ESTIMATING TREATMENT EFFECTS IN DEMAND RESPONSE

Pan Li and Baosen Zhang

Electrical Engineering
University of Washington
Seattle, WA 98195

ABSTRACT

Demand response is designed to motivate electricity customers to modify their loads at critical time periods. Accurately estimating customers response to demand response signals is crucial to the success of these programs. In this paper, we consider signals in demand response programs as a treatment to the customers and estimate the average treatment effect. Specifically, we adopt the linear regression model and derive several consistent linear regression estimators. From both synthetic and real data, we show that including more information about the customers does not always improve estimation accuracy: the interaction between the side information and the demand response signal must be carefully modeled. We then apply the so-called modified covariate method to capture these interactions and show it can strike a balance between having more data and model correctness. Our results are validated using data collected by Pecan Street.

Index Terms— Linear regression, treatment effect, demand response

1. INTRODUCTION

The electrical system is undergoing a transformation in both operation and design. A particular area that is changing dramatically is the balance of supply and demand. Instead of treating demand as inflexible load that must be met by changing generation levels, operators are starting to explore changing demand to balance intermittent generation such as solar and wind. This type of operation is commonly known as *demand response (DR)*. In typical implementation of DR programs, customers receive a DR signal such as modification in prices or simply a message requesting changes in electricity usage. An effective DR program improves the efficiency and sustainability of power systems by allowing utilities and operators to leverage flexibility in the load rather than relying on conventional generators [1, 2, 3, 4, 5, 6].

Demand response have received considerable attention from researchers and operators (e.g. see [7] and the references within). Most of works in the literature have viewed the DR problem from optimization or market design perspectives. For example, authors in [6, 8] considered how to optimize the social welfare; and authors in [9] have considered how to create an efficient market for demand response. In all these setups, customers' responses to demand response signals are either assumed to be known to the operators, or at least known to themselves. However, in practice, the customers often do not know their own utility functions and therefore cannot reveal this information to utilities even if they are perfectly truthful. This uncertainty about customer responses leads to uncertainty about the effectiveness of DR programs, since it is difficult for utilities to judge the impact of DR signals [10]. Namely, it is not trivial to answer basic questions such as whether the DR signal causes any significant change in customer consumption.

In this paper, we focus on the problem of estimating customer responses to DR signals from observational data. We adopt a *treatment model* [11], where we collect consumption data from customers, and DR signal is perceived as a treatment and is applied to some of them. The quantity of interest in this model is the *average treatment effect (ATE)*, representing the average response of the customers to the treatment. The fundamental question we investigate here is to determine if the ATE is statistically significant, and if so, estimating its value. Heuristic estimations such as simple averaging (averaging over the ten previous days for example) are popular and easy to implement [12], but lack model explanation. Here we use regression models, where the output is the measured energy consumption of customers. The inputs consist of the binary treatment indicator variables and *covariates*, denoting other information such as temperature and appliance information.

We derive three consistent linear regression estimators of the ATE through simple linear regression, multiple linear regression and modified covariate method presented in[13]. The first estimator requires the least information on how the covariates interact with the outcome by just considering the treatment variable. The second estimator assumes a much more stringent linear condition but is able to use more information by considering all covariates. The last estimator can be seen a compromise between the first two when the treatment effect is linear in the observed covariates. The estimators are tested based on both synthetic data and real data from Pecan Street[14]. We then compare the estimation performances based on the variance of the estimators. We show that somewhat contrary to common belief, including more data into the model especially when the assumed model is wrong does not perform better. In particular, we show for both data sets that under certain circumstances, simple linear regression is more efficient when the interactions between the covariates and the outcome remain unknown.

The fact that adding more data does not always improve estimation of the ATE is not new. This can be seen as the difference between prediction and inference [15]. Adding more data will almost always improve prediction. But in this context, we are trying to estimate the effect of a single variable, the DR signal, on the output. Having more knowledge about customers may actually "drown out" the relationship between the DR signal and customer consumption. A message from this paper is that the interactions between covariates and the treatment variable need to be carefully modeled to correctly leverage additional information contained in the covariates.

Many results about casual impact estimation has been developed (see [16, 17] and the references within). We choose to emphasize linear models in our analysis because of their ubiquity in both theory and practice, and because they do not involve calculating the distance metrics required in the matching algorithms.

The rest of the paper is organized as follows. Section 2 introduces the linear model and assumptions throughout this paper. Section 3 presents several different estimators based on various forms of linear regression. Section 4 details the case study on the performance of these estimators. Section 5 concludes the paper.

2. PROBLEM SETUP

2.1. Model

The Neyman-Rudin model [17] is commonly adopted when estimating the ATE. It suggests that in each observation of customer i, the outcome Y_i (in our case the consumption data) can only take on one of the two values, either $Y_i(0)$ or $Y_i(1)$. These two values are the potential outcomes either under the DR signal (when $T_i = 1$ and the customer is in treatment group), or with no DR signal (when $T_i = 0$ and the customer is in control group):

$$Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0). \tag{1}$$

We further use an additive model and express (1) as:

$$Y_i = f(\mathbf{x}_i) + g(\mathbf{x}_i) T_i, \tag{2}$$

where $f(\mathbf{x}_i) = Y_i(0)$, $g(\mathbf{x}_i) = Y_i(1) - Y_i(0)$ and the covariates are denoted by $\mathbf{x}_i$. The covariates $\mathbf{x}_i$ in DR programs are the potential predictors of household consumption data, such as temperature and appliance information. In the following sections, we write $f(\mathbf{x}_i)$ as f_i and $g(\mathbf{x}_i)$ as g_i for simplicity.

The ATE modeled in (2) is the empirical mean of g_i:

$$\bar{g} = \frac{1}{N} \sum_{i=1}^{N} g_i. \tag{3}$$

It measures the average effect induced by the treatment to the whole group. We then call f_i as the main effect for customer i and g_i as the treatment effect for customer i.

2.2. Assumptions

We assume a randomized trial scenario, where $T_i \perp \mathbf{x}_i$, so $T_i \perp \{f_i, g_i\}$. Throughout the paper, $\perp$ denotes statistical independence. This scenario guarantees that the treated group is similar to the control group in each trial, so there is no need to balance the covariates in either groups. Note that this assumption might seem strict in reality, but it is not beyond unreasonable. First, the design of the trial is unknown in an observational study, so any assumptions other than randomized trial is even more stringent. Second, the Pecan Street database that we use has customers in and out of the DR programs continuously so it is reasonably to exert such assumption.

3. LINEAR REGRESSION

As discussed in introduction, estimation can be made in several ways. We choose to emphasize linear models in our analysis because of their ubiquity in both theory and practice [18][19]. In this section, we propose three different ways to obtain linear regression estimators, i.e., through simple linear regression (SLR) on treatment variable, multiple linear regression (MLR) on both the treatment variable and the covariates, and a simple linear regression using the modified covariate method (MCM) introduced in [13].

3.1. Simple Linear Regressing on Treatment

Suppose that $\bar{T} = \frac{1}{N}\sum_i T_i = p$, which means that pN customers are treated in a pool of N customers. It can also be interpreted as the probability that each customer gets a treatment. The two explanations are different, but in the presence of large number of samples they share similar statistical properties [20]. We then rewrite (2) into the following form by centering the variables:

$$\begin{aligned} Y_i &= (T_i - p)\bar{g} + \bar{g} + \bar{f} + T_i(g_i - \bar{g}) + (f_i - \bar{f}) \\ &= Z_i \bar{g} + \alpha_0 + \sigma_i, \end{aligned} \tag{4}$$

where $\bar{g} = \frac{1}{N}\sum_{i=1}^{N} g_i$, $\bar{f} = \frac{1}{N}\sum_{i=1}^{N} f_i$, $Z_i = T_i - p$, $\alpha_0 = \bar{g} + \bar{f}$, and $\sigma_i = T_i(g_i - \bar{g}) + (f_i - \bar{f})$.

The least square estimator is consistent, and the estimation for $\bar{g}$ is given by $(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Y}$. With some manipulation, a simpler expression is:

$$\hat{\bar{g}}_{SLR} = \frac{\sum_i T_i(g_i + f_i)}{\sum_i T_i} - \frac{\sum_i (1 - T_i) f_i}{\sum_i (1 - T_i)}. \tag{5}$$

Note that $Y_i(1) = g_i + f_i$ and $Y_i(0) = f_i$, this indicates that the linear regression estimator is exactly the same as difference-in-mean estimator [16]. It simply takes the difference between the average outcome between the treatment group and the control group. The simulation results of the simple linear regression estimator is presented in section 4.

3.2. Multiple Linear Regression on Treatment and Covariates

Now suppose that we know some covariates $\mathbf{x}_i$ about customer i. A multiple linear regression model is carried out when both the treatment variable T_i and $\mathbf{x}_i$ are included as regressors. The simulation results of the multiple linear regression estimator is presented in section 4. The simulations show that even though multiple linear regression still yield a consistent estimator, it does not always improve over the simple linear regression method (measured via variance of the estimators). This is mainly because that the wrong covariates can be selected into the model and that the underlying model may not be linear.

3.3. Modified Covariate Method

As pointed out in the last section, including covariates directly into the linear model to estimate $\bar{g}$ does not necessarily improve the estimator's performance. Here performance is measured by the variance of the estimator. One possible improvement is to relax the assumptions by only assuming linearity in the treatment effect for each customer i, i.e., only g_i is linear in the covariates.

We thus use a new method proposed in [13]. This method works when we assume that the treatment effect is linear in the covariate, i.e., $g_i = \mathbf{x}_i^{\mathrm{T}}\gamma$, but we do not impose any conditions on f_i. In this case, ATE can be presented as $\bar{g}_{MCM} = \frac{1}{N}\sum_i \mathbf{x}_i^{\mathrm{T}}\gamma$.

We then have the following linear regression model [13]:

$$\begin{aligned} Y_i &= f_i + T_i\mathbf{x}_i^{\mathrm{T}}\gamma \\ &= \mathbf{v}_i^{\mathrm{T}}\gamma + \alpha_0 + \sigma_i, \end{aligned} \tag{6}$$

where $\mathbf{v}_i = (T_i - p)\mathbf{x}_i$, $\alpha_0 = \bar{f} + p\bar{\mathbf{x}}^{\mathrm{T}}\gamma$, and $\sigma_i = (f_i - \bar{f}) + p(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}}\gamma$. We refer to $\mathbf{v}_i$ as the modified covariate. The modified covariate method is carried out when regressing the outcome Y_i on $\mathbf{v}_i$ and an intercept. The estimator is still consistent[13]. The corresponding simulation result is presented in section 4.

4. CASE STUDY

We conducted several experiments in two settings: synthetic data and real-world data from Pecan Street [14]. In both settings, SLR stands for simple linear regression, MLR stands for multiple linear regression and MCM is modified covariate method.

4.1. Synthetic Data

We simulate outcome data from four different models, i.e., a linear model, a non-linear model, a model with constant f_i, and a model with non-linear f_i but linear g_i. The covariates $\mathbf{x}$'s are i.i.d. samples from a joint gaussian distribution. The probability of a treatment assignment is set to be 0.8, 0.9, 0.75 and 0.1 respectively for these four models (model 1 to model 4) presented in (7a) to (7d), where d is the dimension of the covariate vector $\mathbf{x}_i$.

$$Y_i = \mathbf{x}_i^{\mathrm{T}}\gamma + \mathbf{x}_i^{\mathrm{T}}\beta T_i \tag{7a}$$

$$Y_i = \sqrt[4]{|\sum_{j=1}^{d} x_{i,j}^3\gamma_j|} + (\sum_{j=1}^{d} x_{i,j}^2\beta_j + \sum_{j=1,k\neq j}^{d} x_{i,j}x_{i,k}\mu_{j,k})T_i \tag{7b}$$

$$Y_i = \alpha_0 + (\sum_{j=1}^{d} x_{i,j}^2\beta_j + \sum_{j=1}^{d}\sum_{k\neq j}^{d} x_{i,j}x_{i,k}\mu_{j,k})T_i \tag{7c}$$

$$Y_i = \sqrt[4]{|\sum_{j=1}^{d} x_{i,j}^3\gamma_j|} + \mathbf{x}_i^{\mathrm{T}}\beta T_i \tag{7d}$$

The variance of the three estimators presented in section 3, i.e., SLR, MLR and MCM is shown in Fig.1. The ϵ^2 shown in Fig.1 is the empirical variance of the estimator and we compare the performance of the estimators based on the magnitude of this variance.

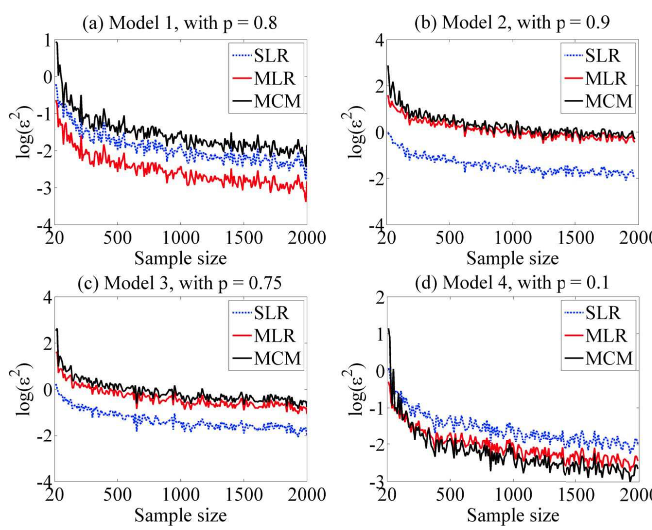

Fig. 1. Variance of the three estimators of synthetic data. SLR performs best in (b) and (c), MLR in (a) and MCM in (d).

From Fig.1, we observe that it is not always the case that MLR yields a better estimator. As seen from Fig.1(a), when the model is linear in the covariates, MLR has the best performance with respect to variance reduction. However, if the model is not linear then MLR does not necessarily reduce the variance of the estimator, as shown in Fig.1(b) through

Fig.1(d). Comparing results from Fig.1(b) and Fig.1(c), we observe that SLR has the lowest estimator variance in both cases, when neither f_i nor g_i is linear in the covariates. Thus we argue that performing SLR is the safest way to yield an estimator with least variance. What is more, if the treatment effect is linear in the covariates and the probability of treatment assignment is small, then MCM outperforms both SLR and MLR, as shown in Fig.1(d). It thus serves as a compromise to use covariate information while keeping the estimator's variance low.

4.2. Pecan Street Data

In this section, we test the estimators on data from Pecan Street [14]. In the tests, we treat the high price signals as treatments. The outcomes are customers' consumption data. To compose the treatment group and control group, we extract the high price signal and include customers whose consumption data is available at that time into the treatment group, and include the other customers into the control group. For the control group, we find their consumption data at the same hour in the date closest to the high price signal date. This mimics the situation where the signals are randomized assigned, since each customer has some chance of receiving a specific signal. Temperature is a primary regressor that researchers use in practice, so we include temperature into the linear regression model as covariate. Other covariates suh as appliance information can easily be added.

Since we do not know the true ATE and the true model for observational data, we use p-values associated with the t-test and the F-test to make comparisons in Pecan Street data. These tests are hypothesis tests for linear regression models with Gaussian noise [21]. The difference between t-test and F-test is that t-test only examines whether including one particular regressor significantly improve the model whereas the F-test examines whether including all regressors significantly improve the model. Suppose that we just include one covariate into the regression model, the MLR model is in the form of $Y_i = Z_i\bar{g} + x_i\beta + \alpha_0 + \delta_i$, where $\bar{g}$ is the ATE to a specific DR signal, $Z_i = T_i - p$ is the centered binary indicator variable for DR signal and x_i is the covariate. The null hypothesis for the t-test in this regression model is given as $H_0 : \bar{g} = 0$ and for the F-test: $H_0 : \bar{g} = \beta = 0$.

For both tests, we examine the significance by setting a confidence level α, normally taken as 0.05. While comparing the values of a certain statistic under different models does not seem intuitive, we can alternatively resort to p-value, which is defined as the probability of obtaining the observed(or more extreme) result under the null hypothesis. Higher p-values suggest that the null hypothesis is true, whereas smaller p-values suggest the opposite. We then can compare the p-value to interpret the significance test under different regression models.

The results are shown in Table 1. The treatment group has 3911 observations and the control group has 893 observations. All estimators are consistent. What is more, from the results in Table 1, we can see that the p-value with the F-test for all methods is generally small, meaning that the consumption data cannot be explained by just an intercept. However, the p-value associated with the t-test is high for MLR, suggesting the insignificance of regressing on the treatment variable. This is mainly due to the lack of information on how the covariates interact with consumption data and that the treatment group is much bigger than the control group. If an utility uses MLR to estimate the ATE, it may conclude that the DR program is ineffectual by mistake. Therefore although including covariates into the model may seem to improve prediction (smaller p-value for the F-test), it does not necessarily lead to a better inference.

From Table 1 we can also see that the p-value for t-test with MCM is small, suggesting that we should regress on the modified covariate. Therefore if we know that the treatment effect is linear in the covariates, then using modified covariate method can on the one hand exploit the covariate information and on the other hand reduce the variance of the estimator. Other interaction terms can be added in similar fashion.

Table 1. Results for pecan street data. Smaller p-values sugsgest significance of the variable(s).

	p-value for t test	**p-value for F test**
SLR	2e-16	2e-16
MLR	0.133	2e-16
MCM	2e-16	2e-16

5. CONCLUSION

In this paper, we estimate the average treatment effect of demand response programs. We adopt a regression model and assume that the DR signal is randomly assigned to customers. We derive a simple linear regression estimator on just the treatment, multiple linear regression estimator on treatment and other covariates, and a modified covariate method. The simulation results show that although including more information may be good for prediction purposes, simple linear regression estimators may have lower variances. In the case where the treatment effect is linear in the covariates, the modified covariate method is able to leverage additional covariate information to reduce estimation error. Thus, the interactions between the covariates and the demand response signal must be carefully modeled. This work can provide a framework for further research in applying causal inference in analyzing consumption data and DR interventions.

6. REFERENCES

[1] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 3, pp. 381–388, 2011.

[2] C. L. Su and D. Kirschen, "Quantifying the effect of demand response on electricity markets," *Power Systems, IEEE Transactions on*, vol. 24, no. 3, pp. 1199–1207, 2009.

[3] D. Wang, X. Guan, J. Wu, P. Li, P. Zan, and Hui. Xu, "Integrated energy exchange scheduling for multimicrogrid system with electric vehicles," *Smart Grid, IEEE Transactions on, preprint*, 2015.

[4] M. H. Albadi and E. F. El-Saadany, "Demand response in electricity markets: An overview," in *IEEE Power and Energy Society General Meeting*. 2007, IEEE.

[5] M. H. Albadi and E. F. El-Saadany, "A summary of demand response in electricity markets," *Electric power systems research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[6] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *IEEE Power and Energy Society General Meeting*. 2011, IEEE.

[7] P. Siano, "Demand response and smart grids:a survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[8] L. Qian, Y. A. Zhang, J. Huang, and Y. Wu, "Demand response management via real-time electricity price control in smart grids," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 7, pp. 1268–1280, 2013.

[9] W. Saad, Z. Han, H. V. Poor, and T. Başar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 86–105, 2012.

[10] C. K. Woo and K. Herter, "Residential demand response evaluation: a scoping study," *Energy and Environmental Economics, Inc. and Lawrence Berkeley National Laboratory*, 2006.

[11] P. M. Aronow and J. A. Middleton, "A class of unbiased estimators of the average treatment effect in randomized experiments," *Journal of Causal Inference*, vol. 1, no. 1, pp. 135–154, 2013.

[12] NewYorkISO, "Caiso demand response resource user guide," Tech. Rep., NewYorkISO.

[13] L. Tian, A. Alizadeh, A. Gentles, and R. Tibshirani, "A simple method for detecting interactions between a treatment and a large number of covariates," *Journal of the American Statistical Association*, vol. 109, no. 508, pp. 1517–1532, 2014.

[14] Pecan Street, "Pecan street online database," `http://www.pecanstreet.org`, 2016.

[15] D. A. Freedman, *Statistical models: theory and practice*, Cambridge University Press, 2009.

[16] E. A. Stuart, "Matching methods for causal inference: A review and a look forward," *Statistical science: a review journal of the Institute of Mathematical Statistics*, vol. 25, no. 1, pp. 1–29, 2010.

[17] G. W. Imbens, "Nonparametric estimation of average treatment effects under exogeneity: A review," *Review of Economics and statistics*, vol. 86, no. 1, pp. 4–29, 2004.

[18] M. H. Amini, O. Karabasoglu, M. D. Ilic, K.G. Boroojeni, and S. S. Iyengar, "Arima-based demand forecasting method considering probabilistic model of electric vehicles' parking lots," in *IEEE Power and Energy Society General Meeting*. 2015, IEEE.

[19] S. Chan, K. M. Tsui, H. C. Wu, Y. Hou, Y. C. Wu, and F. Wu, "Load/price forecasting and managing demand response for smart grid: Methodologies and challenges," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 68–85, 2012.

[20] D. A. Freedman, "On regression adjustments to experimental data," *Advances in Applied Mathematics*, vol. 40, no. 2, pp. 180–193, 2008.

[21] G. A. F. Seber and A. J. Lee, *Linear regression analysis*, John Wiley and Sons, 2012.