Comparison of Methods for Estimating Energy Savings from Home Energy Reports

Brian Arthur Smith, Pacific Gas and Electric Company

Josh Schellenberg, Nexant

October 6, 2015

Key Research Question: How Do Estimates from Randomized Control Trials Compare to those Derived from Other Methods?

Current Situation

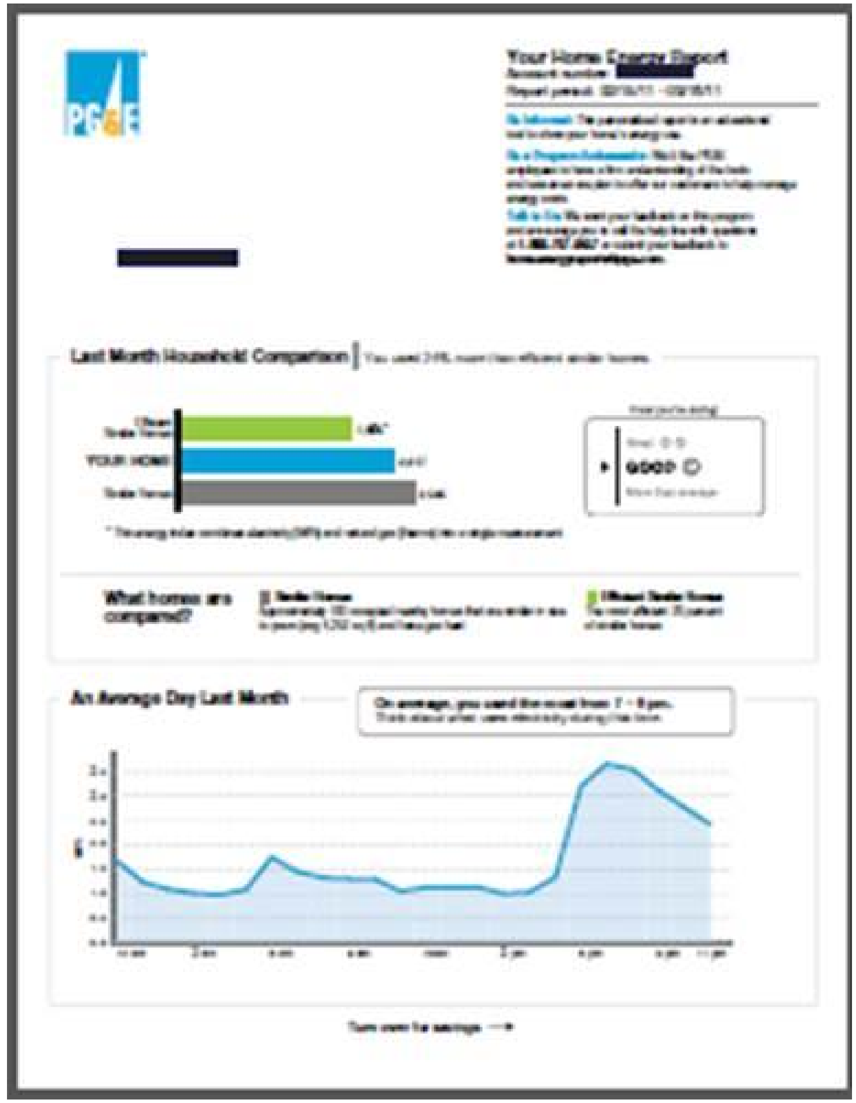
- Home energy reports (HERs) have gained **significant traction** in the utility industry
- Preferred evaluation method is the randomized controlled trial (RCT), but large control groups of non-participants **limit the potential for behavioral programs**
- Alternative statistical methods show **considerable promise** for behavioral program evaluation, but their statistical validity relative to the RCT has yet to be tested rigorously

Study Objective

- **Leverage data** from PG&E's large program
- **Test** promising alternative methods:
  - Bayesian Structured Time Series (BSTS)
  - Regression Tree with Random Effects (RE-EM Tree)
  - Propensity Score Matching (PSM)
- **Compare** the results of a large, multi-year RCT evaluations to the energy savings estimates produced by alternative methods

Overview of PG&E's Home Energy Reports Program

- Launched in 2011 with two experiments
- <u>Expanded steadily since then</u>
- **<u>Currently 1.2 million+</u>** residential customers receiving HERs in a dozen unique experiments
- HERs account for **<u>majority of savings</u>** in PG&E's residential EE portfolio
- **<u>In this analysis</u>**:
  - Approximately 75,000 treatment and 75,000 control customers
  - Participants in one of the first HER RCTs at PG&E (the "Gamma" wave)
  - Analyzed over the course of three post-treatment years (2012-2014)



Savings Estimates Resulting from the RCT Approach at PG&E

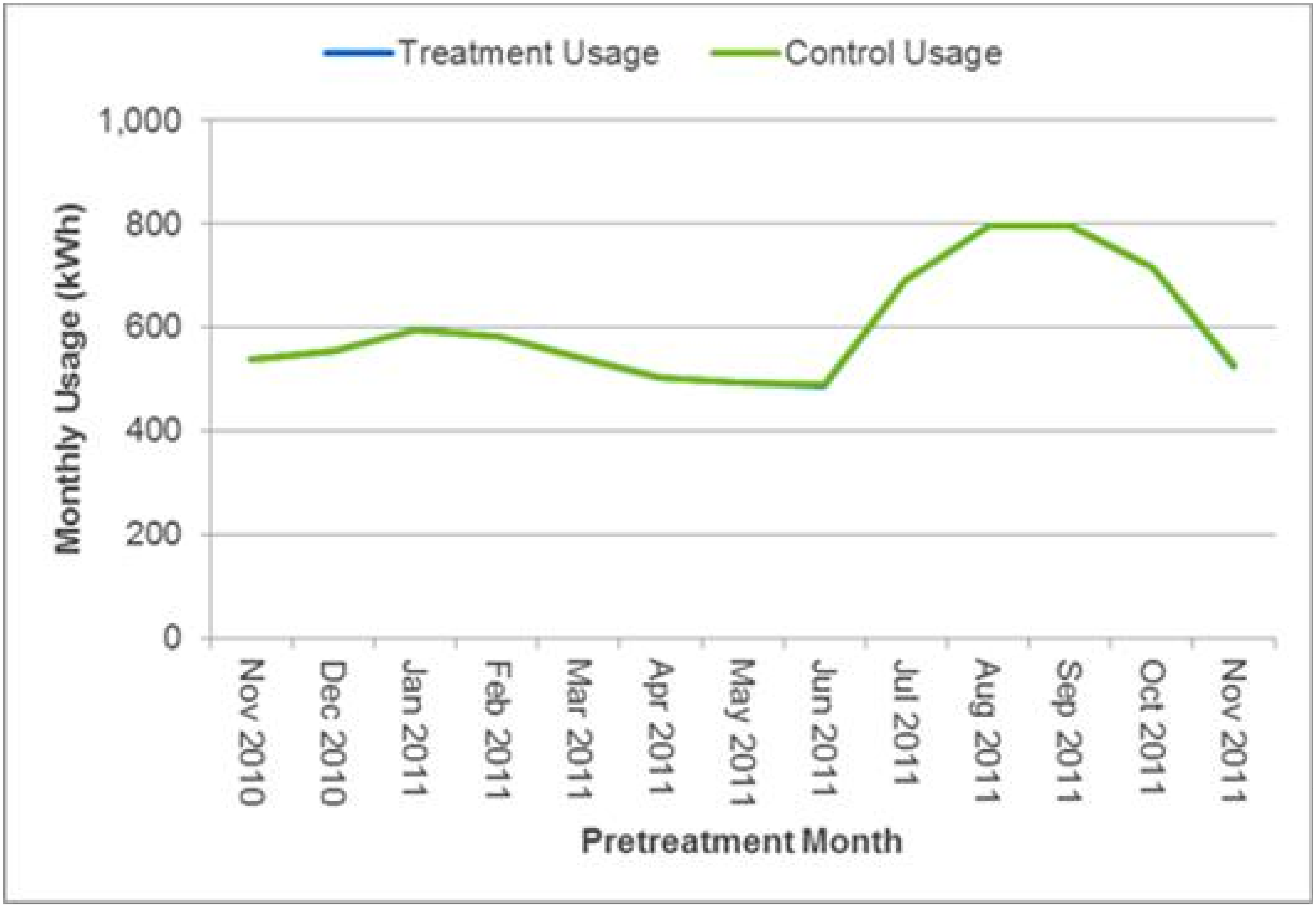
Estimation Methodology	Percent Savings Estimates			Monthly Savings Estimates (kWh)	
	2012	2013	2014	Low	High
Randomized Controlled Trial (RCT)	1.16%	1.58%	1.69%	5.9	12.2

- PG&E's HER program lends itself well to an RCT because it is an **opt-out design** for which random assignment is straightforward
- Given the random assignment, the basic approach for estimating savings is to simply compare the consumption of treatment and control customers using **difference-in-differences**
- Implemented using a **panel regression model** that included an indicator variable for month, a treatment and a customer-level indicator variable (fixed effect)

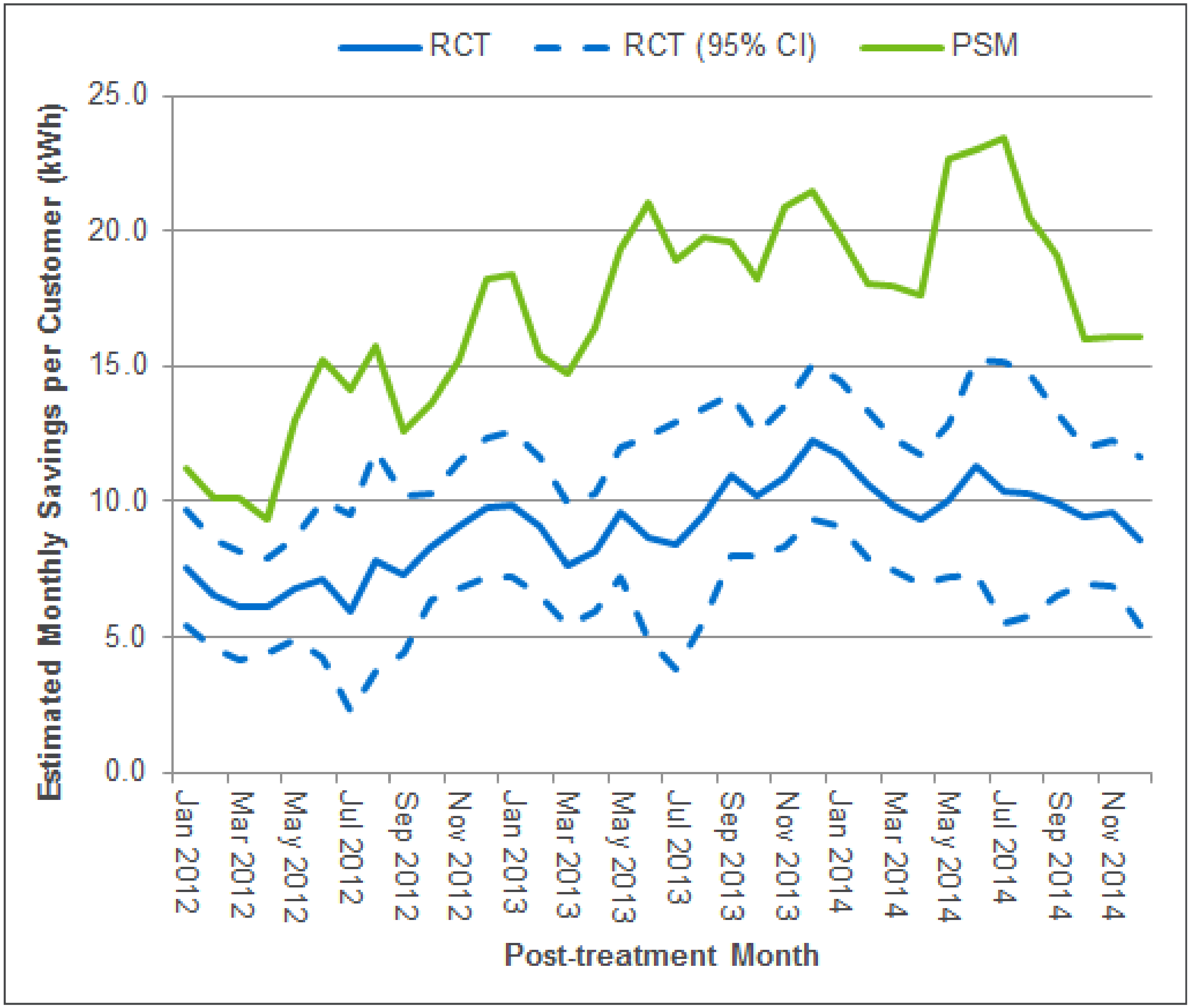
Propensity Score Matching is a Typical Quasi-Experimental Approach

- Propensity Score Matching (PSM) is commonly used to estimate impacts for demand response programs
- Developed a matched control group from a **large dataset of non-participants** (approximately 500,000 customers)
- Each of the 75,000 treatment customers was **matched** to a customer in the large dataset of non-participants using PSM
- Variables included in the propensity score model were simply the **pretreatment monthly usage** amounts for November 2010 through October 2011
- An additional constraint was that each treatment customer had to be matched from **within the same weather station area**
- **Primary disadvantage** of PSM is unobserved selection bias

PSM Uses Pretreatment Usage as a Basis for Verifying the Similarity of the Treated and Matched Controls



PSM Approach Produced Similar Estimates to RCT, But Outside of the 95% Confidence Interval for the RCT (dotted lines)

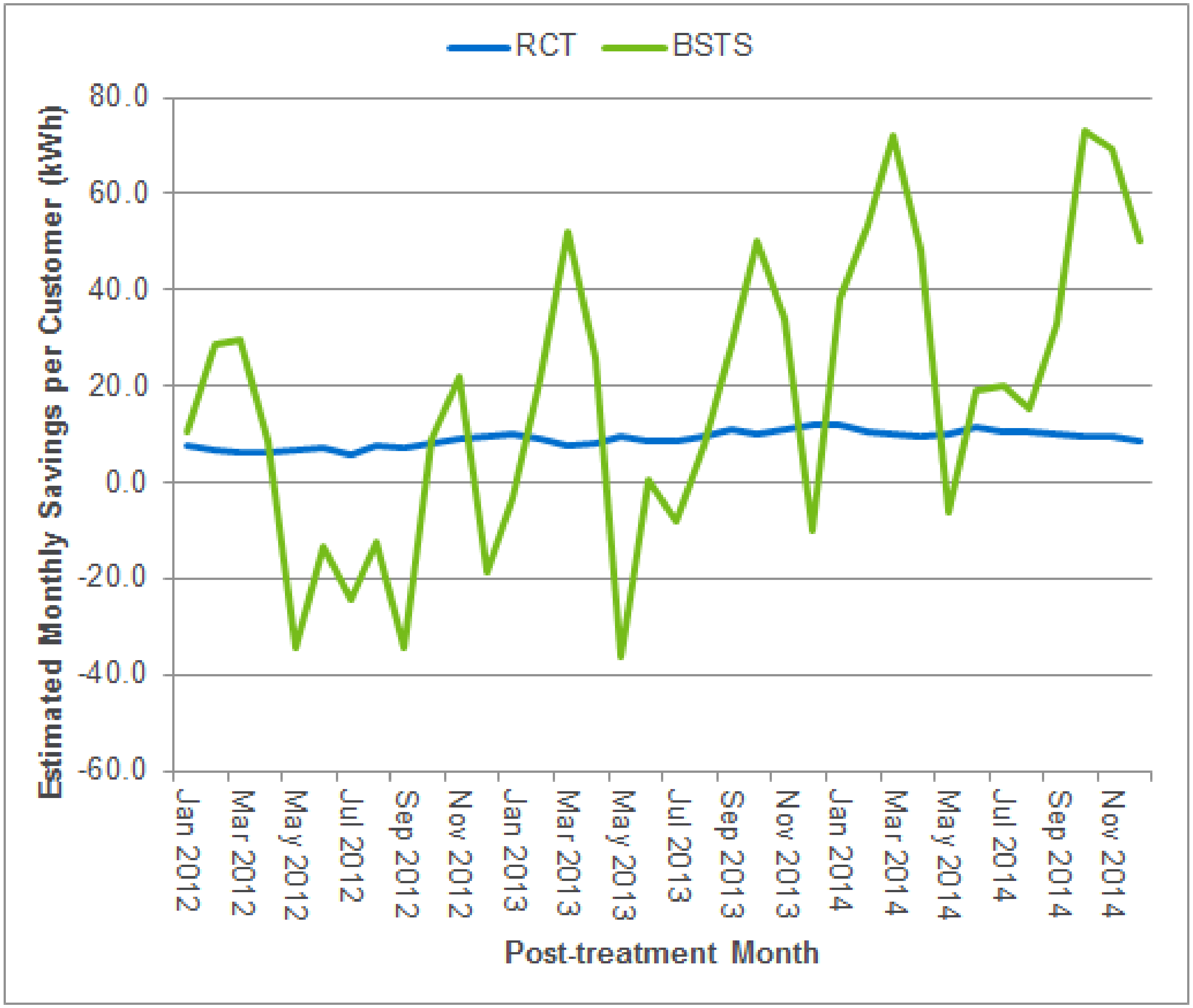


- PSM monthly savings estimates vary from **9.4 kWh to 23.4 kWh**
- Nonetheless, the PSM percent savings estimate is nearly **double** the RCT estimate in 2014
- Upward bias already shows up in the **first post-treatment month**

Bayesian Structured Time Series Allows for Modeling Complex, Non-Linear Relationships

- Bayesian Structured Time Series (BSTS) analysis was conducted using an **R package** called `CausalImpact` using a model with the following variables (including their higher powers and interactions)
  - Average kWh for the treatment group only
  - Heating degree days (HDD)
  - Cooling degree days (CDD)
  - Relative humidity
  - Monthly seasonal effect
- **Disadvantages** of BSTS:
  - Complexity (black box)
  - Statistical learning models have a risk of “overfitting”, whereby too much importance is placed on random patterns in the data, especially when relatively few data points are available

BSTS Produced “Noisy” Results

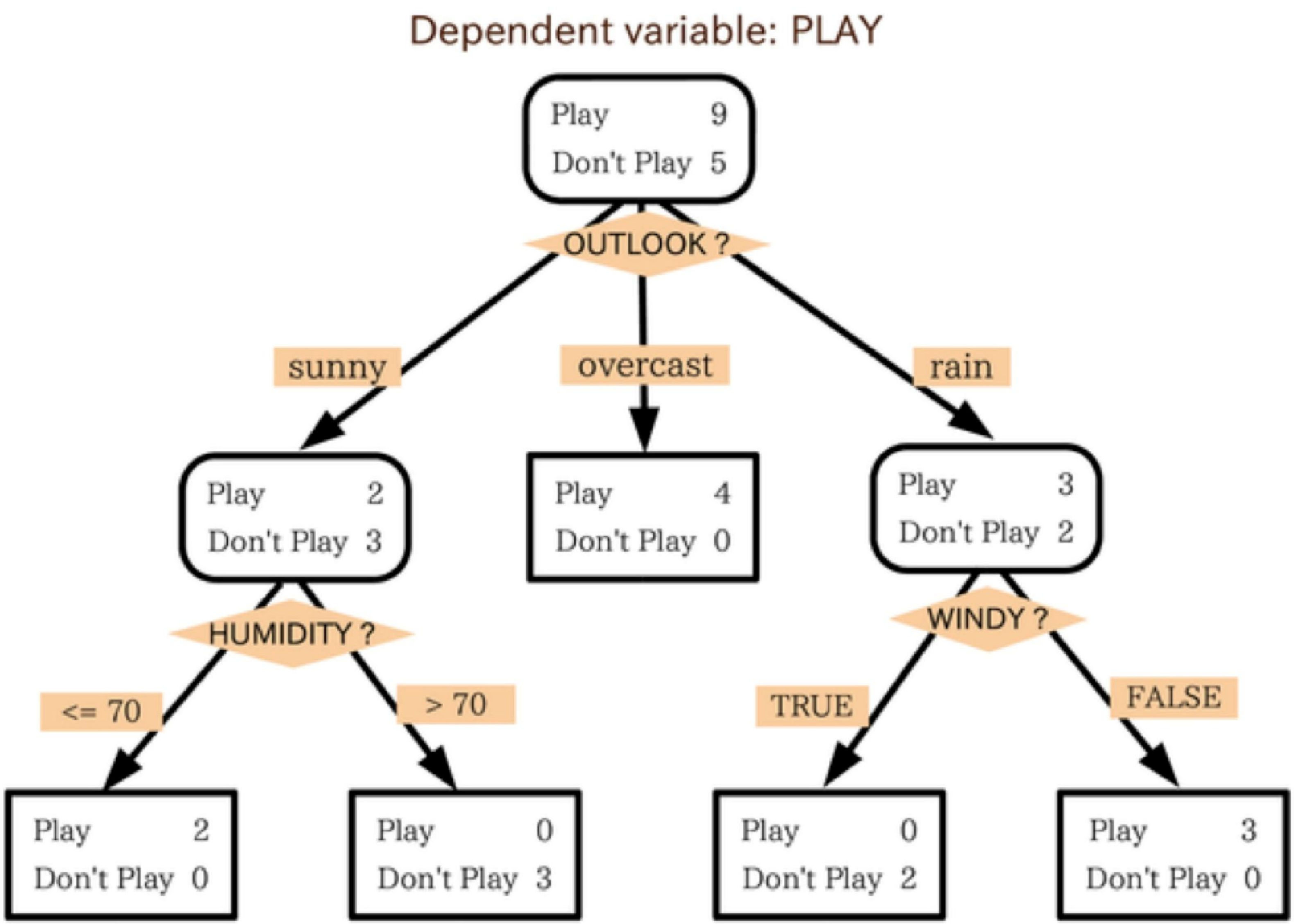


- BSTS monthly savings estimates vary from as low as ***negative* 36.4 kWh** to as high as 73.1 kWh
- In 2014, the BSTS percent savings estimate is nearly **four times higher** than the RCT estimate

Regression Tree with Random Effects Offers the Flexibility of Tree-Based Methods with the Structure of Random Effects Models

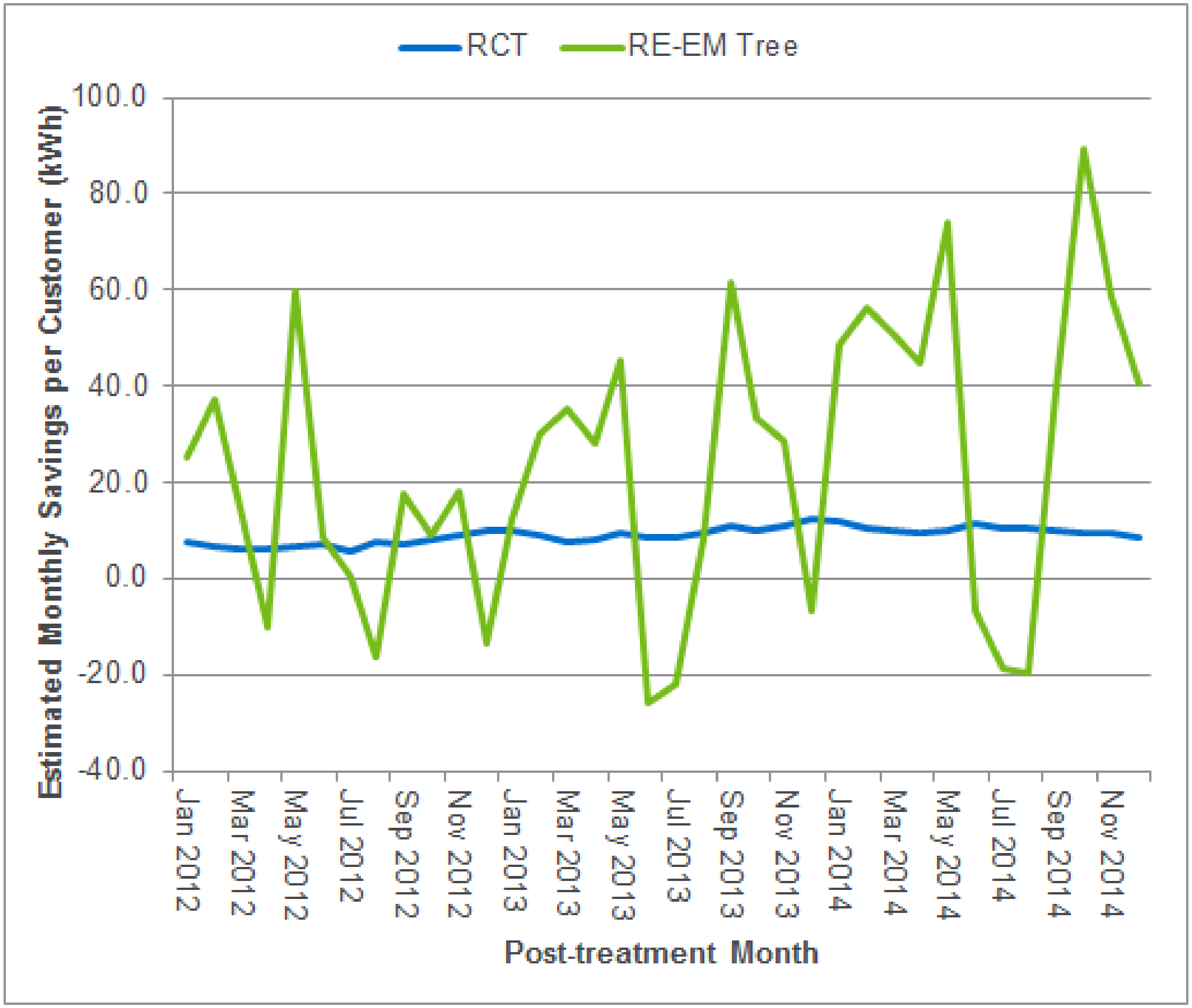
- Tree-based models are rule-based models that partition data based on one or more nested **if-then statements** applied to the independent variables
- Regression Tree with Random Effects (RE-EM) Tree model in this case used **similar variables** to those of the BSTS model
- **Disadvantages** of tree-based models
  - Prone to model instability
  - Poor predictive performance if the relationship between predictors and response cannot be adequately defined, especially when relatively few data points are available

Tree-Based Model Examples



Source: https://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/

RE-EM Tree Produced “Noisy” Results



- RE-EM Tree monthly savings estimates vary from as low as ***negative 25.8 kWh*** to as high as 89.3 kWh
- In 2014, the RE-EM Tree percent savings estimate is nearly **four times higher** than the RCT estimate

Alternative Methods Tested Produced Different Savings Estimates

Estimation Methodology	Percent Savings Estimates			Monthly Savings Estimates (kWh)	
	2012	2013	2014	Low	High
Randomized Controlled Trial (RCT)	1.16%	1.58%	1.69%	5.9	12.2
Bayesian Structured Time Series (BSTS)	-0.40%	2.21%	6.43%	-36.4	73.1
Regression Tree with Random Effects (RE-EM Tree)	2.03%	3.07%	6.08%	-25.8	89.3
Propensity Score Matching (PSM)	2.08%	3.07%	3.21%	9.4	23.4

- PSM performs best, but it **does not resolve the issue** of requiring a large group of customers who do not receive the treatment
- Changes in usage and weather conditions led to a **large upward bias** in the 2014 BSTS and RE-EM estimates, given that both models primarily rely on temperature to estimate usage

Follow-on Research Ideas

- Further research based on several years of hourly interval data is required to **conclusively determine** whether these models are (or are not) a viable alternative to the RCT
  - Nonetheless, a model that primarily relies on temperature patterns may go awry after several years of treatment
  - Key advantage of the PSM approach is that it does not rely on modeling a relationship between temperature and usage, which most likely explains why the PSM results track most closely to the RCT results over multiple years
- Conduct a similar comparative methods analysis for an **opt-in program**

Acknowledgements

- Co-authors
  - Aimee Savage, Nexant, San Francisco, CA
  - Marshall Blundell, Nexant, New York, NY
  - Jonathan Cook, Nexant, Washington, DC
  - Brian Arthur Smith, Pacific Gas and Electric Co., San Francisco, CA
- For their helpful input, we would also like to thank the following people
  - Hunt Allcott, New York University
  - Alex Orfei, Opower
  - Ken Agnew, DNV GL
  - Peter Franzese, California Public Utilities Commission (CPUC)
  - Dan Bush, CPUC
  - IEPEC co-panelists and moderator