ARTICLE IN PRESS
International Journal of Forecasting ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

International Journal of Forecasting

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

A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting

Stephen Haben [a,*], Georgios Giasemidis [b]

[a] Mathematical Institute, University of Oxford, OX2 6GG, UK
[b] CountingLab Ltd, Reading, RG2 8EF, UK

ARTICLE INFO

Keywords:
Probabilistic load forecast
Kernel density estimation
Quantile regression
GEFCom2014

ABSTRACT

We present a model for generating probabilistic forecasts that combines the kernel density estimation (KDE) and quantile regression techniques, as part of the probabilistic load forecasting track of the Global Energy Forecasting Competition 2014. Initially, the KDE method is implemented with a time-decay parameter, but we later improve this method by conditioning on the temperature or period of the week variables in order to provide more accurate forecasts. Secondly, we develop a simple but effective quantile regression forecast. The novel aspects of our methodology are two-fold. First, we introduce symmetry into the time-decay parameter of the kernel density estimation based forecast. Second, we combine three probabilistic forecasts with different weights for different periods of the month.

1. Introduction

In this paper, we present the methodology we used in our winning entry for the probabilistic load forecasting track of the Global Energy Forecasting Competition 2014 (GEFCom2014). The competition consisted of twelve weekly tasks which required historical data to be used to estimate 99 quantiles (0.01, 0.02, . . . , 0.99) for each hour of the following month. Each forecast is evaluated using the pinball function. For further details on the competition structure and the data, the interested reader is referred to the GEFCom2014 introduction paper (Hong et al., this issue). In Section 2, we present a preliminary analysis of the data that motivates the development of the main forecasting methods introduced in Section 3. In Section 4, we provide a short description of our submissions in chronological order, to explain the reasoning behind the forecast selection and the developments of the subsequent forecasts. We present an overall view of the results and conclude in Section 5 with a discussion, lessons learned and future work.

2. Preliminary analysis

We start by performing a preliminary analysis in order to determine our initial forecasting methods. We first tested the competition's initial historical data set to confirm that load and temperature are strongly correlated, as has been shown in other studies (Charlton & Singleton, 2014); see also the GEFCom2014 introduction paper (Hong et al., this issue) for the time series plots of the data. This motivates the development of our kernel density estimation method conditional on the temperature (see Section 3.3). We also found that all of the weather stations were strongly correlated with each other and with the load data. Hence, as an initial estimate of the temperature, we simply took an average over all 25 stations.

The load data have strong daily, weekly and yearly seasonalities, as well as trends (Hong et al., this issue). A visual analysis of the load data showed that certain

* Corresponding author.
E-mail addresses: Stephen.Haben@maths.ox.ac.uk (S. Haben), G.Giasemidis@reading.ac.uk (G. Giasemidis).

http://dx.doi.org/10.1016/j.ijforecast.2015.11.004

Please cite this article in press as: Haben, S., & Giasemidis, G., A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2015.11.004

hours of the day exhibited strong bi-annual seasonalities (such as 11 pm), whereas others did not (e.g., 3 pm). This could be due to the use of heating and cooling appliances through the seasons. This inspires our choice of a biannual model for the quantile regression based forecast (see Section 3.4). A consideration of the autocorrelation and partial autocorrelation plots confirmed the presence of the weekly and daily periodicities. Our forecasts described in the following section are produced with this periodicity in mind.

3. Methodology

In this section, we present the main methods implemented for the competitive tasks of the competition.

3.1. Kernel density estimation (KDE)

Many of the methods we employ are non-parametric kernel density based estimates, similar to those presented by Jeon and Taylor (2012) for probabilistic wind forecasting and (Arora & Taylor, in press) for household-level probabilistic load forecasting. This method is motivated by the strong weekly correlations in the data. A simple kernel density estimate produces an estimate of the probability distribution function $f(X)$ of the load X (at a particular future time period) using past hourly observations $\{X_i\}$ (assuming that $i = 1$ is the beginning of historical load data: 1st Jan 2005). It is given by

$$f(X) = \frac{1}{nh_x} \sum_{i=1}^{n} K\left(\frac{X - X_i}{h_x}\right), \quad (1)$$

where h_x is the load bandwidth. We use a Gaussian kernel function, $K(\bullet)$, for all of our kernel-based forecasting methods. Our first method is a KDE with a time decay parameter, $0 < \lambda \leq 1$. The role of the decay parameter is to give a higher weight to more recent observations. To forecast day D of the week, $D = 1, 2, \ldots, 7$, at hour h, $h = 1, 2, \ldots, 24$, we applied a KDE to all historical observations of the same day D and hour h. This method considers only observations belonging to the same hourly period of the week, denoted w, $w = 1, \ldots, 168$, and we refer to it as *KDE-W*. This can be expressed as

$$f(X) = \frac{1}{nh_x} \sum_{\substack{i=1 \\ \{i \bmod s = w\}}}^{n} \frac{\lambda^{\alpha(i)}}{\sum\limits_{\substack{i=1 \\ \{i \bmod s = w\}}}^{n} \lambda^{\alpha(i)}} K\left(\frac{X - X_i}{h_x}\right). \quad (2)$$

The parameter $s = 168$ is the number of forecasting hours in a week, and $\alpha(i)$ is a periodic function given by[1]

$$\alpha(i) = \min\left(|\mathcal{D} - (\mathcal{D}(i) - \mathbf{1}_A(i))|, \mathcal{T}(i) - |\mathcal{D} - \mathcal{D}(i)|\right), \quad (3)$$

where $\mathcal{D}(i) = 1, 2, \ldots, \mathcal{T}(i)$ is the day of the year (consisting of $\mathcal{T}(i)$ days) corresponding to the historical data X_i, and $\mathcal{D}$ is the day of the year corresponding to the forecasted day. To correct for leap years, we use an indicator function $\mathbf{1}_A(i)$, where $A = \{i | \mathcal{D}(i) > 28 \text{ and } \mathcal{T}(i) = 366\}$. Eq. (3) is simply a periodic absolute value function with an annual period, the minimum values of which occur annually on the same date as the forecasted day.

This method is similar to that presented by Arora and Taylor (in press), with the new feature of the half-yearly symmetry of the time-decay exponential in Eq. (3). Since there is an annual periodicity in the load, we incorporated it into the time-decay parameter such that observations during similar days of the year influence the forecast more than other, less relevant observations. The decay parameter also helps us to take into account the non-stationary behaviour of demand. This method performed better than a similar KDE-W using only a simple monotonically decreasing time-decay parameter across the year. The model parameters were generated using cross-validation on the month prior to the forecasting month. To find the optimal bandwidth, h_x, we used the *fminbnd* function from the optimisation toolbox in Matlab. For the time-decay parameter λ, we considered different values between 0.92 and 1, with 0.01 increments.[2]

The kernel density based estimate has been used as a benchmark in probabilistic forecast methods applied to household level electricity demand, and serves as a useful starting point for our forecasts (Arora & Taylor, in press). The method has the advantage of being quicker to implement than more complicated kernel-based methods, such as the conditional kernel density estimate on independent parameters, which we introduce in the following sections.

3.2. Conditional kernel density estimate on period of week (CKD-W)

A KDE forecast conditional on the period of the week, denoted by w, $w = 1, \ldots, 168$, (CKD-W; see Arora & Taylor, in press) gives a higher weight to observations from similar hourly periods of the week, and can be represented as

$$f(X|w) = \sum_{i=1}^{n} \frac{\lambda^{\alpha(i)} K((w_i - w)/h_w)}{\sum\limits_{i=1}^{n} \lambda^{\alpha(i)} K((w_i - w)/h_w)} K\left(\frac{X - X_i}{h_x}\right) \quad (4)$$

where $\alpha(i)$ is defined in Eq. (3).

This method is similar to that presented by Arora and Taylor (in press), but with the new feature of the half-yearly symmetric time-decay exponential in Eq. (3), which is justified by the yearly periodicity of the load, as was explained in the previous section.

The validation process can be very expensive computationally, especially when searching for multiple optimised parameters (here, there are three parameters: the bandwidths for load and week period variables, and the time

[1] The careful reader should note that Eq. (3) might need a further correction by one when D is in a leap year. However, this does not affect our results, since we did not forecast leap years. In addition, the effect of such an error on the weight would be negligible.

[2] The time-decay parameter must be in the interval (0, 1], where the smaller the value the fewer historical observations that have a significant influence on the final forecast. After testing over several tasks, we found that the decay parameter is bounded below by 0.92.

Please cite this article in press as: Haben, S., & Giasemidis, G., A hybrid model of kernel density estimation and quantile regression for GEFCom2014 probabilistic load forecasting. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2015.11.004