Evaluating the Effectiveness of Conservation Voltage Reduction with Multilevel Robust Regression

Jinhui Yang
Department of Statistics
University of California, Riverside
Riverside CA, USA
jyang065@ucr.edu

Nanpeng Yu*
Department of Electrical and Computer Engineering
University of California, Riverside
Riverside CA, USA
nyu@ece.ucr.edu

Weixin Yao
Department of Statistics
University of California, Riverside
Riverside, CA
weixin.yao@ucr.edu

Alec Wong
Systems and Analytics
Southern California Edison
Rosemead CA, USA
alec.wong@sce.com

Larry Juang
Load and Price Forecasting
Southern California Edison
Rosemead CA, USA
larry.w.juang@sce.com

Raymond Johnson
Systems and Analytics
Southern California Edison
Rosemead CA, USA
raymond.johnson@sce.com

***Abstract*—Conservation voltage reduction (CVR) can effectively reduce electricity consumption and peak demand by keeping the customer voltages in the lower half of the permissible range. To facilitate widespread adoption of CVR, a reliable and robust CVR performance evaluation methodology is in critical need. However, it is difficult to accurately estimate the load reduction impact of CVR in practice. The data quality issues in supervisory control and data acquisition and advanced metering infrastructure make it challenging to distinguish a few percentage of load reduction from measurement errors and bad data. This paper develops a multilevel robust regression model within the framework of statistical experimental design to address the data quality issues. The proposed model is capable of providing robust and reliable estimates of load and voltage reduction from CVR at both distribution feeder and substation levels. The effectiveness of the proposed methodology is validated with field CVR demonstration data provided by a major California electric utility.**

***Index Terms*—Advanced Metering Infrastructure, Conservation Voltage Reduction, Distribution Voltage and VAR Control, Robust Regression**

I. Introduction

Rising electricity costs, stricter environmental regulations, and increased stress of growing distributed energy resources (DERs) on an aging energy infrastructure have attracted continued interest in energy efficiency solutions for power systems. The large infusion of investment in distribution automation around the world has enabled electric utilities to implement various energy efficiency programs in power distribution systems. Besides adopting energy-efficient appliances, conservation voltage reduction (CVR) is another effective energy saving technology for electricity consumers.

We gratefully acknowledge the support from Southern California Edison, NSF under award 1637258, and DOE under award DE-EE0007328. *Corresponding author: Nanpeng Yu.

CVR reduces electricity consumption and peak demand by keeping the customer voltages in the lower half of the permissible range (114V - 126V) allowed under American National Standards Institute (ANSI) standard for utilization voltage C84.1 [1]. A lower voltage will directly lead to reduced consumption from the constant impedance and constant current components of the electric load [2]. With the installation of supervisory control and data acquisition (SCADA) and advanced metering infrastructure (AMI) systems, CVR is now typically implemented as a part of the distribution Voltage/VAR control (DVVC) system at the substation level [3]. The controls of transformer load tap changer, voltage regulators, capacitor banks, and smart inverters are coordinated in DVVC to minimize the total electric loads of substations subject to power factor and customer voltages constraints.

Although CVR has been successfully implemented in many electric utilities around the world, there are still many barriers to its widespread adoption. The most important barrier to the successful evaluation and implementation of CVR technology is poor data quality issues. Missing data and bad data are prevalent in the distribution management system (DMS), geographical information system (GIS), SCADA and AMI systems of electric utilities. Typically, these systems do not have accurate secondary connectivity information [4] which makes it difficult to model the voltage drop from service transformers to customers in DVVC. Distribution feeders constantly undergo reconfiguration due to network faults, changing load and renewable generation. However, the reconfiguration information are not always accessible in real-time from the GIS or SCADA systems. Data gathered from AMI and SCADA systems in distribution networks generally have lower reliability than that of transmission networks due to network clock synchronization errors [5] and communication failures. It is extremely challenging to distinguish a small percentage of load reduction due to CVR from the above mentioned data quality issues.

The existing literature on CVR performance evaluation can be grouped into three categories: statistical experimental design approach, load forecasting based approach, and bottom-up appliance modeling and aggregation approach. In the statistical experimental design approach, treatment and control groups are first selected. The treatment group represents the substations/feeders or time periods where/when CVR is implemented and turned on. The control group represents the substations/feeders or time periods where/when CVR is not implemented or turned off. A simple average comparison [6] or a statistical regression model [2] involving both control and treatment group measurements can then be built to estimate the load and voltage reduction from CVR. In the load forecasting based approach, the electric load under normal voltage condition is estimated with regression models or support vector machine (SVM) [7]. The estimated load under normal voltage condition can then be compared with the load measured under CVR to quantify the load reduction. In the bottom-up appliance modeling and aggregation approach, the energy consumption of individual appliance as a function of service voltage is quantified through physical tests [8]. The electric load reduction produced by CVR for a distribution substation or feeder can then be synthesized based on the load composition of each customer class [9].

None of the existing literature directly addresses the data quality issues in the CVR evaluation problem. In this paper, we propose a multilevel robust regression method to quantify the load and voltage impacts of CVR. The proposed multilevel model is capable of assessing both the feeder and the substation level impacts of CVR. Most importantly, the robust regression method proposed in the paper is not sensitive to outliers or violations of assumptions by the underlying data-generating process. Therefore, it can be easily adopted by utilities to reliably quantify the load and voltage reduction from CVR in the presence of bad and missing data.

The rest of this paper is organized as follows. Section II introduces the CVR evaluation problem and provides the overall framework of our proposed methodology to evaluate the effectiveness of a CVR implementation. Section III presents the technical details of multilevel robust regression model and how it can be applied to solve the CVR evaluation problem. Section IV shows a case study to evaluate the effectiveness of CVR in Southern California Edison's systems. The conclusions are stated in section V.

II. Problem Statement

A. CVR Performance Metrics and Data Sources

The effectiveness of a CVR program is typically evaluated by the following metrics: percentage of voltage reduction (PVR), percentage of load reduction (PLR), and conservation voltage reduction factor (CVR_f), where CVR_f is defined as $CVR_f \triangleq \frac{PLR}{PVR}$. PVR depends on the Volt-VAR control strategy before the CVR implementation, the actual CVR control algorithm, the spatio-temporal distribution of electric loads, and the topology and network parameters of the distribution feeders under the substation. CVR_f depends on the load composition of each customer and how the energy consumption from individual appliance responds to voltage variations.

CVR is usually implemented at the substation level which involves multiple distribution feeders. The control objective of CVR is typically minimizing the substation electric loads or losses subject to voltage and power factor constraints. The operations of equipment such as capacitor banks and smart inverters in one distribution feeder will certainly affect the voltage and electric load of another feeder. Hence, the effectiveness of CVR should be evaluated at both distribution feeder level and substation level. The bi-level evaluation will not only reveal the overall effectiveness of the CVR at the substation level but also the varying degrees of CVR effectiveness at the distribution feeder level.

To conduct a comprehensive evaluation of a CVR implementation, we need to extract information from many data sources which include SCADA and AMI data. The SCADA system provides the electric load, voltage, and current data at the substation and feeder level. It also contains information about the operating status and voltage measurement for individual substation, field capacitors, and smart inverters. The operating schedule of CVR is also included in the SCADA system. The AMI records information about the individual customers' electric power consumption and voltage magnitude.

B. CVR Evaluation Framework

This subsection provides a high-level overview of the proposed CVR evaluation framework. The general framework consists of four steps.

In step one, data preprocessing is performed to remove outliers. For example, the distribution substation/feeders' load may change drastically due to unexpected reconfiguration or restoration activities. The data set recorded in the reconfiguration and restoration periods may be excluded from the analysis.

In step two, the pairs of treatment and control substations/feeders are selected based on geographical location, network topology, voltage level, and electric load profile. The details of the treatment and control substations/feeders selection method are described in Section III.A.

In step three, multilevel robust regression models are developed to quantify the load reduction and voltage reduction impacts of CVR.

In step four, the PLR, PVR and CVR_f are calculated at both feeder and substation levels based on the multilevel robust regression model outputs.

III. Technical Methods

A. Treatment/Control Pair Selection

CVR is implemented on the treatment substations. Each treatment substation is paired with a control substation on which the CVR is not implemented. The control substation is selected based on the similarity between its geographical location, load composition and those of the treatment substation. Specifically, in this study each treatment substation is paired with a control substation of similar load profile within a 15 miles radius. Each treatment substation consists of multiple distribution feeders. Each one of the feeders in the treatment substation is paired with a feeder in the corresponding control substation. The control feeders are selected based on the goodness of fit of a series of regression models. In these regression models, the dependent variable is the treatment feeders' electric load and the independent variables are the feeders' electric load in the control substation. The feeder whose electric load serves as the best predictor for the treatment feeder's electric load in terms of the goodness of fit will be selected as the control feeder.

B. Multilevel Regression Models

1) Background and Notations: Multilevel regression models will be developed in this subsection to estimate PLR, PVR and CVR_f. A treatment and control substations pair is selected by the criterion described in Section III.A. Suppose there are J pairs of treatment and control feeders in the treatment and control substation pair.

We define the following notations which will be used in the multilevel regression models. Let t denote hour t during the testing period. Let $U^T_{j,t}$ and $V^T_{j,t}$ represent the average electric load and average voltage magnitude of the jth treatment feeder in hour t. Similarly, in the corresponding jth control feeder, $U^C_{j,t}$ and $V^C_{j,t}$ represent the average electric load and average voltage magnitude in hour t. Let D_t denote the CVR operation status of the treatment feeder. If $D_t = 1$, then the treatment feeder is operated under CVR in hour t. If $D_t = 0$, then CVR is turned off on the treatment feeder in hour t.

Next multilevel regression models will be constructed to quantify PLR and PVR respectively.

2) Percentage Load Reduction: For each substation, the following multilevel regression model is constructed to estimate the treatment feeder electric load $U^T_{j,t}$:

$$\begin{aligned}\log(U^T_{j,t}) = \beta_0 + \beta_1 \log(U^C_{j,t}) + \beta_2 D_t + \beta_{3,j} C_{j,t} \\ + \beta_{4,j} C_{j,t} \cdot D_t + \beta_{5,j} \log(U^C_{j,t}) \cdot C_{j,t} + \epsilon_{U^T_{j,t}}\end{aligned} \quad (1)$$

where $C_{j,t}$ is the indicator of jth pair of treatment/control feeders. If the data collected in hour t is from the jth treatment/control pair, then $C_{j,t} = 1$. Otherwise, $C_{j,t} = 0$. $\epsilon_{U^T_{j,t}}$ are independent and identically distributed errors with zero mean. For the identifiability of coefficients, we assume $\sum_j \beta_{3,j} = \sum_j \beta_{4,j} = \sum_j \beta_{5,j} = 0$. By including the interactions between $C_{j,t}$ and $\{D_t, \log(U^C_{j,t})\}$ in Equation (1), we allow the effects of conservation voltage reduction D_t and control feeder load $\log(U^C_{j,t})$ on the treatment feeder load to be different in each treatment feeder. The feeder indicators $C_{j,t}$ are introduced to unify the regression models for different feeders in the same substation so that we can leverage Equation (1) to model the entire substation directly instead of fitting a separate regression model for each feeder. In this way, we estimate the CVR impact at the substation level with each feeder as a block factor.

The $\log$ transformation is applied on the feeder loads in Equation (1) so that the regression parameters can be easily interpreted. For example, the coefficient $\beta_2 + \beta_{4,j}$ represents the impact of CVR on load reduction in the jth treatment feeder. In order to express PLR_j in terms of $\beta_2 + \beta_{4,j}$, we define two more load time series. Denote $U^{T_{On}}_{j,t}$ and $U^{T_{Off}}_{j,t}$ as the average electric load of the jth treatment feeder in hour t if CVR was turned on and off respectively.

From Equation (1), we have

$$\log\left(\frac{U^{T_{On}}_{j,t}}{U^{T_{Off}}_{j,t}}\right) = \beta_2 + \beta_{4,j} \Rightarrow \frac{U^{T_{On}}_{j,t}}{U^{T_{Off}}_{j,t}} = \exp\left(\beta_2 + \beta_{4,j}\right) \quad (2)$$

Then, the percentage load reduction on the jth feeder PLR_j due to CVR can be quantified as,

$$\begin{aligned}PLR_j &= 100 \cdot \frac{U^{T_{Off}}_{j,t} - U^{T_{On}}_{j,t}}{U^{T_{Off}}_{j,t}} \\ &= 100 \cdot [1 - \exp\left(\beta_2 + \beta_{4,j}\right)]\end{aligned} \quad (3)$$

After PLR is calculated for all feeders under the substation, the percentage load reduction for the treatment substation can be calculated by Equation (4) below.

$$PLR = \frac{\sum_{j=1}^{J} (\bar{U}^T_j \cdot PLR_j)}{\sum_j^J \bar{U}^T_j} \quad (4)$$

where $\bar{U}^T_j$ is the average load of the jth treatment feeder. The PLR at the substation level is calculated as a weighted sum of individual feeder's PLR where the weights are the electric loads of each feeder.

Note that based on the unified model (1), we can also test if the impacts of CVR on the load reduction varies across different feeders for each substation. This can be done by running a statistical test with the null hypothesis that $\beta_{4,j} = 0$ for $j = 1, ..., J$.

3) Percentage Voltage Reduction: Similarly, for each substation, the following multilevel regression model is constructed to estimate treatment circuit voltage $V^T_{j,t}$

$$\begin{aligned}\log(V^T_{j,t}) = b_0 + b_1 \log(V^C_{j,t}) + b_2 D_t + b_{3,j} C_{j,t} \\ + b_{4,j} C_{j,t} \cdot D_t + b_{5,j} \log(V^C_{j,t}) \cdot C_{j,t} + \epsilon_{V^T_{j,t}}\end{aligned} \quad (5)$$

where $C_{j,t}$ is the indicator of jth pair of treatment/control feeders, and $\epsilon_{V^T_{j,t}}$ are independent and identically distributed errors with zero mean. For the identifiability of coefficients, we assume $\sum_j b_{3,j} = \sum_j b_{4,j} = \sum_j b_{5,j} = 0$.

The $\log$ transformation is applied on the feeder voltages in Equation (5) so that the regression parameters can be easily interpreted.

Similar to the derivation in the multilevel regression model for load reduction, the percentage voltage reduction on the jth feeder PVR_j due to CVR can be quantified as

$$PVR_j = 100 \cdot [1 - \exp\left(b_2 + b_{4,j}\right)] \quad (6)$$

After PVR is calculated for all feeders under the substation, the percentage voltage reduction for the treatment substation can be calculated by Equation (7) below:

$$PVR = \frac{\sum_{j=1}^{J}(\bar{V}_j^T \cdot PVR_j)}{\sum_j^J \bar{V}_j^T} \tag{7}$$

Where $\bar{V}_j^T$ is the average voltage of the jth treatment feeder. The PVR at the substation level is calculated as a weighted sum of individual feeder's PVR where the weights are the average voltage of each feeder.

Note that based on the unified model (5), we can test if the impact of CVR on the voltage reduction varies across different feeders for each substation. This can be done by running a statistical test with the null hypothesis that $b_{4,j} = 0$ for all $j = 1, .., J$.

C. Model Parameter Estimation

The methods for estimating the coefficients in Equations (1) and (5) are presented in this subsection. Both the ordinary least squares (OLS) method and robust regression methods will be discussed.

1) OLS Method: To better explain how to estimate the model parameters in (1) and (5), we write them in a unified model

$$y_i = \boldsymbol{x_i'}\boldsymbol{\beta} + \epsilon_i, \tag{8}$$

where $\boldsymbol{x_i}$ is the observed $p \times 1$ covariate vector, $\boldsymbol{\beta}$ is an unknown $p \times 1$ vector and the $\epsilon_i's$ are iid with E$(\epsilon) = 0$. The most popular and simplest method to estimate $\boldsymbol{\beta}$ is the OLS estimator which minimizes the sum of squared residuals. The OLS estimates are given by

$$\hat{\boldsymbol{\beta}}_{OLS} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n}(y_i - \boldsymbol{x_i'}\boldsymbol{\beta})^2 \tag{9}$$

The OLS estimate $\hat{\boldsymbol{\beta}}_{OLS}$ has an explicit formula and thus is simple to compute and use.

2) Robust Regression Method: As explained in Section I, real-world data most likely contain outliers or problematic observations. It is widely known that the OLS estimate is highly sensitive to the outliers that do not follow the relationship/model assumption of the majority of the data samples. Even a single outlier can have large effect on the OLS estimate and may give a misleading result.

A practical method to deal with the outliers is to first remove them and then apply the OLS to the "clean data". However, in most of situations, the outliers can not be fully or easily identified and some of them may occur due to unexpected or unknown reasons. One could use the residual plot from the OLS estimates to identify outliers. However, the OLS estimates can be severely affected by the outliers and thus its resulting residuals can not be effectively used to identify outliers. Due to the masking effect, some outliers might even have small residuals based on the OLS estimates and thus will be missed from the residual plot of OLS [10], [11].

We propose to apply the M-estimate [12], which replaces OLS criterion (9) with a robust criterion, to estimate regression parameters in the presence of outliers. M-estimate of $\boldsymbol{\beta}$ is given by the equation below.

$$\hat{\boldsymbol{\beta}}_M = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{y_i - \boldsymbol{x_i'}\boldsymbol{\beta}}{\hat{\sigma}}\right) \tag{10}$$

where $\rho(\cdot)$ is called the robust loss function and its derivative $\psi(\cdot) = \rho'(\cdot)$, is called the influence function. $\hat{\sigma}$ is an error scale estimate. One commonly used robust influence function is called Huber's ψ function [13], where

$$\psi_c(t) = \max\{-c, \min(c, t)\} \tag{11}$$

The value of $c = 1.345$ is recommended in practice and provides approximately a 95% relative efficiency when the error density is normally distributed. Therefore, the Huber's M-estimate can provide comparable performance to the OLS estimate where there are no outliers, but much better performance than OLS when there are outliers.

One of the major advantages of the Huber's M-estimate (10) over OLS is that it is not necessary to detect outliers in advance and can be applied to the original source data directly in the presence of outliers. The Huber's M-estimate can downweight the outliers automatically when estimating the regression coefficients.

Note that $\hat{\boldsymbol{\beta}}_M$ in (10) is a solution of the following equation

$$0 = \sum_{i=1}^{n} \psi(r_i)\boldsymbol{x_i} = \sum_{i=1}^{n} w_i \boldsymbol{x_i}(y_i - \boldsymbol{x_i'}\boldsymbol{\beta}) \tag{12}$$

where $r_i = (y_i - \boldsymbol{x_i'}\boldsymbol{\beta})/\hat{\sigma}$ is standardized residual and $w_i = \psi(r_i)/r_i$. The weight w_i plays an important role in (12) and can downweight the effect of outliers with suitably chosen ψ function. For OLS, $w_i = 1$ since $\rho(t) = \frac{1}{2}t^2$ and $\psi(t) = t$. Therefore, OLS can not downweight the effects of outliers. For Huber's M-estimate,

$$w_i = \begin{cases} 1, & |r_i| \le c; \\ c/|r_i|, & |r_i| > c. \end{cases} \tag{13}$$

Therefore, the weights of Huber's M-estimate are the same as those of OLS (equal to 1) for small residuals and are smaller for larger residuals. In addition, the weights go to 0 when the residuals go to infinity (Figure 1). Therefore, Huber's M-estimate can effectively downweight the effect of outliers and thus achieve robust parameter estimates. One might also use some other robust regression methods. See, for example, [14] for a review of some popular robust regression methods.

IV. Case Study

A. Data Description

Before Southern California Edison (SCE)'s territory-wide implementation of CVR, it first started a CVR performance evaluation project. 5 representative substations are selected as treatment substations where CVR is implemented and then turned on and off on a weekly basis. The CVR performance evaluation project started on August 1, 2016 and ended on June 30, 2017. The 5 representative substations in the treatment

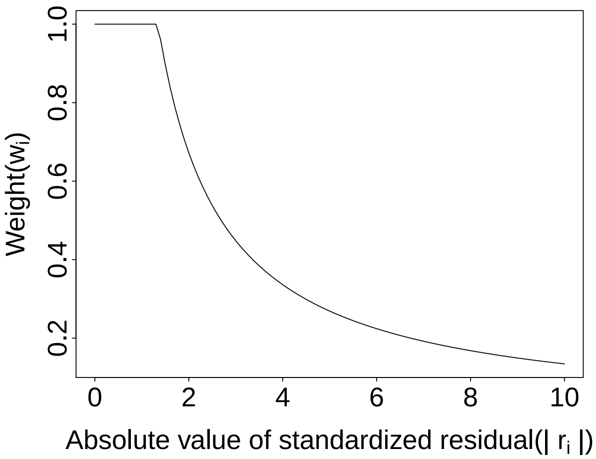

Figure 1: The weights of Hubers M-estimate as a function of standardized residuals

group include 28 distribution feeders. According to the control substations and feeders selection methodology specified in Section III.A, another 5 substations and 28 feeders are selected as part of the control group.

Electric load and voltage data are collected from SCADA and AMI systems. The electric load and voltage information are aggregated at both distribution feeder and substation levels. Both OLS estimation method and robust regression method are leveraged to quantify the impact of CVR on electric load and voltage at the feeder and substation level. In the OLS estimation approach, a set of dates specified by the system operators are excluded from the analysis based on network reconfiguration schedule and manual identification of outliers. In the robust regression approach, the M-estimates is leveraged to automatically identify outliers.

B. CVR Evaluation Results

1) OLS estimation approach: For each of the treatment and control substation pairs, a set of outliers specified by the system operators are removed from the analysis. The coefficients in Equations (1) and (5) are estimated by OLS estimation based on Equation (9). The estimates for PLR, PVR and CVR_f of all 5 treatment substations are reported in Table I. According to the OLS estimation approach, CVR resulted in 0.766% of voltage reduction and 0.716% of load reduction on average over the 5 treatment substations. The CVR factors (CVR_f) of the 5 treatment substations range from 0.84 to 1.29.

2) Robust regression approach: In the robust regression approach, we do not manually remove outliers. Instead Huber's M-estimate is leveraged to downweight the effect of outliers automatically. The coefficients in Equations (1) and (5) are estimated by M-estimates (with Huber's ψ function) from Equation (10). The estimates for PLR, PVR and CVR_f of all 5 treatment substations are calculated and reported in Table

Table I: PLR, PVR and CVR_f of the Treatment Substations Estimated by OLS Estimation Approach

Substation	$PLR(\%)$	$PVR(\%)$	CVR_f
1	0.59	0.53	1.11
2	0.38	0.75	0.51
3	1.25	1.23	1.02
4	0.74	0.58	1.29
5	0.62	0.74	0.84

II. Under the robust regression approach, CVR is expected to result in 0.746% reduction in voltage and 0.8% reduction in load. The CVR factors (CVR_f) of the 5 treatment substations range from 0.8 to 1.46. The robust regression approach with Huber's M-estimate provides reasonable estimates for CVR_f for all representative treatment substations. The PLR and PVR estimates from the robust regression approach are very similar to that of the OLS estimation approach with manual outlier removal.

Table II: PLR, PVR and CVR_f of the Treatment Substations Estimated by M-estimate (with Huber's ψ Function)

Substation	$PLR(\%)$	$PVR(\%)$	CVR_f
1	0.64	0.5	1.27
2	0.66	0.73	0.91
3	1.03	1.28	0.8
4	0.86	0.59	1.46
5	0.81	0.63	1.28

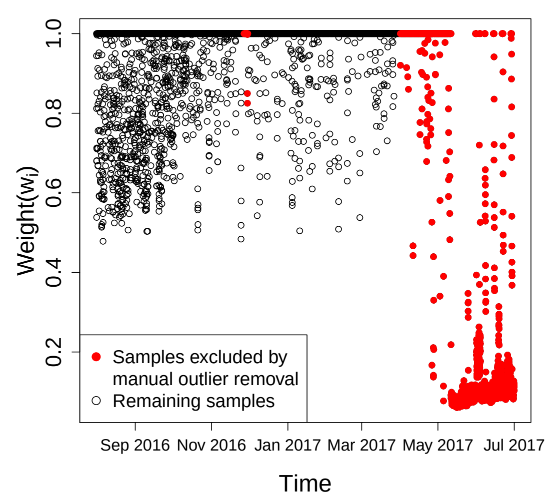

Figure 2: Weights of the Huber's M-estimate for excluded and remaining data points.

In order to illustrate the usefulness of the robust regression methods, the weights of the Huber's M-estimate are depicted in Fig. 2. As shown in the figure, the weights of the manually excluded hours are much lower on average than that of the remaining hours. Although the manual outlier removal procedure in OLS estimation excludes some problematic data, it deletes much more observations than necessary for conservativeness.

In order to evaluate the effectiveness of manual outlier removal and the M-estimate method, we calculate the proportions of outliers in excluded and remaining data set, respectively. Any observation with a residual $|r_i| > 3$ is considered as an outlier from Huber's M-estimate. Based on the estimates from the robust regression, we found that in most of substations the proportions of outliers in the excluded data are much larger than the proportions of outliers in the remaining "clean data". For example, in substation 1, the proportion of outliers in the excluded data from the manual outlier removal step is 7.29%. However, the proportion of outliers in the remaining "clean data" is only 0.47%. Therefore, the manual outlier removal procedure did exclude a set of data with a much larger proportion of outliers. However, there are still some outliers in the remaining data set after the manual outlier removal procedure.

3) The Homogeneity Effects of CVR on Electric Load and Voltage: Based on the unified model (1), we can test whether the effects of CVR on electric loads are the same across different feeders of the same substation. One way to check this is using an F-test. The null hypothesis is $H_0 : \beta_{4,j} = 0$ for $j = 1, ..., J$. The alternative hypothesis is H_a: at least one $\beta_{4,j} \neq 0$. Failure to reject H_0 means that the effect of CVR on electric load reduction is the same across all the feeders in a substation.

The F-test results for the 5 treatment substations are shown in Table III. As shown in the table with a significance level of $\alpha = 0.05$, H_0 is rejected only in substation 3 where two $\beta_{4,j}$s are nonzero. Therefore, the effect of CVR on load reduction differs by feeders only for substation 3.

Table III: F-test Results for Homogeneity Effects of CVR on Electric Load in Treatment Substations

Substation	F-test Results	J	F value	P-value
1	Fail to Reject	8	1.2535	0.2693
2	Fail to Reject	4	1.3108	0.2689
3	Reject H_0	5	3.8351	0.00405
4	Fail to Reject	5	2.0695	0.08193
5	Fail to Reject	6	1.8467	0.1001

Table IV: F-test Results for the Homogeneity Effects of CVR on Voltage in Treatment Substations

Substation	F-test Results	J	F value	P-value
1	Reject H_0	8	111.96	$< 2.2 \times 10^{-16}$
2	Reject H_0	4	269.45	$< 2.2 \times 10^{-16}$
3	Reject H_0	5	63.207	$< 2.2 \times 10^{-16}$
4	Reject H_0	5	437.61	$< 2.2 \times 10^{-16}$
5	Reject H_0	6	48.2	$< 2.2 \times 10^{-16}$

A similar F-test was conducted to test the homogeneity effects of CVR on voltage in all feeders under the same treatment substations. The F-test results are reported in Table IV. As shown in the table, the effects of CVR on voltage are the same across different feeders in every substation.

V. CONCLUSIONS

A multilevel robust regression method with Huber's M-estimate is proposed to estimate the impact of CVR on electric load and voltage reductions at both distribution feeder and substation level. A real-world CVR evaluation study shows that the proposed robust regression method can successfully address the data quality issues by downweighting the outliers automatically. The CVR evaluation results show that the robust regression method produces reliable and reasonable estimates of the percentage load reduction, percentage voltage reduction, and conservation voltage reduction factor for all treatment substations. Furthermore, the multilevel model structure is very flexible in evaluating whether the impacts of CVR on voltage and electric load are the same across different distribution feeders of a treatment substation.

In the future, we would like to include the DERs in the CVR evaluation framework as their penetration level in the distribution network continues to increase. With access to granular AMI data, the impacts of CVR on different types of customers in different hours will be studied in detail. From the methodology perspective, other robust regression methods such as LMS estimates and S-estimates will be evaluated and compared with the proposed M-estimate [14].

REFERENCES

[1] A. Standard, "C84. 1-1982," *American National Standard for Electric Power Systems and Equipment-Voltage Ratings (60Hz).*

[2] H. J. Liu, R. Macwan, N. Alexander, and H. Zhu, "A methodology to analyze conservation voltage reduction performance using field test data," in *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on.* IEEE, 2014, pp. 529–534.

[3] Z. Wang and J. Wang, "Review on implementation and assessment of conservation voltage reduction," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1306–1315, 2014.

[4] W. Wang, N. Yu, B. Foggo, J. Davis, and J. Li, "Phase identification in electric power distribution systems by clustering of smart meter data," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec 2016, pp. 259–265.

[5] Y. Gao and N. Yu, "State estimation for unbalanced electric power distribution systems using AMI data," in *Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017 IEEE.* IEEE, 2017, pp. 1–5.

[6] B. Kennedy and R. Fletcher, "Conservation voltage reduction (CVR) at Snohomish County PUD," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 986–998, 1991.

[7] Z. Wang, M. Begovic, and J. Wang, "Analysis of conservation voltage reduction effects based on multistage SVR and stochastic process," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 431–439, 2014.

[8] M. Chen, R. Shoults, J. Fitzer, and H. Songster, "The effects of reduced voltages on the efficiency of electric loads," *IEEE Transactions on Power Apparatus and Systems*, no. 7, pp. 2158–2166, 1982.

[9] J. De Steese, S. Merrick, and B. Kennedy, "Estimating methodology for a large regional application of conservation voltage reduction," *IEEE transactions on power systems*, vol. 5, no. 3, pp. 862–870, 1990.

[10] P. J. Huber, "Robust statistics," in *International Encyclopedia of Statistical Science.* Springer, 2011, pp. 1248–1251.

[11] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection.* John wiley & sons, 2005, vol. 589.

[12] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, pp. 73–101, 1964.

[13] P. J. Huber, "Robust statistics," *Wiley Series in Probability and Mathematical Statistics, New York: Wiley*, 1981.

[14] C. Yu and W. Yao, "Robust linear regression: a review and comparison," *Communications in Statistics-Simulation and Computation*, pp. 1–22, 2017.