A Robust Segmented Mixed Effect Regression Model for Baseline Electricity Consumption Forecasting

Xiaoyang Zhou, Yuanqi Gao, Weixin Yao, and Nanpeng Yu

Abstract—**Renewable energy production has been surging around the world in recent years. To mitigate the increasing uncertainty and intermittency of the renewable generation, proactive demand response algorithms and programs are proposed and developed to further improve the utilization of load flexibility and increase the efficiency of power system operation. One of the biggest challenges to efficient control and operation of demand response resources is how to forecast the baseline electricity consumption and estimate the load impact from demand response resources accurately. In this paper, we propose a mixed effect segmented regression model and a new robust estimate for forecasting the baseline electricity consumption in Southern California, USA, by combining the ideas of random effect regression model, segmented regression model, and the least trimmed squares estimate. Since the log-likelihood of the considered model is not differentiable at breakpoints, we propose a new backfitting algorithm to estimate the unknown parameters. The estimation performance of the new estimation procedure has been demonstrated with both simulation studies and the real data application for the electric load baseline forecasting in Southern California.**

Index Terms—**Segmented regression model, mixed effects, trimmed maximum likelihood, demand response, electric load.**

I. Introduction

THE renewable energy sector has experienced exponential growth in the past five to ten years. The global annual growth rates of solar photovoltaic and wind energy are 42% and 17% from 2010 through 2015, respectively [1]. The renewable penetration level in certain parts of the world is much higher than the global average penetration level. For example, the renewable energy penetration level in California reached 30% in 2017. The recently passed California Senate Bill No. 100 will further boost renewable penetration level up to 60% by 2030 and 100% by 2045. To mitigate the increasing uncertainty and intermittency of renewable generation, demand response resources are in critical need. In the past ten years, traditional and passive price-based and incentive-based demand response programs have been implemented throughout the USA. In recent years, proactive demand response algorithms and programs are proposed and developed to improve the utilization of load flexibility and dispatchability further [2]. Accurate load impact forecasting is needed to leverage the load flexibility from the demand response resources effectively. The load impact from a demand response resource is defined as the difference between load baselines and metered load when a demand response event is triggered. In practice, it is very challenging to develop a good estimation of the load baseline which represents the electric load that would have occurred without demand response event [3].

A sound baseline estimation methodology should represent an appropriate tradeoff between simplicity and accuracy. The existing baseline methodology can be categorized into two types: Type-I and Type-II. In Type-I methodology, the baseline is estimated by using a similar day-based algorithm, which depends on historical interval meter data and similarity metrics such as weather and calendar. Simplicity is the most significant advantage of Type-I methodology [4], [5]. In Type-II methodology, more sophisticated statistical methods are adopted to estimate and forecast the baseline electricity consumption. Type-II methodology typically yields better forecasting accuracy and is undergoing rapid development. It can be further divided into three groups: statistical methods, machine learning/deep learning methods, and hybrid methods. In the first group, [6] proposes a refined multiple linear regression model. Reference [7] proposes a method to coherently convert a set of lower-level node forecasting to aggregate nodes using empirical copula and Monte-Carlo sampling. In the same vein, [8] proposes an aggregation of random forest load forecasting framework. The second group utilizes deep learning algorithms. Reference [9] proposes support vector regressions models to forecast the demand response baseline. In [10], an ensemble ResNet deep neural network model is proposed. The sequence-to-sequence recurrent neural network with attention mechanism is adopted in [11]. In the third group, hybrid methods have been devel-

Manuscript received: January 14, 2020; revised: July 4, 2020; accepted: August 17, 2020. Date of CrossCheck: August 17, 2020. Date of online publication: November 4, 2020.

The research of W. Yao was supported by National Science Foundation (No. DMS-1461677) and Department of Energy (No. DE-EE0007328). The research of N. Yu was supported by National Science Foundation (No. 1637258) and Department of Energy (No. DE-EE0007328).

X. Zhou and W. Yao are with the Department of Statistics, University of California, Riverside, CA, USA (e-mail: xzhou010@ucr.edu; weixin.yao@ucr.edu).

Y. Gao and N. Yu (corresponding author) are with the Department of Electrical and Computer Engineering, University of California, Riverside, USA (e-mail: ygao024@ucr.edu; nyu@ece.ucr.edu).

DOI: 10.35833/MPCE.2020.000023

oped, in which more than one forecasting algorithms serve as building blocks for an overall model. Reference [12] proposes a cooperative quantile regression forest and multivariate quantile regression framework. Reference [13] proposes a two-level hybrid ensemble of deep belief network model.

There are several limitations with the existing approaches. First, some of the methods do not exploit the structure of the forecasting problem effectively. For example, the segmented nature of calendar variables on the load profile is not well addressed. Second, deep learning based forecasting algorithms are typically computationally expensive to train. In addition, they yield un-interpretable results and can be sensitive to the selection of hyperparameters. Third, hybrid methods are generally complicated to build, thus can be error-prone to implement and benchmark. Lastly, most of the existing work build and train a separate model for each time series. This significantly limits the scalability of model, especially for large service territories operated by electric utilities.

In this paper, we propose a mixed effect segmented regression (MESR) model, which is a Type-II methodology, to forecast the hourly electric load baseline in Southern California, USA at the 220 kV transformer bank level. One commonly used method for electric power demand forecasting at each hour is the multiple linear regression with hour as a categorical variable and weather data as continuous covariates. An alternative model for hour is to include it as a linear predictor. However, it is expected that the linear effect of hour on electric demand does not hold in the whole range of time. To this end, we propose to model the hour effect by a segmented regression model [14]-[17], which can be considered as a compromise between modelling the hour as a global linear predictor and a categorical variable. The nonlinear relationship with breakpoints is piece-wise, segmented, broken-line, or multi-phased. The breakpoints are also called change-points, transition-points or switch-points in some applications. Using a segmented regression model for the covariate hour, the effect on the electric consumption changes continuously across time so that we can borrow the information from other hours when estimating the impact of hour. The estimated breakpoints can also tell us how the linear effect of hour changes across different segmented areas. Segmented regression models have been widely used in many areas. In medication, segmented regression is a powerful statistical tool for estimating the intervention effects of interrupted time series studies [18]. The segmented regression is also used to identify the changes in the recent trend of cancer mortality and incidence data analysis [19]. In ecology area, the segmented regression is a widely used statistical tool to model ecological thresholds [20]. For the geometric purpose, the segmented regression statistically models the trends at groundwater levels [21]. Many other examples with piecewise linear terms have been studied in the literature including mortality studies [22], Stanford heart transplant data [23], and mouse leukemia [24]. Note that electric consumption data are essentially longitudinal/panel data. They exhibits very strong spatio-temporal dependencies [25]. To incorporate the correlation among the observations and the individual-specific heterogeneity from each transformer bank, we propose to use the random effect regression model [26], [27].

Note that it is not trivial to compute the maximum likelihood estimate (MLE) for the MESR, since its log-likelihood is not differentiable at breakpoints. Many standard computation algorithms such as the Newton-Raphson algorithm can not be used directly. In this paper, we propose a backfitting algorithm to combine the segmented regression estimation method proposed in [28] and the mixed effect regression estimation method proposed in [29] to maximize the non-differentiable log-likelihood of the mixed effect segmented regression model. Note that the MLE is sensitive to outliers, which is the case of our electric consumption data collected in the Southern California. We further propose a robust estimation procedure for the considered model by extending the idea of the least trimmed squares (LTS) estimate [30]. Simulation studies demonstrate the effectiveness of the proposed estimation procedures. The LTS also provides much better forecasting performance than the standard MLE for the testing data when forecasting the hourly electric power consumption in Southern California.

The rest of the paper is organized as follows. Section II introduces the MESR and describes the proposed robust estimation algorithms. Section III illustrates the finite sample performance of the proposed method using a simulation study. In Section IV, we apply the new estimation procedure to forecast the hourly electric power demand in Southern California, USA. Section V concludes the paper with some discussions.

II. MESR and Proposed Robust Estimation Algorithm

Given a random sample $\{y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{s}_{ij}, z_{ij}\}, i=1,2,\ldots,n, j=1,2,\ldots,n_i$, where n is the number of subjects; n_i is the number of observations collected for the i^{th} subject; y_{ij} is the response variable; $\boldsymbol{x}_{ij}$ is the p-dimensional fixed-effect covariate; $\boldsymbol{s}_{ij}$ is the q-dimensional random-effect covariate; and z_{ij} is the breakpoint variable with the breakpoints $\{\varphi_k\}, k=1,2,\ldots,l$. The proposed MESR model for the load baseline estimation can be written as:

$$y_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\phi} + \boldsymbol{s}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i + \beta_0 z_{ij} + \sum_{k=1}^{l}\beta_k (z_{ij}-\varphi_k)_+ + \varepsilon_{ij} \tag{1}$$

where $\boldsymbol{\phi}$ is the regression coefficient for the random effect covariates; β_0 and β_k are the regression coefficients for the breakpoint variables; and the quantities with a subscript "+" means taking the positive part. For example, t_+ equals t if $t \ge 0$ and 0 otherwise; $\boldsymbol{\gamma}_i \sim N_q(0, \boldsymbol{\Sigma}_\gamma)$; $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i}) \sim N_{n_i}(0, \boldsymbol{\Sigma}_\varepsilon)$. In this paper, we assume that $\boldsymbol{\Sigma}_\varepsilon = \sigma^2 \boldsymbol{I}_{n_i}$, where σ is the standard deviation of the error variable; and $\boldsymbol{I}_{n_i}$ is the $n_i \times n_i$ identity matrix. The MESR (1) consists of three parts: multiple linear regression $\boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\phi}$, random effects $\boldsymbol{s}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i$, and segmented regression $\beta_0 z_{ij} + \sum_{k=1}^{l}\beta_k (z_{ij}-\varphi_k)_+$, which models the heterogeneous linear effect of z_{ij} on y_{ij} across different areas of the breakpoint variable. β_k measures the difference of slopes (linear effect of z_{ij} on y_{ij}) before and after the breakpoint φ_k. We mainly focus on the situation where the segmented parts are fixed effects. But the proposed estimation procedure can be extended to the situation where the segmented parts also contain random effects [31]-[33].

Suppose that $\boldsymbol{y}_i=(y_{i1},y_{i2},\ldots,y_{in_i})^{\mathrm{T}}$, $\boldsymbol{X}_i=(\boldsymbol{x}_{i1},\boldsymbol{x}_{i2},\ldots,\boldsymbol{x}_{in_i})^{\mathrm{T}}$, $\boldsymbol{S}_i=(\boldsymbol{s}_{i1},\boldsymbol{s}_{i2},\ldots,\boldsymbol{s}_{in_i})^{\mathrm{T}}$, and $\boldsymbol{Z}_i=(\boldsymbol{z}^*_{i1},\boldsymbol{z}^*_{i2},\ldots,\boldsymbol{z}^*_{in_i})^{\mathrm{T}}$, where $\boldsymbol{z}^*_{ij}=(z_{ij},(z_{ij}-\varphi_1)_+,\ldots,(z_{ij}-\varphi_l)_+)^{\mathrm{T}}$. Then, (1) can be rewritten in matrix format as:

$$\boldsymbol{y}_i=\boldsymbol{X}_i\boldsymbol{\phi}+\boldsymbol{S}_i\boldsymbol{\gamma}_i+\boldsymbol{Z}_i\boldsymbol{\beta}+\boldsymbol{\varepsilon}_i \tag{2}$$

where $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_l)^{\mathrm{T}}$. Based on (2), $E(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{Z}_i,\boldsymbol{S}_i)=\boldsymbol{X}_i\boldsymbol{\phi}+\boldsymbol{Z}_i\boldsymbol{\beta}$ and $var(\boldsymbol{y}_i|\boldsymbol{X}_i,\boldsymbol{Z}_i,\boldsymbol{S}_i)=\boldsymbol{S}_i\boldsymbol{\Sigma}_\gamma\boldsymbol{S}_i^{\mathrm{T}}+\sigma_\varepsilon^2\boldsymbol{I}_{n_i}\triangleq\boldsymbol{\Sigma}_i$, where $E(\cdot)$ and $var(\cdot)$ denote conditional expectation and conditional variances, respectively. Therefore, the random effects $\boldsymbol{\gamma}_i$ make the observations within each correlated subject. The log-likelihood function of $\{y_{ij},\boldsymbol{x}_{ij},\boldsymbol{s}_{ij},z_{ij}\}, i=1,2,\ldots,n,\ j=1,2,\ldots,n_i$ is:

$$l(\boldsymbol{\theta})\propto\sum_{i=1}^{n}\ln(|\boldsymbol{\Sigma}_i|^{-1/2})-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\phi}-\boldsymbol{Z}_i\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\phi}-\boldsymbol{Z}_i\boldsymbol{\beta}) \tag{3}$$

where $\boldsymbol{\theta}$ collects all the unknown parameters $\{\boldsymbol{\phi},\boldsymbol{\beta},\boldsymbol{\varphi},\sigma,\boldsymbol{\Sigma}_\gamma\}$ in model (1). Unlike the traditional mixed effect model, maximizing (3) is not trivial since it is not differentiable at φ_k. We propose a backfitting algorithm to maximize (3) by alternately updating the segmented regression part and the linear mixed effect part. Next, we discuss in detail how to perform such two estimation procedures.

A. Estimation of Breakpoints

Given the estimate $\{\boldsymbol{\phi},\hat{\boldsymbol{\Sigma}}_i\}$, (1) will be a segmented regression model. The breakpoints and slopes in segmented regression can be estimated through many ways such as regression spline as well as Bayesian Markov chain Monte Carlo (MCMC) methods [34], [35]. We will extend the linearization technique proposed in [28] to MESR (1) due to its simplicity of computation. According to the definition of breakpoints, the log-likelihood is not differentiable at φ_k. The breakpoint estimation can be performed via a first-order Taylor expansion of $(z_{ij}-\varphi_k)_+$ around an initial value $\varphi_k^{(0)}$:

$$(z_{ij}-\varphi_k)_+\approx(z_{ij}-\varphi_k^{(0)})_++(\varphi_k-\varphi_k^{(0)})(-1)I(z_{ij}>\varphi_k^{(0)}) \tag{4}$$

where $I(\cdot)$ is the indicator function. It equals 1 if the condition inside the parenthesis is true and 0 otherwise; $(-1)I(z_{ij}>\varphi_k^{(0)})$ is the first derivative of $(z_{ij}-\varphi_k)_+$ assessed in $\varphi_k^{(0)}$.

Let $\boldsymbol{v}_{ij}=((-1)I(z_{ij}>\varphi_1^{(0)}),(-1)I(z_{ij}>\varphi_2^{(0)}),\ldots,(-1)I(z_{ij}>\varphi_l^{(0)}))^{\mathrm{T}}$, $\tilde{\boldsymbol{z}}_{ij}=(z_{ij},(z_{ij}-\varphi_1^{(0)})_+,(z_{ij}-\varphi_2^{(0)})_+,\ldots,(z_{ij}-\varphi_l^{(0)})_+)^{\mathrm{T}}$, and $\delta_k=\beta_k(\varphi_k-\varphi_k^{(0)})$. Define $\boldsymbol{V}_i=(\boldsymbol{v}_{i1},\boldsymbol{v}_{i2},\ldots,\boldsymbol{v}_{in_i})^{\mathrm{T}}$, $\boldsymbol{\delta}=(\delta_1,\delta_2,\ldots,\delta_l)^{\mathrm{T}}$, and $\tilde{\boldsymbol{Z}}_i=(\tilde{\boldsymbol{z}}_{i1},\tilde{\boldsymbol{z}}_{i2},\ldots,\tilde{\boldsymbol{z}}_{in_i})^{\mathrm{T}}$. Given the estimate $\{\boldsymbol{\phi},\hat{\boldsymbol{\Sigma}}_i\}$, the log-likelihood (3) can be simplified as:

$$l_1(\boldsymbol{\beta},\ \boldsymbol{\delta})\propto-\frac{1}{2}\sum_{i=1}^{n}(\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{Z}}_i\boldsymbol{\beta}-\boldsymbol{V}_i\boldsymbol{\delta})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_i^{-1}(\tilde{\boldsymbol{y}}_i-\tilde{\boldsymbol{Z}}_i\boldsymbol{\beta}-\boldsymbol{V}_i\boldsymbol{\delta}) \tag{5}$$

where $\tilde{\boldsymbol{y}}_i=\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\phi}$. Therefore, $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ in (5) can be easily found by weighted least squares estimate. Note that $\varphi_k=(\delta_k/\beta_k)+\varphi_k^{(0)}$. The iterative algorithm will terminate at $\delta_k=0$. Given the estimate $\{\boldsymbol{\phi},\hat{\boldsymbol{\Sigma}}_i\}$, the algorithm to estimate the breakpoints is summarized in Algorithm 1.

Algorithm 1: segmented regression estimation

1. Set initial values of all breakpoints $\varphi_k^{(0)}$ for $k=1,2,\ldots,l$, and calculate the variable $\tilde{\boldsymbol{Z}}_i$ and the variable $\boldsymbol{V}_i$
2. Repeat
3. Fit the regression model of $\tilde{\boldsymbol{y}}_i$ on $\tilde{\boldsymbol{Z}}_i$ and $\boldsymbol{V}_i$ by maximizing the log-likelihood (5). Update the breakpoint with equation $\varphi_k^{(s+1)}=\left(\delta_k^{(s)}/\beta_k^{(s)}\right)+\varphi_k^{(s)}$, where $\varphi_k^{(s)}$ is the estimate of φ_k at the s^{th} iteration
4. Until converge

B. Estimation of Mixed Effect Regression Models

In this sub-section, we discuss how to maximize (3) given the estimate $\boldsymbol{\beta}$ and $\boldsymbol{\varphi}$, where $\boldsymbol{\varphi}=(\varphi_1,\varphi_2,\ldots,\varphi_l)^{\mathrm{T}}$. Let $\hat{\boldsymbol{Z}}_i$ be the estimate of $\boldsymbol{Z}_i$ after replacing φ_k by $\hat{\varphi}_k$. Plugging in the estimate $\{\hat{\boldsymbol{Z}}_i,\boldsymbol{\beta}\}$ into the model (1), we can obtain:

$$\boldsymbol{y}_i^*=\boldsymbol{X}_i\boldsymbol{\phi}+\boldsymbol{S}_i\boldsymbol{\gamma}_i+\boldsymbol{\varepsilon}_i \tag{6}$$

where $\boldsymbol{y}_i^*=\boldsymbol{y}_i-\hat{\boldsymbol{Z}}_i\boldsymbol{\beta}$. Therefore, the model (6) is simply a traditional mixed effect regression model. We propose to employ the penalized weighted least square (PWLS) method to estimate the unknown parameters in (6). More details of computing the linear mixed effect regression model are given in [29], which are also implemented in R package *lme4*.

C. Estimation of Mixed Effect Breakpoints

By combining the estimation procedures in Section II-A and II-B, we propose Algorithm 2 to maximize the log-likelihood (3) for the model (2).

Algorithm 2: MLE

1. Set initial values of breakpoint $\varphi_k^{(0)}$ and $\boldsymbol{\beta}^{(0)}$
2. Repeat
3. Given current breakpoint values $\varphi_k^{(s)}$ and slopes $\boldsymbol{\beta}^{(s)}$, calculate $\boldsymbol{y}_i^{*(s)}=\boldsymbol{y}_i-\hat{\boldsymbol{Z}}_i^{(s)}\boldsymbol{\beta}^{(s)}$
4. Fit mixed effect regression model by the PWLS estimation procedure introduced in Section II-B to obtain covariance matrix $\boldsymbol{\Sigma}_r^{(s)}$ and the fixed effect regression estimate $\boldsymbol{\phi}^{(s)}$
5. Calculate $\tilde{\boldsymbol{y}}_i^{(s)}=\boldsymbol{y}_i-\boldsymbol{X}_i\boldsymbol{\phi}^{(s)}$
6. Fit segmented regression model with $\tilde{\boldsymbol{y}}_i^{(s)}$ and $\boldsymbol{\Sigma}_r^{(s)}$ using Algorithm 1 and update segmented regression parameter estimate to $\boldsymbol{\varphi}^{(s+1)}$ and $\boldsymbol{\beta}^{(s+1)}$
7. Until converge

D. Estimation of Robust Segmented Mixed Effect Regression

It is well known that the MLE is sensitive to outliers and might give misleading results when there are outliers in the data, which is the case for our collected electric power demand data in Southern California. More details will be given in Section IV. The issue of outlier is well recognized in the field of load forecasting, and is typically solved using robust regression algorithms. For example, [36] considers Huber's robust regression; [37] advocates the use of L_1 regression model. In the statistics literature, many robust regression methods have been proposed, although not all of them are investigated in the load forecasting literature, e.g., M-estimates [38], R-estimates [39], least median of squares (LMS) estimates [40], LTS estimates [30], S-estimates [41], MM-estimates [42], robust and efficient weighted least squares estimator (REWLSE) [43], mean shift method [44]-[46]. Reference [47] provides a good review of some commonly used robust regression estimation methods. Next, we propose the idea of least trimmed squares estimate [30] to provide a robust estimate of the model (1). Given an integer trimming parameter $h \le N$, where N is the total number of training samples, the least trimmed squares minimize the sum of the smallest h squared residuals with objective function:

$$\sum_{k=1}^{h} r_{(k)}^2 \tag{7}$$

where $r_{(k)}^2 \in [r_{(1)}^2, r_{(N)}^2]$ are the ordered squared residuals $\{y_{ij} - \hat{y}_{ij}, i=1,2,\dots,n, j=1,2,\dots,n_i\}$ with $\hat{y}_{ij} = \boldsymbol{x}_{ij}^{\mathrm{T}}\boldsymbol{\phi} + \boldsymbol{s}_{ij}^{\mathrm{T}}\boldsymbol{\gamma}_i + \hat{\beta}_0 z_{ij} + \sum_{k=1}^{l}\hat{\beta}_k (z_{ij} - \hat{\varphi}_k)_+$. The robust MESR estimation based on LTS is described in Algorithm 3.

Algorithm 3: LTS

1. A subsample of size h^* is selected randomly from the data and then the model (1) is fitted to the subsample using Algorithm 2 of Section II-C. Let $\boldsymbol{\theta}^{(0)}$ be the initial parameter estimate
2. Repeat
3. Based on current model parameter estimate $\boldsymbol{\theta}^{(s)}$, forecast N responses $\hat{y}_{ij}^{(s)}$, and calculate the residuals $r_{ij}^{(s)} = y_{ij} - \hat{y}_{ij}^{(s)}$. Rank the squared residuals from smallest to largest and select the first h observations that correspond to the smallest h squared residuals
4. Fit the model (1) to the subsample selected in *Step 3* using Algorithm 2 and get the model parameter estimate $\boldsymbol{\theta}^{(s+1)}$
5. Until converge

To increase the chance of finding the global minimum, one might run Algorithm 3 from many random subsamples and choose the solution which has the smallest trimmed squares. Let r be the dimension of $\boldsymbol{\theta}$. The initial sample size h^* can be any small number larger than r as long as the initial parameter estimate $\boldsymbol{\theta}^{(0)}$ can be computed based on the subsample. The maximum breakpoint of LTS [48], i. e., the smallest fraction of contamination that can cause the estimator to take arbitrary large values, is 0.5, which is attained when $h = [(N+r+1)/2]$, where $[\cdot]$ means rounding to the nearest integer. If we have the prior that the proportion of outliers is no more than α, we can also set $h = [N(1-\alpha)+1]$, where α is he trimming proportion. In practice, one might try several α values to evaluate LTS and check how the estimate behaves with different trimming proportions [49]-[51]. In our real data application, we use a validation data to choose the trimming proportion.

III. Simulation Study

In this section, we use a simulation study to illustrate the performance of the proposed estimation procedure for the MESR. All the computations are implemented in R. We use R package *segmented::segmented* [52] for breakpoint estimation and *lme4::lmer* [53] for random-effect estimation. We generate observations $\{y_{ij}, \boldsymbol{x}_{ij}, \boldsymbol{s}_{ij}, z_{ij}, i=1,2,\dots,n, j=1,2,\dots,n_i\}$ from the following model:

$$y_{ij} = \phi_0 + \phi_1 x_{ij} + \gamma_{i0} + s_{ij}\gamma_{i1} + \beta_0 z_{ij} + \beta_1 (z_{ij} - \varphi_1)_+ + \beta_2 (z_{ij} - \varphi_2)_+ + \varepsilon_{ij} \tag{8}$$

where $x_{ij} \sim Pois(10)$, the Poisson distribution with rate parameter 10; $s_{ij} \sim Uniform(5,10)$, the uniform distribution with lower and upper limits 5 and 10, respectively. The breakpoint variables z_{ij} are n_i arithmetic sequence ranging in $(0,20)$, $\varepsilon_{ij} \sim N(0,0.5)$, $\begin{bmatrix}\gamma_{i0}\\ \gamma_{i1}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma_{r1}^2 & \rho\sigma_{r1}\sigma_{r2}\\ \rho\sigma_{r1}\sigma_{r2} & \sigma_{r2}^2\end{bmatrix}\right)$, with $\sigma_{r1} = \sigma_{r2} = 1$, $\rho = 0.5$, $i=1,2,\dots,n$, $j=1,2,\dots,n_i$. The other parameters in (8) are set to be: $\phi_0 = -2.5$; $\phi_1 = 1.5$; $\beta_0 = 1.5$; $\beta_1 = 1.5$; $\beta_2 = -2.5$; $\varphi_1 = 6.67$; and $\varphi_2 = 13.33$.

We consider the following four simulation scenarios:

1) $n=50$, and n_i is randomly chosen in $(90,110)$.
2) $n=50$, and n_i is randomly chosen in $(18,22)$.
3) $n=200$, and n_i is randomly chosen in $(450,550)$.
4) $n=200$, and n_i is randomly chosen in $(18,22)$.

First, we utilize (8) to simulate the dataset without outliers. The model is estimated using MLE. In Tables I-IV, we report the mean, median, and standard deviation (SD) for the estimates of fixed effect regression parameters, breakpoints, segmented regression parameters, and random effect covariance matrix, respectively, based on 500 replications.

TABLE I
Simulation Results of Fixed Effect Parameter Estimates by MLE for Situation Without Outliers

Parameter	$\phi_0=-2.5$			$\phi_1=1.5$		
	Mean	Median	SD	Mean	Median	SD
$n=50$, $n_i \sim U(90,110)$	–2.505	–2.508	0.125	1.500	1.499	0.002
$n=50$, $n_i \sim U(18,22)$	–2.500	–2.502	0.169	1.500	1.500	0.005
$n=200$, $n_i \sim U(450,550)$	–2.498	–2.497	0.064	1.500	1.500	0.001
$n=200$, $n_i \sim U(18,22)$	–2.491	–2.498	0.125	1.500	1.500	0.004

TABLE II
Simulation Results of Breakpoints Estimates by MLE for Situation Without Outliers

Parameter	$\varphi_1=6.667$			$\varphi_2=13.333$		
	Mean	Median	SD	Mean	Median	SD
$n=50$, $n_i \sim U(90,110)$	6.667	6.666	0.022	13.334	13.332	0.012
$n=50$, $n_i \sim U(18,22)$	6.667	6.664	0.053	13.334	13.333	0.034
$n=200$, $n_i \sim U(450,550)$	6.667	6.667	0.006	13.333	13.333	0.003
$n=200$, $n_i \sim U(18,22)$	6.667	6.666	0.023	13.332	13.332	0.023

From Tables I-IV, we can see that the proposed MLE algorithm performs well when the dataset does not contain any outliers. Also, when the sample size increases, the SD of each parameter estimate decreases.

Next, we simulate the dataset with outliers based on model (8). The model parameters are estimated by both Algorithm 2 and Algorithm 3. In order to check how robust each estimate is against the outliers, we randomly choose 5% of each simulated data and add 30 to the response Y (the range of Y is $(15,69)$) and 10 to the value of X (the range of X is $(0,10)$). When applying LTS, we need to choose the trimming proportion α, which has long been a difficult problem. However, LTS can provide a robust model estimate as long as the proportion of outliers is less than α but with low efficiency if α is too large. Usually, a conservative choice of α is recommended in practice. In this paper, we report the results for both $\alpha=0.1$ and $\alpha=0.2$. Note that the results of LTS will be better if $\alpha=0.05$ is used.

TABLE III
SIMULATION RESULTS OF BREAKPOINT SLOPE ESTIMATES BY MLE FOR SITUATION WITHOUT OUTLIERS

Parameter	$\beta_0 = 1.5$			$\beta_1 = 1.5$			$\beta_2 = -2.5$		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
$n=50,\ n_i \sim U(90,110)$	1.499	1.500	0.006	1.499	1.499	0.008	–2.499	–2.499	0.008
$n=50,\ n_i \sim U(18,22)$	1.499	1.500	0.013	1.502	1.502	0.020	–2.502	–2.502	0.019
$n=200,\ n_i \sim U(450,550)$	1.500	1.500	0.001	1.500	1.500	0.001	–2.500	–2.500	0.002
$n=200,\ n_i \sim U(18,22)$	1.499	1.499	0.010	1.502	1.501	0.014	–2.501	–2.502	0.014

TABLE IV
SIMULATION RESULTS OF RANDOM EFFECT ESTIMATES BY MLE FOR SITUATION WITHOUT OUTLIERS

Parameter	$\sigma_{r1} = 1$			$\sigma_{r2} = 1$			$\rho = 0.5$		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
$n=50,\ n_i \sim U(90,110)$	0.969	0.959	0.108	0.978	0.976	0.101	0.504	0.503	0.121
$n=50,\ n_i \sim U(18,22)$	0.988	0.988	0.141	0.993	0.990	0.096	0.490	0.501	0.150
$n=200,\ n_i \sim U(450,550)$	0.990	0.991	0.050	0999	0.999	0.056	0.499	0.499	0.001
$n=200,\ n_i \sim U(18,22)$	0.987	0.990	0.101	0.997	1.000	0.068	0.500	0.506	0.094

Tables V-VIII present the simulation results for the estimates of fixed effect regression parameters, breakpoints, segmented regression parameters, and random effect covariance matrix, respectively, based on 200 replications. From the tables, it is observed that the standard MLE fails to provide the reasonable estimates of fixed effect regression parameters and random effect covariance matrix when the data contain 5% outliers while LTS can provide reasonable estimates for all parameters with both $\alpha = 0.1$ and $\alpha = 0.2$.

TABLE V
SIMULATION RESULTS OF FIXED EFFECT ESTIMATES BY MLE AND LTS WITH DIFFERENT α LEVELS FOR SITUATION WITH OUTLIERS

Parameter	Method	$\phi_0 = -2.5$			$\phi_1 = 1.5$		
		Mean	Median	SD	Mean	Median	SD
$n=50,\ n_i \sim U(90,110)$	MLE	3.332	3.334	0.663	1.019	1.017	0.026
	LTS $\alpha=0.2$	–2.535	–2.539	0.283	1.500	1.500	0.004
	LTS $\alpha=0.1$	–2.521	–2.522	0.210	1.500	1.500	0.003
$n=50,\ n_i \sim U(18,22)$	MLE	3.293	3.298	1.069	1.490	1.487	0.045
	LTS $\alpha=0.2$	–2.471	–2.438	0.616	1.500	1.500	0.015
	LTS $\alpha=0.1$	–2.496	–2.485	0.539	1.499	1.500	0.014
$n=200,\ n_i \sim U(450,550)$	MLE	3.310	3.314	0.180	1.017	1.017	0.006
	LTS $\alpha=0.2$	–2.502	–2.507	0.130	1.500	1.500	0.001
	LTS $\alpha=0.1$	–2.502	–2.505	0.089	1.500	1.500	0.001
$n=200,\ n_i \sim U(18,22)$	MLE	3.359	3.417	1.049	1.493	1.496	0.046
	LTS $\alpha=0.2$	–2.464	–2.437	0.290	1.500	1.500	0.007
	LTS $\alpha=0.1$	–2.488	–2.491	0.267	1.500	1.499	0.007

IV. REAL DATA ANALYSIS

In this Section, we illustrate the application of the proposed estimation procedure of MESR to forecast the electric load in Southern California, USA.

A. Data

The electric consumption data are aggregated to fifty-two 220 kV transformer banks from December 31, 2012 to November 1, 2013 in Southern California Edison's service territory. The task is to build a forecasting model for the total electricity consumption of residential customers at each 220 kV transformer bank on weekdays.

The data cleansing of the raw dataset is done in two steps. First, we exclude daily observations for commercial customers and remove zero-usage records from the electric consumption data file. Second, we add daily temperature and humidity information for each bank according to its zipcodes.

The response variable U_t is the aggregated customers' hourly electricity consumption recorded by the smart meters. We use the following transformation to make it comparative among 52 subgroups:

TABLE VI
SIMULATION RESULTS OF BREAKPOINT ESTIMATES BY MLE AND LTS WITH DIFFERENT α LEVELS FOR SITUATION WITH OUTLIERS

Parameter	Method	$\varphi_1=6.667$			$\varphi_2=13.333$		
		Mean	Median	SD	Mean	Median	SD
$n=50,\ n_i\sim U(90,110)$	MLE	6.647	6.679	0.546	13.322	13.317	0.324
	LTS $\alpha=0.2$	6.670	6.380	0.755	13.351	13.342	0.519
	LTS $\alpha=0.1$	6.670	6.677	0.415	13.323	13.332	0.283
$n=50,\ n_i\sim U(18,22)$	MLE	6.319	6.317	0.204	12.656	12.651	0.139
	LTS $\alpha=0.2$	6.425	6.445	0.185	13.027	13.037	0.237
	LTS $\alpha=0.1$	6.603	6.616	0.170	13.116	13.117	0.145
$n=200,\ n_i\sim U(450,550)$	MLE	6.671	6.673	0.172	13.333	13.337	0.098
	LTS $\alpha=0.2$	6.670	6.685	0.293	13.325	13.330	0.166
	LTS $\alpha=0.1$	6.670	6.677	0.166	13.328	13.330	0.094
$n=200,\ n_i\sim U(18,22)$	MLE	6.307	6.316	0.209	12.643	12.649	0.144
	LTS $\alpha=0.2$	6.427	6.433	0.082	13.030	13.039	0.118
	LTS $\alpha=0.1$	6.606	6.610	0.076	13.190	13.116	0.067

TABLE VII
SIMULATION RESULTS OF BREAKPOINT SLOPE ESTIMATES BY MLE AND LTS WITH DIFFERENT α LEVELS FOR SITUATION WITH OUTLIERS

Parameter	Method	$\beta_0=1.5$			$\beta_1=1.5$			$\beta_2=-2.5$		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
$n=50,\ n_i\sim U(90,110)$	MLE	1.493	1.494	0.109	1.521	1.518	0.154	–2.516	–2.522	0.163
	LTS $\alpha=0.2$	1.501	1.502	0.123	1.510	1.500	0.166	–2.519	–2.505	0.168
	LTS $\alpha=0.1$	1.506	1.503	0.070	1.505	1.501	0.083	–2.509	–2.507	0.071
$n=50,\ n_i\sim U(18,22)$	MLE	1.590	1.590	0.039	1.544	1.536	0.074	–2.610	–2.602	0.135
	LTS $\alpha=0.2$	1.556	1.556	0.027	1.584	1.564	0.102	–2.602	–2.580	0.109
	LTS $\alpha=0.1$	1.575	1.576	0.029	1.557	1.557	0.069	–2.590	–2.588	0.070
$n=200,\ n_i\sim U(450,550)$	MLE	1.500	1.502	0.035	1.500	1.499	0.040	–2.500	–2.502	0.042
	LTS $\alpha=0.2$	1.500	1.502	0.054	1.502	1.500	0.051	–2.499	–2.502	0.054
	LTS $\alpha=0.1$	1.495	1.497	0.022	1.506	1.505	0.026	–2.506	–2.499	0.028
$n=200,\ n_i\sim U(18,22)$	MLE	1.587	1.586	0.037	1.549	1.544	0.075	–2.605	–2.611	0.138
	LTS $\alpha=0.2$	1.574	1.574	0.014	1.557	1.556	0.048	–2.499	–2.502	0.054
	LTS $\alpha=0.1$	1.553	1.553	0.014	1.571	1.566	0.035	–2.506	–2.499	0.028

TABLE VIII
SIMULATION RESULTS OF RANDOM EFFECT ESTIMATES BY MLE AND LTS WITH DIFFERENT α LEVELS FOR SITUATION WITH OUTLIERS

Parameter	Method	$\sigma_{r1}=1$			$\sigma_{r2}=1$			$\rho=0.5$		
		Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
$n=50,\ n_i\sim U(90,110)$	MLE	0.518	0.379	0.582	1.019	1.026	0.117	0.843	0.999	0.612
	LTS $\alpha=0.2$	1.000	0.993	0.097	0.993	1.003	0.097	0.509	0.510	0.115
	LTS $\alpha=0.1$	0.992	0.998	0.111	0.994	0.999	0.097	0.511	0.519	0.107
$n=50,\ n_i\sim U(18,22)$	MLE	0.529	0.529	0.111	0.533	0.523	0.318	0.815	0.816	0.158
	LTS $\alpha=0.2$	0.823	0.825	0.121	0.779	0.779	0.066	0.498	0.485	0.066
	LTS $\alpha=0.1$	0.746	0.744	0.114	1.080	1.073	0.079	0.494	0.493	0.079
$n=200,\ n_i\sim U(450,550)$	MLE	0.769	0.897	0.435	1.013	1.015	0.056	0.619	0.592	0.272
	LTS $\alpha=0.2$	0.992	0.990	0.048	0.998	0.998	0.047	0.495	0.496	0.050
	LTS $\alpha=0.1$	0.992	0.989	0.048	0.998	0.997	0.047	0.496	0.496	0.049
$n=200,\ n_i\sim U(18,22)$	MLE	0.535	0.531	0.118	0.553	0.550	0.315	0.787	0.787	0.148
	LTS $\alpha=0.2$	0.884	0.848	0.060	0.781	0.782	0.033	0.476	0.473	0.085
	LTS $\alpha=0.1$	0.763	0.763	0.056	1.082	1.082	0.038	0.493	0.493	0.071

$$L_t = \ln \frac{U_t}{A_{total}} \tag{9}$$

In (9), the transformed response variable is derived as follows. First, we divide the aggregated usage by the total air conditioning tonnage of a residential customer. The customer is in the air conditioning cycling program. Second, we apply the log-transformation. The electricity consumption is divided by the total AC tonnage because the latter determines the numerical magnitude of the load measurements. Since the new response variable represents the electricity consumption level per unit of air conditioning tonnage, the effects of other explanatory variables are comparable among different transformer banks, which allows to use common slopes to simplify the model. Figure 1 shows the transformed response variable for a few sample banks over 5 days.

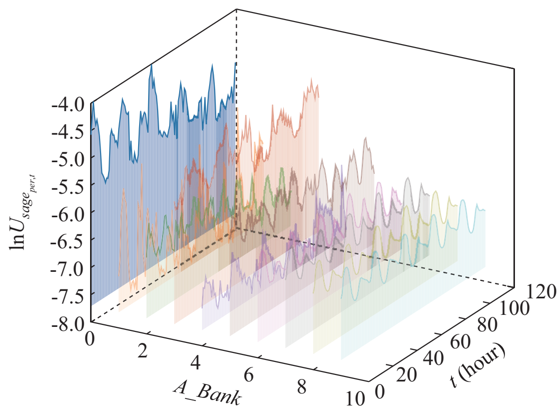

Fig. 1. Data visualization of transformed response variable L_t (shown in the figure as $\ln U_{sage_{per,t}}$) versus transformer bank indicator variable B (shown in the figure as *A_Bank*) and time.

The explanatory variables are collected and listed in Table IX. A two-day lagged electricity consumption variable is selected rather than the one-day lagged variable, because the estimates of the load impact of the demand response resources need to be submitted to the independent system operator one day before the actual operations. The average temperature and humidity are included because they are highly correlated with the electricity consumption. The duty cycle option variable indicates the percentage of participation rate of air conditioning load in the program, and has a substantial influence over the load impact for air conditioning cycling demand response program. The transformer bank indicator variable B is chosen as the random effect, because it contains information about the data from different geographic areas, which is thus expected to be heterogeneous with different baselines. A random effect model, assuming that B are sampled from a larger population, is able to incorporate the individual-specific heterogeneity of B while allowing to borrow information across B with much smaller number of parameters (compared with one fixed effect parameter for each of 52 values of B). In addition, this allows to extend the model to additional transformer banks.

In this paper, the training dataset is chosen as the samples in the first 205 observed weekdays for all transformer banks. The testing dataset consists of the samples from the 10 observed weekdays immediately following the training dataset. The total number of testing sample is 12480.

TABLE IX
SEVEN EXPLANATORY VARIABLES IN REAL DATA APPLICATION

Notation	Explanatory variable
L_t^{lag}	Base e log of two-day lagged electricity consumption
T_t	Daily average ambient temperature
H_t	Humidity of the day
Hr_t	Hour/time of the day
A_t	Duty cycle percentage
A_{total}	Total AC tonnage under the same transformer bank
B	Indicator variable of transformer bank

Note: B is the random effect variable; and Hr_t is the segmented variable.

B. Model and Result

We apply the proposed estimation procedure of MESR to forecast the electricity consumption. Figure 2 shows the hourly trend for average electric consumption and its forecasting results for a typical day. Two breakpoints locate between 02:00-3:00 a.m. and 06:00-08:00 p.m., respectively.

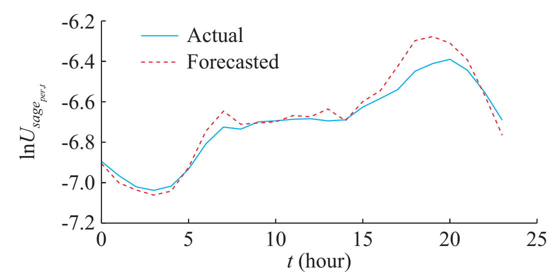

Fig. 2. Hourly trend between average hourly electric consumption L_t (shown in the figure as $\ln U_{sage_{per,t}}$) with variable Hr_t (shown in the figure as t) averaged over all transformer bank B.

It seems that the curve corresponding to the actual consumption (after the log transformation) indicates three segments with two breakpoints. The first breakpoint locates between 02:00 a.m. and 03:00 a.m., and the second breakpoint lies between 06:00 p.m. and 08:00 p.m.. We have also tried the model with three breakpoints (one more breakpoint in the middle segmented area), but the BIC for two breakpoints is smaller. The forecasting curve in Fig. 2 corresponds to the forecasted values across all transformer banks for the same forecasting day. As shown in Fig. 2, we can see that the proposed model can forecast the actual values well and the fitted values also match the breakpoint relationship.

The observations collected over time within the same transformer bank are correlated. The auto-correlation function of the observation time series of each transformer bank is shown in Fig. 3, where the time series demonstrates strong auto-correlation patterns.

Ignoring such correlation by fixed effect model would result in inefficient estimates and lose forecasted power. To incorporate such correction, the transformer bank is treated as random effects. Using a random effect model can also drastically reduce the number of unknown parameters in the model, and thus lead to more efficient parameter estimates.

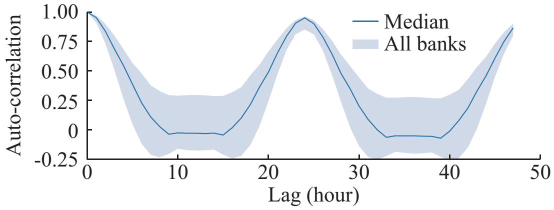

Fig. 3. Auto-correlation function of observation time series of each transformer bank.

Next, we describe the construction of the fixed effects. The first six explanatory variables described in Table IX, along with their two-way and three-way interactions, are considered as potential explanatory variables. The model variables are selected via least absolute shrinkage and selection operator (LASSO) regression method [54], which improves the forecasting accuracy and interpretability of the model. The final selected MESR is expressed as:

$$\begin{aligned} L_t = B + {} & Hr_t + (Hr_t - \varphi_1)_+ + (Hr_t - \varphi_2)_+ + T_t + H_t + A_t + L_t^{lag} + \\ & [Hr_t + (Hr_t - \varphi_1)_+ + (Hr_t - \varphi_2)_+]T_t + [Hr_t + (Hr_t - \varphi_2)_+]H_t + \\ & [Hr_t + (Hr_t - \varphi_1)_+ + (Hr_t - \varphi_2)_+]A_t + (T_t + H_t)A_t + T_t H_t \end{aligned} \tag{10}$$

where $B \sim N(0, \sigma^2_{A\_Bank})$ is the normal distribution with mean 0 and variance $\sigma^2_{A\_Bank}$. We apply both MLE and LTS algorithms to estimate the model and compare their forecasting performances. Since the true proportion of outliers is unknown, three proportions $\alpha = 0.15, 0.10, 0.05$ are selected for LTS to fit the model (10). In addition, the proposed algorithms are compared with two benchmarks, i.e., the multiple linear regression model [55] and the cooperative quantile regression forest (QRF)/multivariate quantile regression (MQR) method [12]. We set the training/testing data partitioning of these benchmarks to be the same as the setup discussed in Section IV-A. We compare their performances by mean absolute percentage error (MAPE) and root mean squared error (RMSE) on the testing dataset. The formula for MAPE and RMSE are given by $MAPE = \frac{1}{N}\sum \frac{|y_{ij} - \hat{y}_{ij}|}{y_{ij}}$ and $RMSE = \sqrt{\frac{1}{N}\sum (y_{ij} - \hat{y}_{ij})^2}$, where N is the total number of testing samples. For better comparison, we also report three quartiles of absolute percentage error (APE) and absolute error (AE).

From Tables X and XI, the proposed robust segmented mixed effect regression model (LTS in the tables) outperforms the multiple linear regression model [55] and the QRF/MQR model [55]. The improvements are more significant in terms of the RMSE and AE. The reason why the LTS has a slightly higher MAPE compared with the QRF/MQR baseline is that the LTS produces a bit larger estimation errors for some transformer banks with lighter loading. This results in a higher MAPE due to the small denominator. Within the LTS method, each evaluation criterion reaches the lowest value when $\alpha = 0.1$ and is much smaller than that of MLE. The breakpoint estimates shown in Table XII confirm the locations of the breakpoints shown in Fig. 2.

TABLE X
FORECASTING RESULTS EVALUATED BY APE FOR LAST 10 DAYS IN OCTOBER 2013 WITH MLE COMPARED WITH LTS WITH DIFFERENT α LEVELS

Method	MAPE (%)	25% APE (%)	50% APE (%)	75% APE (%)
GEFcom2012 [55]	17.97	5.95	12.06	19.80
QRF/MQR [12]	10.15	2.36	5.25	10.70
MLE	13.94	4.55	8.48	13.66
LTS $\alpha = 0.05$	11.08	2.78	5.45	9.37
LTS $\alpha = 0.1$	10.75	2.46	4.95	8.77
LTS $\alpha = 0.15$	10.88	2.55	5.10	9.01

TABLE XI
FORECASTING RESULTS EVALUATED BY RMSE FOR LAST 10 DAYS IN OCTOBER 2013 WITH MLE COMPARED WITH LTS WITH DIFFERENT α LEVELS

Method	RMSE (kWh)	25% AE (kWh)	50% AE (kWh)	75% AE (kWh)
GEFcom2012 [55]	1305.65	26.92	168.96	684.95
QRF/MQR [12]	742.73	10.01	67.77	298.06
MLE	672.88	5.78	42.19	164.08
LTS $\alpha = 0.05$	449.68	3.98	27.01	97.45
LTS $\alpha = 0.1$	414.42	3.65	24.73	86.33
LTS $\alpha = 0.15$	420.00	4.75	25.26	88.84

TABLE XII
BREAKPOINT ESTIMATION FOR ELECTRIC POWER DEMAND DATASET VIA MLE AND LTS WITH DIFFERENT α LEVELS

Method	φ_1 (hour)	φ_2 (hour)
MLE	2.64	20.47
LTS $\alpha = 0.05$	2.27	20.47
LTS $\alpha = 0.1$	2.27	20.77
LTS $\alpha = 0.15$	2.27	20.47

Table XIII displays the fixed effect and breakpoints slope estimates for LTS with $\alpha = 0.1$. The variance estimates of the random effects and the error term are 0.0015 and 0.0052, respectively.

According to Table XIII, all the parameters are significant with $\alpha = 0.05$. When calculating the p-values, Satterthwaite method is used for approximating degrees of freedom of the t-distribution for the t-test statistics. The variable Hr_t and its breakpoints have both positive and negative slopes, and the signs match the curves in Fig. 2. Also, there is a sensible positive relationship between A_t and electric load U_t.

V. CONCLUSION

In this paper, we propose a robust segmented mixed effect regression model to forecast the electric load baseline in Southern California. When estimating unknown parameters, we propose a backfitting algorithm by combining the ideas of the penalized least square method for random-effects regression model and the linearization technique [28] for seg-

TABLE XIII
PARAMETER ESTIMATES FOR ELECTRIC POWER DEMAND DATASET EVALUATED BY LTS WITH $\alpha = 0.1$ AND SIGNIFICANCE LEVEL OF 0.05

Parameter	Estimate	*p*-value
Intercept	1.063×10^{0}	<0.0001
Hr_t	2.100×10^{-2}	0.0003
$(Hr_t-\varphi_1)_+$	-4.713×10^{-2}	<0.0001
$(Hr_t-\varphi_2)_+$	6.665×10^{-2}	<0.0001
T_t	-2.198×10^{-2}	<0.0001
H_t	2.556×10^{-2}	0.0010
A_t	8.838×10^{-1}	<0.0001
L_t^{lag}	-2.223×10^{0}	<0.0001
$Hr_t H_t$	-2.493×10^{-5}	<0.0001
$(Hr_t-\varphi_2)_+ H_t$	2.017×10^{-4}	<0.0001
$Hr_t T_t$	-8.165×10^{-4}	<0.0001
$(Hr_t-\varphi_1)_+ T_t$	1.018×10^{-3}	<0.0001
$(Hr_t-\varphi_2)_+ T_t$	-1.606×10^{-3}	<0.0001
$Hr_t A_t$	2.012×10^{-2}	0.0002
$T_t H_t$	-6.684×10^{-5}	<0.0001
$T_t A_t$	2.763×10^{-2}	<0.0001
$H_t A_t$	2.763×10^{-2}	<0.0001

mented regression. In addition, we extend the idea of LTS to MESR to provide a robust model estimate. Both simulation study and real data application demonstrate the effectiveness of the proposed new estimation procedures.

Since the model is built up with hourly data, we could also aggregate the data and construct a daily electric load model. In this paper, we assume that the number of breakpoints is known. If the number of breakpoints is unknown, one could apply the selection techniques proposed by [56]-[59] to our model. In (1), all random effects are assumed to have a multivariate normal distribution. It will be interesting to extend the work of [60] to relax the normality assumption of the random effects in (1). In addition, for LTS, although an conservation α or serval α values can be used in practice, it requires some careful tuning of α so that LTS can have both robustness and high efficiency. Since the response variable by AC tonnage is normalized, it is expected that there is not too much heterogeneity for the effects of other variables after controlling the heterogeneity of different banks. But it is worthy of more research trying some more complicated models such as random slopes for all other variables and their interactions as well as nonparametric regression for hour, humidity, or temperature variables.

REFERENCES

[1] R. Adib, H. Murdock, F. Appavou *et al.* (2016, Dec.). Renewables 2016 global status report. Global Status Report Renewable Energy Policy Network for the 21st Century (REN21). [Online]. Available: https://www. ren21. net/wp-content/uploads/2019/05/REN21_GSR2016_FullReport_en_11.pdf

[2] Q. Wang, C. Zhang, Y. Ding *et al.*, "Review of real-time electricity markets for integrating distributed energy resources and demand response," *Applied Energy*, vol. 138, pp. 695-706, Jan. 2015.

[3] S. Nolan and M. O'Malley, "Challenges and barriers to demand response deployment and evaluation," *Applied Energy*, vol. 152, pp. 1-10, Aug. 2015.

[4] T. Wei, Q. Zhu, and N. Yu, "Proactive demand participation of smart buildings in smart grid," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1392-1406, May 2016.

[5] N. Yu, T. Wei, and Q. Zhu, "From passive demand response to proactive demand participation," in *Proceedings of 2015 IEEE International Conference on Automation Science and Engineering (CASE)*, Gothenburg, Sweden, pp. 1300-1306, Aug. 2015.

[6] N. Charlton and C. Singleton, "A refined parametric model for short term load forecasting," *International Journal of Forecasting*, vol. 30, no. 2, pp. 364-368, Apr. 2014.

[7] S. B. Taieb, J. W. Taylor, and R. J. Hyndman, "Hierarchical probabilistic forecasting of electricity demand with smart meter data," *Journal of the American Statistical Association*, vol. 6, pp. 1-17, Mar. 2020.

[8] B. Goehry, Y. Goude, P. Massart *et al.*, "Aggregation of multi-scale experts for bottom-up load forecasting," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 1895-1904, May 2020.

[9] Y. Chen, P. Xu, Y. Chu *et al.*, "Short-term electrical load forecasting using the support vector regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659-670, Jun. 2017.

[10] K. Chen, K. Chen, Q. Wang *et al.*, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943-3952, Jul. 2019.

[11] L. Sehovac and K. Grolinger, "Deep learning for load forecasting: sequence to sequence recurrent neural networks with attention," *IEEE Access*, vol. 8, no. 8, pp. 36411-36426, Feb. 2020.

[12] A. Bracale, P. Caramia, P. De Falco *et al.*, "Multivariate quantile regression for short-term probabilistic load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 628-638, Jan. 2020.

[13] Z. Cao, C. Wan, Z. Zhang *et al.*, "Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1881-1897, May 2020.

[14] P. I. Feder, "The log likelihood ratio in segmented regression," *The Annals of Statistics*, vol. 3, no. 1, pp. 84-97, Jan. 1975.

[15] R. Beckman and R. Cook, "Testing for two-phase regressions," *Technometrics*, vol. 21, no. 1, pp. 65-69, Feb. 1979.

[16] J. E. Ertel and E. B. Fowlkes, "Some algorithms for linear spline and piecewise multiple linear regression," *Journal of the American Statistical Association*, vol. 71, no. 355, pp. 640-648, Apr. 1976.

[17] A. Tishler and I. Zang, "A maximum likelihood method for piecewise regression models with a continuous dependent variable," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 30, no. 2, pp. 116-124, Jun. 1981.

[18] A. K. Wagner, S. B. Soumerai, F. Zhang *et al.*, "Segmented regression analysis of interrupted time series studies in medication use research," *Journal of Clinical Pharmacy and Therapeutics*, vol. 27, no. 4, pp. 299-309, Aug. 2002.

[19] H.-J. Kim, M. P. Fay, E. J. Feuer *et al.*, "Permutation tests for joinpoint regression with applications to cancer rates," *Statistics in Medicine*, vol. 19, no. 3, pp. 335-351, Jan. 2000.

[20] J. D. Toms and M. L. Lesperance, "Piecewise regression: a tool for identifying ecological thresholds," *Ecology*, vol. 84, no. 8, pp. 2034-2041, Aug. 2003.

[21] Q. Shao and N. Campbell, "Applications: modelling trends in groundwater levels by segmented regression with constraints," *Australian & New Zealand Journal of Statistics*, vol. 44, no. 2, pp. 129-141, Dec. 2002.

[22] A. E. Kunst, C. W. Looman, and J. P. Mackenbach, "Outdoor air temperature and mortality in the netherlands: a time-series analysis," *American Journal of Epidemiology*, vol. 137, no. 3, pp. 331-341, Feb. 1993.

[23] N. Molinari, J.-P. Daurès, and J.-F. Durand, "Regression splines for threshold selection in survival data analysis," *Statistics in Medicine*, vol. 20, no. 2, pp. 237-247, Jan. 2001.

[24] R. Rigby and D. Stasinopoulos, "Detecting break points in the hazard function in survival analysis," *Statistical Modelling*, vol. 1992, pp. 303-311, Dec. 1992.

[25] J. Shi, Y. Liu, and N. Yu, "Spatio-temporal modeling of electric loads," in *Proceedings of North American Power Symposium (NAPS)*, Morgantown, USA, Sept. 2017, pp. 1-6.

[26] N. M. Laird and J. H. Ware, "Random-effects models for longitudinal data," *Biometrics*, vol. 38, no. 4, pp. 963-974, Dec. 1982.

[27] P. Diggle, P. J. Diggle, P. Heagerty *et al.*, *Analysis of Longitudinal Data*. Oxford: Oxford University Press, 2002.

[28] V. M. Muggeo, "Estimating regression models with unknown breakpoints," *Statistics in Medicine*, vol. 22, no. 19, pp. 3055-3071, Sept. 2003.

[29] D. Bates. (2020, Nov.). Computational methods for mixed models. [Online]. Available: https://cran. r-project. org/web/packages/lme4/vignettes/Theory.pdf

[30] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871-880, Jan. 1984.

[31] V. M. Muggeo, D. C. Atkins, R. J. Gallop *et al.*, "Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study," *Statistical Modelling*, vol. 14, no. 4, pp. 293-313, May 2014.

[32] M. Muggeo. (2016, Feb.). Segmented mixed models with random changepoints in R. [Online]. Available: https://www. researchgate. net/publication/292629179

[33] A. Dominicus, S. Ripatti, N. L. Pedersen *et al*., "A random change point model for assessing variability in repeated measures of cognitive function," *Statistics in Medicine*, vol. 27, no. 27, pp. 5786-5798, Nov. 2008.

[34] T. Hastie and R. Tibshirani, *Generalized Additive Models*. Virginia Beach: Chapman and Hall/CRC, 1990.

[35] C. Gössl and H. Küchenhoff, "Bayesian analysis of logistic regression with an unknown change point and covariate measurement error," *Statistics in Medicine*, vol. 20, no. 20, pp. 3109-3121, Oct. 2001.

[36] J. Jiao, Z. Tang, P. Zhang *et al.*, "Ensuring cyberattack-resilient load forecasting with a robust statistical method," in *Proceedings of 2019 IEEE PES General Meeting (PESGM)*, Atlanta, USA, pp. 1-5, Aug. 2019.

[37] J. Luo, T. Hong, and S. Fang, "Robust regression models for load forecasting," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5397-5404, Sept. 2019.

[38] P. J. Huber, *Robust Statistics*. New York: John Wiley and Sons, 1981.

[39] L. A. Jaeckel, "Estimating regression coefficients by minimizing the dispersion of the residuals," *The Annals of Mathematical Statistics*, vol. 43, no. 5, pp. 1449-1458, Oct. 1972.

[40] A. F. Siegel, "Robust regression using repeated medians," *Biometrika*, vol. 69, no. 1, pp. 242-244, Apr. 1982.

[41] P. Rousseeuw and V. Yohai, "Robust regression by means of *s*-estimators," in *Robust and Nonlinear Time Series Analysis.* Berlin: Springer, 1984, pp. 256-272.

[42] V. J. Yohai, "High breakdown-point and high efficiency robust estimates for regression," *The Annals of Statistics*, vol. 15, no. 2, pp. 642-656, Jun. 1987.

[43] D. Gervini and V. J. Yohai, "A class of robust and fully efficient regression estimators," *The Annals of Statistics*, vol. 30, no. 2, pp. 583-616, Apr. 2002.

[44] Y. She and A. B. Owen, "Outlier detection using nonconvex penalized regression," *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 626-639, Jun. 2011.

[45] Y. Lee, S. N. MacEachern, and Y. Jung, "Regularization of case-specific parameters for robustness and efficiency," *Statistical Science*, vol. 27, no. 3, pp. 350-372, Aug. 2012.

[46] C. Yu, K. Chen, and W. Yao, "Outlier detection and robust mixture modeling using nonconvex penalized likelihood," *Journal of Statistical Planning and Inference*, vol. 164, no. 1, pp. 27-38, Sept. 2015.

[47] C. Yu and W. Yao, "Robust linear regression: a review and comparison," *Communications in Statistics-Simulation and Computation*, vol. 46, no. 8, pp. 6261-6282, Mar. 2017.

[48] D. L. Donoho and P. J. Huber, "The notion of breakdown point," in *A Festschrift for Erich L. Lehmann*. Belmon: Wadsworth International Group, 1983.

[49] N. Neykov, P. Filzmoser, R. Dimova *et al.*, "Robust fitting of mixtures using the trimmed likelihood estimator," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 299-308, Sept. 2007.

[50] M. Li, S. Xiang, and W. Yao, "Robust estimation of the number of components for mixtures of linear regression models," *Computational Statistics*, vol. 31, no. 4, pp. 1539-1555, Aug. 2016.

[51] L. Yang, S. Xiang, and W. Yao, "Robust fitting of mixtures of factor analyzers using the trimmed likelihood estimator," *Communications in Statistics–Simulation and Computation*, vol. 46, no. 2, pp. 1280-1291, Feb. 2017.

[52] V. M. Muggeo, "Segmented: an R package to fit regression models with broken-line relationships," *R News*, vol. 8, no. 1, pp. 20-25, Jan. 2008.

[53] D. Bates, M. Mächler, B. Bolker *et al*., "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1-48, Oct. 2015.

[54] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267-288, Jan. 1996.

[55] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 357-363, Apr. 2014.

[56] J. Liu, S. Wu, and J. V. Zidek, "On segmented multivariate regression," *Statistica Sinica*, vol. 7, no. 2, pp. 497-525, Apr. 1997.

[57] M. S. Ben Aïssa, M. Boutahar, and J. Jouini, "Bai and Perron's and spectral density methods for structural change detection in the US inflation process," *Applied Economics Letters*, vol. 11, no. 2, pp. 109-115, Feb. 2004.

[58] B. Strikholm and T. Teräsvirta, "Determing the number of regimes in a threshold autoregressive model using smooth transition autoregressions," Tech. Rep. SSE/EFI Working Paper Series in Economics and Finance, Jan. 2005.

[59] R. Prodan, "Potential pitfalls in determining multiple structural changes with an application to purchasing power parity," *Journal of Business & Economic Statistics*, vol. 26, no. 1, pp. 50-65, Jan. 2008.

[60] K. J. Lee and S. G. Thompson, "Flexible parametric models for random-effects distributions," *Statistics in Medicine*, vol. 27, no. 3, pp. 418-434, May 2008.

Xiaoyang Zhou received the M.S. degree in statistics, and Ph.D. degree in applied statistics from University of California, Riverside, USA, in 2014 and 2017, respectively. She currently works for Uber, San Francisco, USA, as a Data Scientist dedicating in measurement and planning. Her research interests include unsupervised machine learning algorithm, estimation algorithm using nonparametric and semi-parametric methods, and multi-attribution model of digital marketing.

Yuanqi Gao received the B.E. degree in electrical engineering from Donghua University, Shanghai, China in 2015, and the Ph.D. degree in electrical engineering from the University of California, Riverside, USA, in 2020, where he is currently a Postdoctoral Scholar in the Department of Electrical and Computer Engineering. His research interests include big data analytics and machine learning applications in smart grids.

Weixin Yao received the B.S. degree in statistics from the University of Science and Technology of China, Hefei, China, in 2002, and the Ph.D. degree in statistics from Pennsylvania State University, State College, USA, in 2007. He is currently a Full Professor of Statistics at the University of California, Riverside, USA. He is an Elected Member of International Statistical Institute from 2018 and an Editorial Board Member of Biometrics, Journal of Computational and Graphical Statistics, and The American Statistician. His main research interests include mixed models, nonparametric and semiparametric modeling, robust data analysis, and statistical applications in smart grid.

Nanpeng Yu received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2006, and the M.S. and Ph.D. degrees in electrical engineering from Iowa State University, Ames, USA, in 2007 and 2010, respectively. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of California, Riverside, USA. He is an Editor of the IEEE Transactions on Smart Grid and IEEE Transactions on Sustainable Energy. His research interests include machine learning and big data analytics, optimization and control of smart grid, and electricity market design.