reduction in MAPE is 25%.

To generate 95% forecast intervals, we perform 1,000 bootstrap replications for each zone. We use a block bootstrap with each block being a calendar week. Our in-sample coverage is nearly exact, with a coverage rate of 95.1. The out-of-sample performance is not as accurate, with a coverage rate of 86.8.

[21] The in-sample MPE is relatively uninteresting, as our estimation procedure ensures that the MPE for each zone is 0 in the OFQR model.

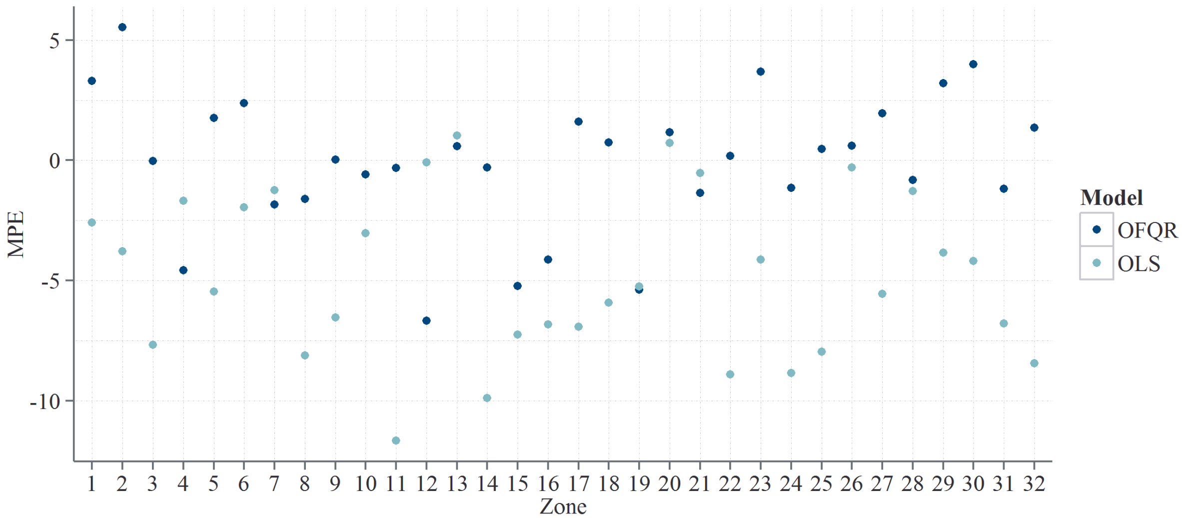

Figure 9: MPE of the oracle forecast by zone

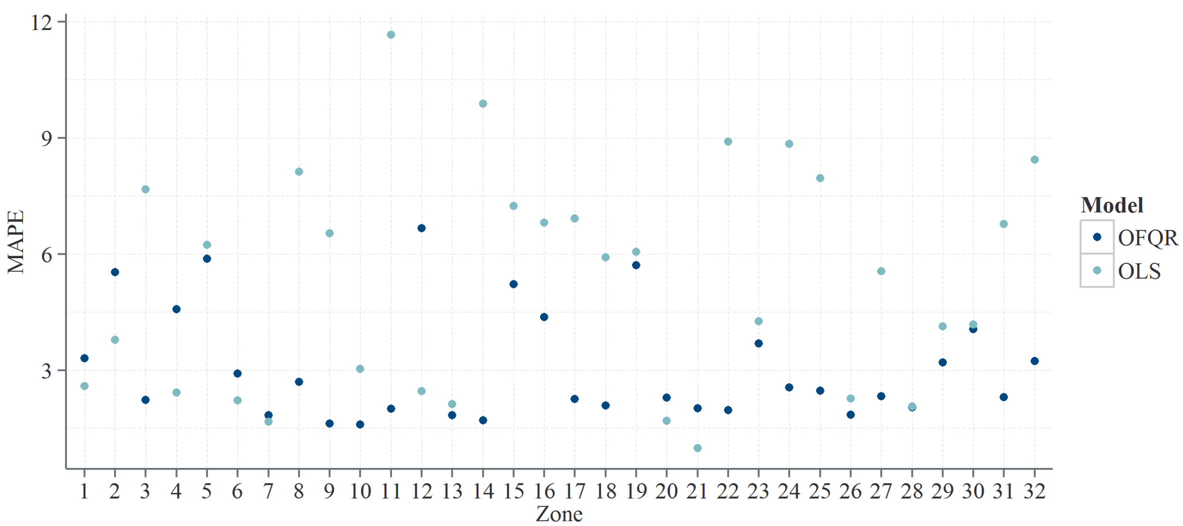

Figure 10: MAPE of the oracle forecast by zone

7 Implications and conclusions

Peak demand forecasts guide utility resource planning over medium- to long-term horizons. Inaccuracies in these forecasts can lead to substantial economic costs. If peak demand is over-forecasted, more capacity would be built than is required, needlessly raising electric rates for all customers in order to recuperate the excess costs. And if demand is under-forecasted, too little capacity would be built, leading to brownouts and blackouts with their attendant economic costs. An accurate forecasting approach is necessary to avoid inefficient investments and insufficient load.

The modeling problem has become more complicated since the onset of the Great Recession in 2008. Even though the recession lasted only 18 months, the economic recovery has been weak. Sales growth has been negative in some years and positive in others, exacerbating uncertainty in peak demand models that include sales as a predictor (Faruqui and Shultz, 2012). Accuracy is at a premium when it comes to demand forecasting and it has become more challenging to achieve in the post-recessionary environment.

The standard approach throughout the utility industry to forecast peak demand has been OLS. In our paper, we introduce a new approach, optimal forecast quantile regression (OFQR), and evaluate its performance relative to OLS using data from 32 zones, where a zone could be a utility, an entire ISO/RTO, or a region within an ISO/RTO. These 32 zones cover eight of the nine U.S. Census divisions and span a range of load and sales levels.

A common model specification is estimated for all zones. Daily peak demand is expressed as a function of monthly energy sales, daily and monthly weather measurements, an economic activity index, and several sets of fixed effects for calendar variables. The models are estimated using daily data from the summers of 1999 through 2008. Their performance is evaluated using the in-sample period and using an out-of-sample period that runs from 2009 to 2013. In the out-of-sample period, we use actual values of the predictors in our evaluation to focus on the accuracy of the forecasting approach and avoid confounding from inaccurate forecasts of the predictors. We find that OFQR dominates OLS in 24 of the 32 zones, with a median reduction in MAPE of 25% in the out-of-sample period.

We describe theoretically and demonstrate empirically that annual peak days are distinguished from non-peak days due to extreme values of the predictors in the model and due to the presence of large, unobservable shocks to daily peak demand. Because OLS accounts for the former, though not the latter feature of peak demand, it tends to have a downward bias. In contrast, OFQR takes both of these features into account in an optimal way and offers an accurate forecast. This difference explains the divergence in the empirical performance of the two approaches.

In producing their forecasts, some modelers have attempted to account for the uniqueness of peak demand days by limiting their sample to days that are peak-like due to extreme weather conditions or by performing other ad hoc adjustments. Thus, considerable judgment can be introduced into the peak demand forecasting process when the model is based on OLS. OFQR removes this subjectivity and produces an accurate forecast automatically using the complete data set.

We suggest that utilities and regulatory bodies consider re-estimating their models using OFQR and compare its performance to that of their current approaches. It is important to recognize that no additional data is required to carry out the comparison. If OFQR gives better results, they might consider adopting this approach as their forecasting algorithm.

References

Bremnes, John Bjrnar. 2004. "Probabilistic wind power forecasts using local quantile regression." *Wind Energy* 7(1):47–54.

Bunn, Derek, Arne Andresen, Dipeng Chen and Sjur Westgaard. 2013. Analysis and Forecasting of Electricity Price Risks with Quantile Factor Models. Working paper.

Cabrera, Brenda López and Franziska Schulz. 2014. Forecasting Generalized Quantiles of Electricity Demand: A Functional Data Approach. SFB 649 Discussion Paper Series 2014-030 Humboldt-Universität zu Berlin.

Davison, A. C. and D. V. Hinkley. 1997. *Bootstrap Methods and their Application.* Cambridge; New York, NY, USA: Cambridge University Press.

Faruqui, Ahmad and Eric Shultz. 2012. "Demand Growth and the New Normal." 2012(12):23–28.

Hendricks, Wallace and Roger Koenker. 1992. "Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity." *Journal of the American Statistical Association* 87(417):58–68.

Hyndman, R.J. and Shu Fan. 2010. "Density Forecasting for Long-Term Peak Electricity Demand." *Power Systems, IEEE Transactions on* 25(2):1142–1153.

Hyndman, Rob J. and Anne B. Koehler. 2006. "Another look at measures of forecast accuracy." *International Journal of Forecasting* 22(4):679–688.

ISO New England. 2014. Forecast Model Structures of the ISO New England Long-Run Energy and Seasonal Peak Load Forecasts. Technical report ISO New England.

Jebaraj, S. and S. Iniyan. 2006. "A review of energy models." *Renewable and Sustainable Energy Reviews* 10(4):281–311.

Jónsson, Tryggvi, Pierre Pinson, Henrik Madsen and Henrik Aa. Nielsen. 2013. Predictive Densities for Day-Ahead Electricity Prices Using Time-Adaptive Quantile Regression. Working paper.

Kaza, Nikhil. 2010. "Understanding the spectrum of residential energy consumption: A quantile regression approach." *Energy Policy* 38(11):6574–6585.

Koenker, Roger and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica* 46(1):33–50.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *The Journal of Economic Perspectives* 15(4):143–156.

Menne, Matthew J., Imke Durre, Russell S. Vose, Byron E. Gleason and Tamara G. Houston. N.d. "An Overview of the Global Historical Climatology Network-Daily Database." *Journal of Applied Meteorology and Climatology.* Forthcoming.

Resource Adequacy Planning Department. 2014. PJM Load Forecast Report. Annual load forecast report PJM Interconnection.

Soyiri, Ireneous N., Daniel D. Reidpath and Christophe Sarran. 2013. "Forecasting peak asthma admissions in London: an application of quantile regression models." *International Journal of Biometeorology* 57(4):569–578.

Suganthi, L. and Anand A. Samuel. 2012. "Energy models for demand forecasting–A review." *Renewable and Sustainable Energy Reviews* 16(2):1223–1240.

Worthington, Andrew C. and Helen Higgs. 2013. Forecasting the impact of generation mix on wholesale electricity prices in Australia. Discussion Paper Series 2013-06 Griffith Business School.