Time Will Tell: Using Smart Meter Time Series Data to Derive Household Features and Explain Heterogeneity in Pricing Programs

Siddharth Patel, Sam Borgeson, Ram Rajagopal, Stanford University
C. Anna Spurlock, Ling Jin, and Annika Todd, Lawrence Berkeley National Laboratory

ABSTRACT

The recent nationwide adoption of smart meters provides a new source of rich data about individual household electricity consumption. Data science techniques can extract a variety of high temporal resolution, household-specific features from the hourly electricity time series itself and in combination with other readily available relevant information, like weather or census data. This allows us to observe or estimate important characteristics of household electricity use that were previously unobservable. These characteristics and household features have the advantage of representing the actual choices and behaviors of households, which can differ substantially from stated preferences and subjective information from traditional survey or interview methods. The use of data-derived numerical features is common in machine learning but not in traditional engineering and econometric models.

In this paper, we use this technique to help answer a question that is important to program implementers: who is likely to respond to my program? In other words, program implementers typically use rules of thumb to identify target households (e.g., top 25% of usage from monthly bills), and program evaluation typically only identifies the overall average effect. Understanding the heterogeneity of program response can help shed light on how and why different households respond in different ways, allowing implementers to focus on specific groups, tailor programs to speak to the way that households actually behave, and predict the effectiveness of future programs for portfolio planning purposes.

We identify household specific features that explain heterogeneity in response to experimental time-of-use and critical-peak-pricing electricity rates. The experiment was performed using randomized controlled trials with treatment groups encouraged to enroll into new rates. The household responses of interest are metrics related to energy consumption during peak hours. We use numerical features derived from pre-treatment smart meter data as covariates in an instrumental variable regression, and we find, promisingly, that they can explain considerable heterogeneity of treatment outcomes. These results lay the groundwork for using smart meter data along with data science techniques to improve program uptake, evaluation, design, and targeting.

Introduction

A primary goal of demand side management programs is to provide grid resources (i.e. avoided consumption) with costs lower than supply-side alternatives. The value of demand side resources are calculated by comparing program implementation costs to their benefits. Thus, techniques that can either lower implementation costs through more effective planning and recruitment or improve savings through the strategic targeting of interventions can significantly boost the value of demand side resources. One strategy for doing so is to develop models that use customer characteristics to predict enrollment in and savings under specific programs. Such models can improve program design in three important ways. First, programs can focus recruitment efforts on households that are likely to accept program offers or predicted to provide better than average benefits to the program. This improved targeting would drive down customer acquisition costs and improve program savings. Second, programs can tailor messages and framing using household characteristics to improve the salience of offers made to customers. Third, program outcomes can be predicted more accurately on an hour-, week-, or year-ahead basis. This information can serve grid planners and aid in the design of future programs.

The challenge is that these household characteristics may be difficult to observe, may only be observable indirectly, and typically only represent one snapshot in time rather than information on an ongoing basis. Some characteristics, like appliance inventories or demographics, can be observed by administering questionnaires, but this is a relatively laborious process. Household usage on a monthly basis is much less useful for DR and pricing programs than usage on an hourly basis. Through the application of data science techniques to smart meter data, household characteristics and their effects on consumption can be estimated before a program begins. Figure 1 illustrates the relationship between pre-treatment household meter data and estimated features of consumption on its left hand side.

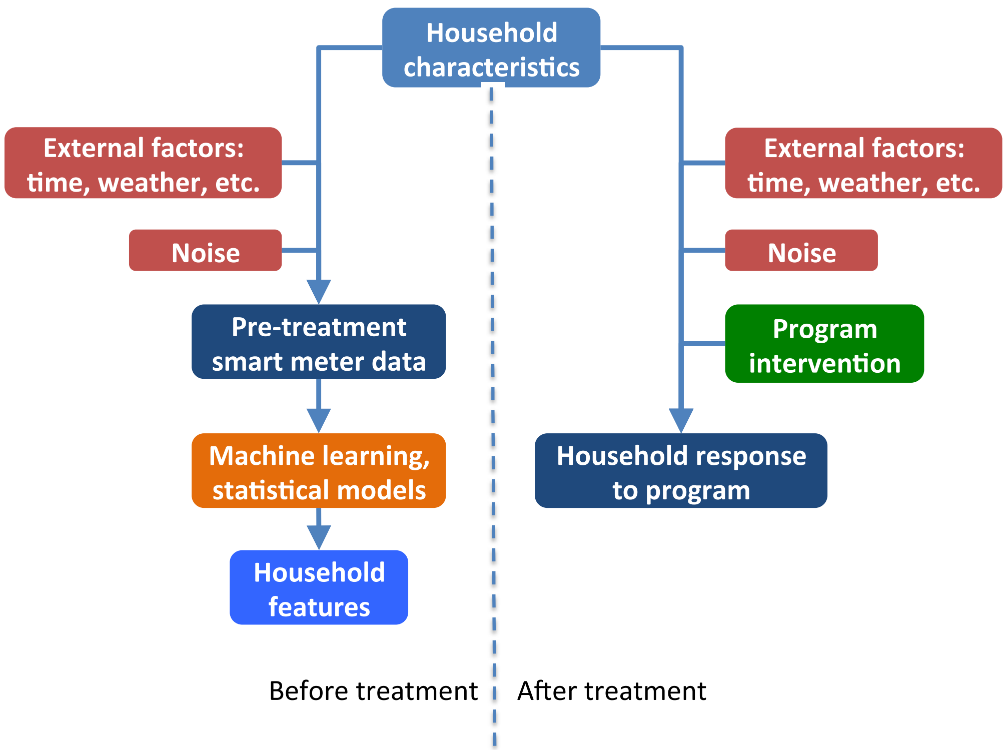

Figure 1. Relationship between household characteristics, features, and program response. The smart meter data in the pre-treatment period is influenced by household characteristics, external factors, and random variation. Machine learning and statistical models can extract features from this smart meter data that contain information about household characteristics. Those household characteristics also influence the response to demand side management programs, so the household features can be used to explain heterogeneity in household responses to programs.

The pre-treatment smart meter time series for a household is a rich data source, spanning working days and holidays, seasonal changes in weather, and variations in other conditions. Statistical and machine learning techniques can be applied to this large dataset to extract a variety of features related to household characteristics (Borgeson et al. 2015). For example, the average daily minimum consumption contains information about the magnitude of a household's always-on load. Change point regression and load disaggregation models can estimate the temperature sensitivity of a household's electricity consumption, which provides insight into the extent to which the household uses electricity for cooling (Borgeson 2013). Customer daily load profiles can be clustered into representative load shapes through optimized pattern recognition with statistical learning algorithms (Kwac 2014), which leads to segmentation of customers based on their daily consumption schedules. There are many such features that summarize key aspects of how a household consumes electricity.

The goal of this work is to use smart meter data and econometric techniques to identify which features, statistically derived from consumption data, have bearing on how households respond to demand side management programs. Some of these features may be related to the capability of a household to respond – for example, households with larger consumption may provide a larger absolute response. Other features may be related to the willingness or ease with which homes can respond – perhaps homes with a higher degree of variability are more able to adjust their schedules to shift load away from peak hours. Finally, features can be related to incentives, identifying households that stand to save the most money by altering their consumption in the manner desired by the program. This information could improve program uptake, improve program impact through directed targeting, and lead to better estimates of program potential.

This paper develops a methodology for using features to shed light on heterogeneity in household responses to programs. Specifically, we incorporate features into econometric models in order to estimate the causal impact the features have on household response. The method is applied to data from a large randomized controlled trial for time-of-use (TOU) and critical-peak-pricing (CPP) programs. We find that household features can explain considerable heterogeneity between different types of households. This finding lays the groundwork for improved program design, uptake, and evaluation.

Background

The incorporation of greater shares of renewably generated electricity will require a more flexible electric grid (Mai et al. 2012). Demand side management has an important role to play in providing that flexibility (Strbac 2008; van Renssen 2014; Lund et al. 2015). A national assessment of demand response potential in the United States found that the residential sector has the greatest capacity for peak load reduction but is the least tapped sector (FERC 2009). There are a few key reasons for this gap between potential and actuality. First, the consumption of a typical household is small when compared to an industrial or commercial facility. Thus, a large number of households must be treated to provide a meaningful amount of flexibility, imposing additional requirements for communication, measurement, and administration. Another hurdle is that the electrical consumption of a single household is highly uncertain, which makes an individual household a less dependable resource (Sevlian and Rajagopal 2014). Thus, there is a need to develop and test methods for estimating how much response can be elicited from a large group of households and in what ways. Estimating program potential is a difficult task – FERC's assessment of demand response reported major reductions in demand response potential in two regions due to lower than expected program performance (FERC 2015).

Time-varying rates have gained acceptance as a method for eliciting residential flexibility (Cappers, Goldman, and Kathan 2010), and there have been a variety of experimental studies conducted to evaluate how well these rates work (Faruqui and Sergici 2010). These experiments, when conducted with a control group, typically focus on estimating the average treatment effect on treated groups. However, this approach cannot differentiate between treated households with relatively weak or strong responses. The ability to identify such households is of great value. Program targeting, and therefore costs, as well as overall program effectiveness could be greatly improved with the ability to identify households expected to respond well to a particular type of program. There is a developing literature related to explaining the *heterogeneity* of individual responses in a randomized controlled trial (Imai and Ratkovic 2013; Athey and Imbens 2015).

A few of the studies reviewed in Faruqui and Sergici 2010 attempt to explain which households responded better. They rely on demographic data obtained from surveys instead of metered consumption. This paper proposes using features derived from household smart meter data to estimate the heterogeneity of household responses to pricing programs. This approach has two advantages – smart meter data is already readily available throughout much of the U.S., whereas collecting survey and demographic data is a more laborious and costly process. Furthermore, survey data relies on what people state about their preferences related to electricity consumption, whereas smart meter consumption history reveals how occupants actually chose to consume.

The remainder of the paper is organized as follows. The Data section explains the pricing program experiment, the data available, and the household responses of interest. The Model section describes the instrumental variable regression model used to test the causal effect of the pricing treatment heterogeneously across households based on features. The Results and Discussion section discusses three specific hypotheses about which features can explain heterogeneity and demonstrates their validity in the data. Finally, the Conclusions section summarizes key findings and points out directions for extending this work.

Data

The data for this paper comes from an experiment conducted by an electric utility covering over 100,000 households. The experiment was one type of randomized controlled trial, called a randomized encouragement design ("RED"; LBNL and EPRI 2013), in which households are randomly assigned to either a treatment group, where they are encouraged to enroll in the program, or the control group, in which households are not contacted. Household smart meter data is available from before the treatment and during the treatment. The treatment began on June 1, 2012. Households were treated using either a TOU plan or a CPP plan. Both plans were structured to make consumption during peak hours, defined as 4pm to 7pm, more expensive.

The TOU rates applied the increased peak hour costs on all business days, while the CPP rates applied the increased costs on just a dozen critical event days per summer, which were communicated to the households a day in advance. Table 1 describes the treatment groups and the days analyzed for those groups. Households assigned to an opt-in treatment group had to actively opt-in to receive the treatment rate plan. Households assigned to an opt-out treatment group were defaulted into the treatment and had to actively opt-out to avoid receiving the treatment rate plan.

Table 1. Description of treatment groups and days analyzed for these groups.

Treatment group	Period used to compute household features	Days analyzed for computing household response
Opt-in TOU	6/1/2011 – 5/31/2012	Business days in summers of 2011, 2012, 2013
Opt-out TOU	6/1/2011 – 5/31/2012	Business days in summers of 2011, 2012, 2013
Opt-in CPP	6/1/2011 – 5/31/2012	Critical event-like[1] days in summer 2011, Critical event days in summers of 2012 and 2013
Opt-out CPP	6/1/2011 – 5/31/2012	Critical event-like[1] days in summer 2011, Critical event days in summers of 2012 and 2013

1. There were no actual critical event days in summer 2011 because it was prior to the start of the program, so summer business days with weather similar to the critical event days of 2012 and 2013 were chosen.

The household features are computed using the year's worth of pre-treatment smart meter data. For example, the mean consumption of a given household over the year would be a feature for that household. The median of the daily minimum load over the year would be another feature. Household features will be denoted by x_i in this paper.

The goal of the treatment rate plans is to get households to reduce their consumption during peak hours. To quantify the effectiveness of the program, a number of different measures of household electricity consumption were considered. For each household (i), for each day (t), the following metrics were computed:

- e_{it} = total electrical energy consumed during the day (kWh)
- p_{it} = total electrical energy consumed during peak hours (kWh)
- m_{it} = maximum hourly electrical energy consumption during peak hours (kWh)
- f_{it} = p_{it} / e_{it}, fraction of electrical energy consumed during peak hours

Only f_{it} is a relative quantity; the other three are absolute quantities. All three metrics are computed for each day and each household directly from smart meter data from the pre-treatment summer (2011) and the two treatment summers (2012 and 2013). This contrasts with the household features, which are derived for each household solely from the pre-treatment smart meter data. A feature is a household-specific quantity that is known prior to the beginning of the treatment. A metric is a measure of how a specific household consumed electrical energy on a specific day, either before or during the treatment period.

These particular metrics capture aspects of consumption that TOU and CPP programs try to influence. In particular, p_{it} and m_{it} are directly related to the primary goal of reducing electricity usage during peak hours. We may expect to see decreases in f_{it} if households shift consumption away from peak hours and towards off-peak hours, and f_{it} captures this for both large and small households because it is a normalized quantity.

For a given metric, the response of a treated household is how much that metric changed from the pre-treatment period to the treatment period, relative to the same difference in the control group. For example, when considering p_{it}, the response of a treatment household is how much its energy usage during peak hours on a typical day changed between the pre-treatment and treatment periods, beyond the change observed in control group households. This is a difference-in-difference approach, explained in more detail in the next section. The goal of this work is to demonstrate and quantify meaningful relationships between household features, which are known *a priori*, and household responses to treatment as measured by changes in the metrics.

Model

Let y_{it} be any of the metrics described in the prior section (e.g., f_{it} or p_{it}). Let x_i be a numerical feature for household i (e.g., mean of consumption from pre-treatment smart meter data), and let $X_i = 1$ if x_i is greater than the median value for all households, and $X_i = 0$ otherwise. Using the high/low indicator X_i in the subsequent models greatly simplifies their interpretation. To determine whether a given feature can explain heterogeneity in how enrolled households responded to the pricing program, which was a randomized encouragement design, the following regression specification is employed.

To account for any selection bias in the treatment group, i.e. to correct for the fact that enrolled customers are likely to be more enthusiastic responders to the treatment than randomly selected customers, the regression model is estimated using two-stage least squares (Imbens and Wooldridge 2009; Cappers et al. 2014). In the first stage, the treatment indicator T_{it} is estimated by A_{it}, where A_{it} is an indicator variable for whether household i was encouraged to be in treatment on day t, producing $\hat{T}_{it}$. Similarly, the interaction term between the treatment and the feature indicator, $X_i T_{it}$, is estimated by $X_i A_{it}$, producing $\overline{X_i T_{it}}$. In the second stage, using $\hat{T}_{it}$ and $\overline{X_i T_{it}}$, the heterogeneity of the treatment effect on the treated is estimated as ϕ in the following specification:

$$y_{it} = \alpha + \delta(X_i P_t) + \zeta \hat{T}_{it} + \phi(\overline{X_i T_{it}}) + \gamma_i + \tau_t + \varepsilon_{it}$$

y_{it} is the impact metric of interest (e.g., peak hour electricity usage)
X_i is an indicator variable for whether household i has a high value of feature x_i
P_t is an indicator variable for whether day t is before or during the treatment period
$\hat{T}_{it}$ is an indicator variable for whether household i is in treatment on day t, estimated in the first stage
$\overline{X_i T_{it}}$ is the interaction term between X_i and T_{it}, estimated in the first stage
γ_i is a fixed effect for household i
τ_t is a fixed effect for day t
ε_{it} is an error term assumed to be i.i.d. normal, conditional on the covariates, clustered at the household level.

This regression produces an unbiased estimate of ζ, the treatment effect on y_{it} for treatment households with low values of x_i; and of ϕ, the additive treatment effect of having a high value of x_i on the change in y_{it} for treatment households. Thus, the total treatment effect for households with high values of x_i (i.e., $X_i = 1$) is $\zeta + \phi$, where ϕ indicates how much greater the effect was for households with high values of x_i. The model also produces an unbiased estimate of δ, which is the effect of X_i on the change in the metric y_{it} for control households between the pre-treatment and treatment periods. A statistically significant value of our coefficient of interest, ϕ, is an indication that x_i can explain heterogeneity in treatment household responses to the program in terms of the given metric y_{it}.

Results and Discussion

With this model, hypotheses about relationships between household features and program responses can be tested. While the questions of greatest interest will vary between programs and program planners, the next three subsections explain particular hypotheses and report the results of testing them in the data.

Hypothesis 1

The first hypothesis tested is that households with a higher mean consumption during the pre-treatment summer (x_i) would exhibit a greater response in terms of reduction of electrical energy consumption during peak hours (p_{it}). The basic idea is that households that consume more have a greater capacity to reduce peak hour consumption in response to the pricing program. Note that this is not an obvious relationship. It could be the case, for example, that households that consume more are also wealthier, with electricity comprising a smaller part of their budget, so they may be unlikely to respond much to a pricing program.

The hypothesis that greater pre-treatment mean consumption corresponds to greater reductions in energy during peak hours was found to be valid for all treatment groups. That is, ϕ is negative and statistically significant, indicating that treatment households with higher mean consumption reduced p_{it} more, as shown in Table 2. Thus, the intuition that larger households should respond more is confirmed in the data. Furthermore, the *magnitude* of the relationship is quantified.

Table 2. Estimates of the effect of greater household consumption on heterogeneity in absolute responses to treatment.

Metric	p_{it}			
	x_i = mean consumption during summer			
Group	ζ kWh	Sig.[1]	ϕ kWh	Sig.[1]
Opt-in TOU	-.54	***	-.77	***
Opt-out TOU	-.18	***	-.33	***
Opt-in CPP	-1.2	***	-1.5	***
Opt-out CPP	-.57	***	-.99	***

1. Sig. is a symbolic representation of the statistical significance of β. Let s denote the probability under the null hypothesis that β attains a magnitude at least as large as that estimated by the model. Then * indicates $s < 0.05$; ** indicates $s < 0.01$; *** indicates $s < 0.001$; a dash in both the β and Sig. columns indicates that $s \geq 0.05$.

For all treatment groups, $\phi > \zeta$, meaning that the additional treatment effect on households with high pre-treatment mean consumption was greater than the entire treatment effect on households with low pre-treatment mean consumption. In other words, households with higher mean consumption exhibited, on average, twice the reduction in consumption during peak hours as did households with low mean consumption. For example, on summer business days, in the Opt-in TOU group, treatment households with low values of x_i reduced their consumption during peak hours by 0.54 kWh on average, whereas treatment households with high values of x_i reduced by 1.31 kWh on average. Thus, if a program planner considering TOU and CPP plans were primarily concerned with reducing energy consumption during peak hours, they would do well to target households with higher mean consumption.

Hypothesis 2

The second hypothesis is that households with higher peak consumption during the pre-treatment summer (x_i) would exhibit a greater response in terms of reduction of maximum hourly consumption during peak hours (m_{it}). Specifically, x_i is the 97th percentile of the daily maximum consumption in the pre-treatment smart meter data for household *i*. The basis for this hypothesis is similar in nature to that for Hypothesis 1 – households with greater peaks during the pre-treatment period have a greater capacity to reduce their peaks in response to the pricing program. The hypothesis was found to be valid for all treatment groups – ϕ is negative and statistically significant, as shown in Table 3.

Table 3. Estimates of the effect of greater household consumption on heterogeneity in absolute responses to treatment.

Metric	m_{it}			
	x_i = 97th percentile of daily maximum hourly consumption			
Group	ζ kWh	Sig.[1]	ϕ kWh	Sig.[1]
Opt-in TOU	-.23	***	-.28	***
Opt-out TOU	-.08	***	-.12	***
Opt-in CPP	-.46	***	-.59	***
Opt-out CPP	-.23	***	-.35	***

1. This table uses the same symbolic convention for statistical significance as Table 2.

Once again, for all treatment groups, $\phi > \zeta$, meaning that treatment households with larger peaks in their pre-treatment data exhibited, on average, twice the reduction in daily peaks as did treatment households with smaller peaks in their pre-treatment data. Managing peak load is of critical importance for system operators and utilities. Thus, it is noteworthy that among the Opt-in CPP group, on critical event days, treatment households with high values of x_i reduced their maximum hourly consumption during peak hours by over 1 kWh, on average, whereas those with low values of x_i reduced by less than 0.5 kWh. A program planner tasked with enrolling households in a pricing program to curtail maximum hourly load during peak hours could use this data to justify targeting households with higher daily peaks.

Hypothesis 3

The last hypothesis considered is that households that consumed a higher fraction of their electrical energy during peak hours in the pre-treatment period (x_i) would shift a greater portion of their consumption away from peak hours in response to the pricing program (i.e. would have a greater reduction in f_{it}). Note that x_i and f_{it} measure relative quantities, so differences in the scale of consumption between different households are leveled out.

This hypothesis was confirmed in the data: ϕ is negative, indicating a stronger treatment effect for those who consumed a greater fraction of their energy during peak hours in the pre-treatment period. However, the effect sizes and statistical significance are weak for some groups. The additional treatment effect here was not as large as it was for Hypotheses 1 and 2, and it was only strongly statistically significant for households on TOU rates. This means that a program planner primarily concerned with reducing f_{it} should expect relatively small improvements from targeting households with higher fractions of electrical energy consumed during peak hours in pre-treatment smart meter data. There may be other more effective targeting criteria for obtaining reductions in this particular metric.

Table 4. Estimates of the effect of greater household fraction of consumption during peak hours on heterogeneity in relative response to treatment.

Metric	f_{it}			
	x_i = fraction of consumption during peak hours			
Group	ζ	Sig.[1]	ϕ	Sig.[1]
Opt-in TOU	-.021	***	-.012	***
Opt-out TOU	-.005	***	-.006	***
Opt-in CPP	-.035	***	-.014	**
Opt-out CPP	-.017	***	-.009	*

1. This table uses the same symbolic convention for statistical significance as Table 2.

Conclusion

The main finding of this work is that household features derived using statistical techniques on pre-treatment smart meter data can explain a considerable amount of heterogeneity in treatment outcomes between different types of households. For example, we found that when considering electricity consumption during peak hours as the program metric, the effect of treatment on households with high pre-treatment mean and maximum consumption was over double that of households with low mean and maximum consumption. The intent of this work is to provide proof of the existence of such relationships and to demonstrate a principled method for using pre-treatment smart meter data to analyze program outcomes and identify households that responded most strongly. These results open the door to further exploration for other significant relationships between household features and heterogeneity in outcomes among treated households. This kind of modeling and analysis is of practical use for program planners seeking to improve overall program efficacy through better targeting, which reduces customer acquisition costs and increases program benefit per customer enrolled, and through more consistent and refined program evaluation.

A key extension of this work is to develop hypotheses about *how* consumers responded to the treatments and then to test those hypotheses in the data. This would allow for an interpretable, data-driven segmentation of the treatment households that could provide insight into which households responded best and how they did so. The second extension is the development of predictive models that use pre-treatment smart meter data to better predict which households will respond the best to a particular treatment.

References

Athey, S. and G. Imbens. 2015. “Machine learning methods for estimating heterogeneous causal effects.” Unpublished manuscript, last modified December 30, 2015. arXiv preprint arXiv:1504.01132.

Borgeson, S.D. 2013. “Targeted Efficiency: Using Customer Meter Data to Improve Efficiency Program Outcomes.” Dissertation, Berkeley, CA: UC Berkeley.

Borgeson, S., J.A. Flora, J. Kwac, C.W. Tan, and R. Rajagopal. 2015. “Learning from Hourly Household Energy Consumption: Extracting, Visualizing and Interpreting Household Smart Meter Data.” In *Design, User Experience, and Usability: Interactive Experience Design*. Springer International Publishing: 337-354.

Cappers, P., C. Goldman, D. Kathan. 2010. “Demand response in US electricity markets: Empirical evidence.” *Energy* 35 (4): 1526-1535.

Cappers, P., A. Todd, M. Perry, B. Neenan, and R. Boisvert. 2014. *Quantifying the Impacts of Time-based Rates, Enabling Technology, and Other Treatments in Consumer Behavior Studies: Protocols and Guidelines*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Faruqui, A. and S. Sergici. 2010. “Household response to dynamic pricing of electricity: a survey of 15 experiments.” *Journal of Regulatory Economics* 38 (2): 193-225.

FERC (Federal Energy Regulatory Commission). 2009. *A National Assessment of Demand Response Potential*. Prepared by The Brattle Group, Freeman Sullivan, & Co, and Global Energy Partners.

FERC (Federal Energy Regulatory Commission). 2015. *2015 Assessment of Demand Response and Smart Metering*.

Imai, K. and M. Ratkovic. 2013. "Estimating treatment effect heterogeneity in randomized program evaluation." *The Annals of Applied Statistics* 7 (1): 443-470.

Imbens, G. M., and J. M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5-86.

Kwac, J., J. Flora, and R. Rajagopal. 2014. "Household Energy Consumption Segmentation Using Hourly Data." *Smart Grid, IEEE Transactions on* 5 (1): 420-430.

LBNL (Lawrence Berkeley National Laboratory) and EPRI (Electric Power Research Institute, Inc.). 2013. *Quantifying the Impacts of Consumer Behavioral Study Experiments and Pilots: Protocols and Guidelines*. Berkeley, CA: Lawrence Berkeley National Laboratory.

Lund, P., J. Lindgren, J. Mikkola, and J. Salpakari. 2015. "Review of energy system flexibility measures to enable high levels of variable renewable electricity." *Renewable and Sustainable Energy Reviews* 45 (2015): 785-807.

Mai, T., D. Sandor, R. Wiser, and T. Schneider. 2012. *Renewable Electricity Futures Study: Executive Summary*. Golden, CO: National Renewable Energy Laboratory.

Sevlian, R.A. and R. Rajagopal. 2014. "A model for the effect of aggregation on short term load forecasting." *PES General Meeting| Conference & Exposition, 2014 IEEE*. National Harbor, MD (2014): 1-5.

Strbac, G. (2008). "Demand side management: Benefits and challenges." *Energy Policy* 36 (12): 4419-4426.

van Renssen, S. 2014. "People power to the rescue." *Nature Climate Change* 4 (6): 417-419.