Losing Control: What Will Happen if Randomized Controlled Trials are Phased Out of Behavioral Program Evaluation?

Josh Schellenberg, Nexant

2015 IEPEC Conference — Long Beach, California

International Energy Program Evaluation Conference

Reimagine tomorrow.

First off, I would like to thank the many contributors to this paper

- 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 from New York University
  - Alex Orfei from Opower
  - Ken Agnew from DNV GL
  - Peter Franzese from the California Public Utilities Commission (CPUC)
  - Dan Bush from the CPUC
  - IEPEC co-panelists and moderator

Utilities and other stakeholders have started to question the future viability of the RCT

Current Situation

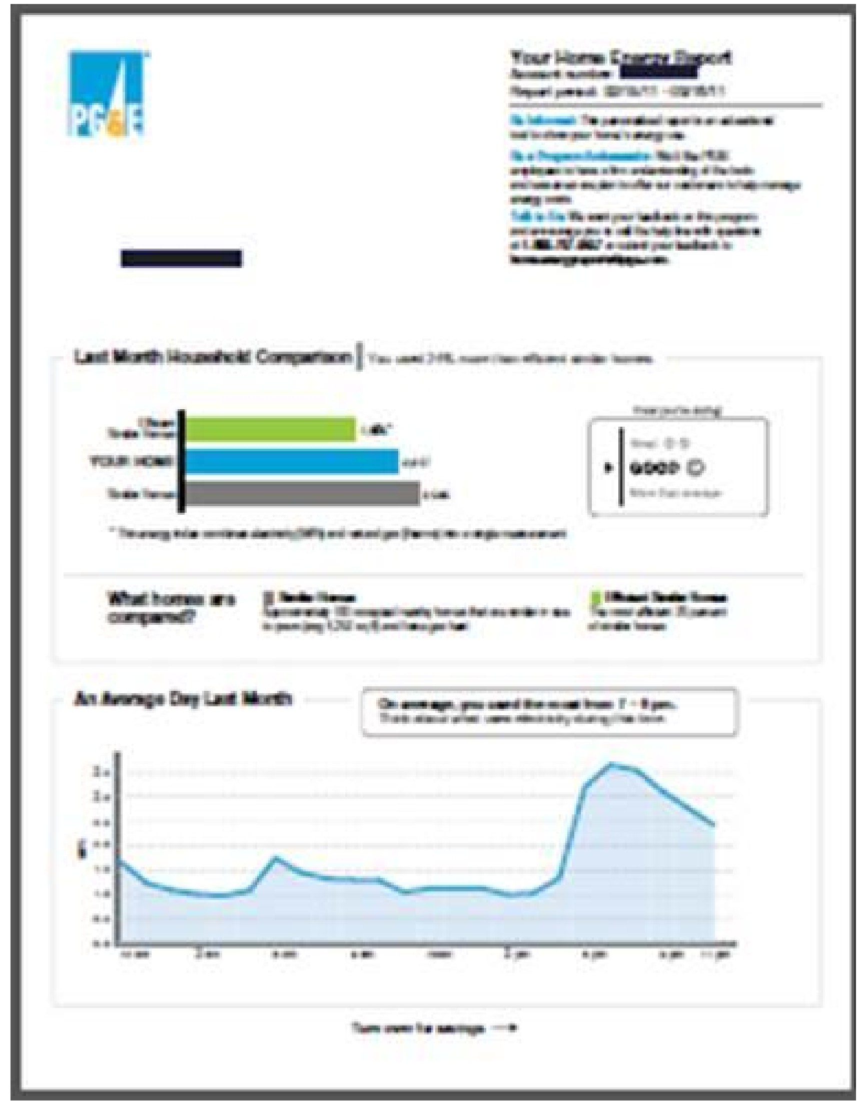
- Home energy reports (HERs) have gained **significant traction** in the utility industry
- Preferred evaluation method has been the randomized controlled trial (RCT), but large control groups of non-participants **limit the potential for behavioral programs**
- Innovative statistical methods show **considerable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

Study Objective

- **Leverage data** from one of the largest HER programs
- **Test** 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 evaluation to the energy savings that evaluators would have estimated using alternative methods

HERs provided by Opower have expanded rapidly at Pacific Gas and Electric Co. (PG&E) since 2012

- **1.2 million** residential customers receiving HERs at this point
- HERs account for **majority of savings** in residential EE portfolio
- In **this analysis**:
  - 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)



RCT has produced savings estimates that are consistent across years and months

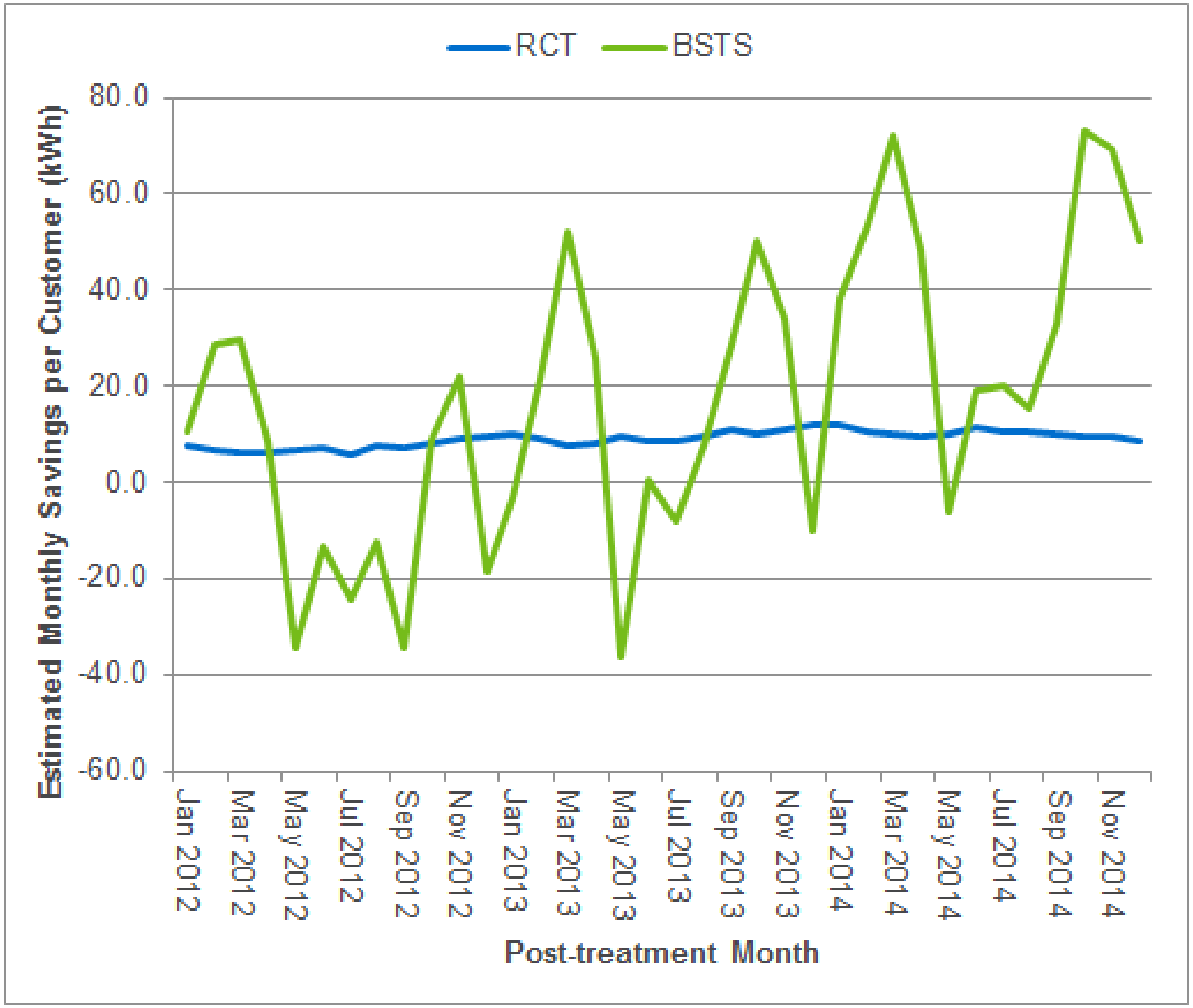
Estimation Methodology	Percent Savings Estimates			Monthly Savings Estimates (kWh)	
	2012	2013	2014	Low	High
Randomized Controlled Trial (RCT)	1.16%	1.58%	1.69%	5.9	12.2

- 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)

BSTS provides flexibility in modeling trend, seasonality and regression effects that allow for complex, non-linear relationships

- An **R package** called `CausalImpact` was obtained online and applied 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 results that were extremely noisy from month to month and not accurate in any given year

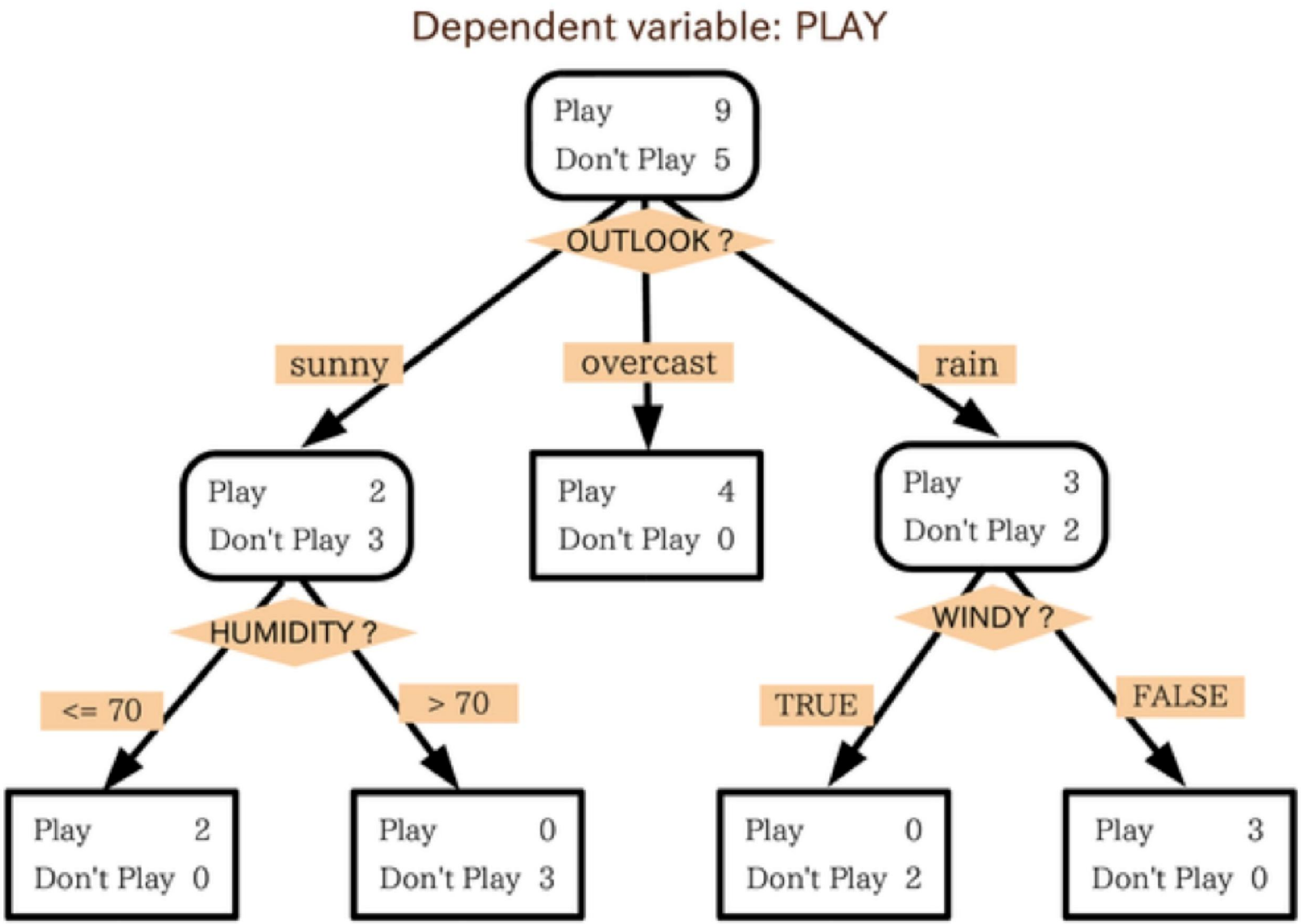


- BSTS monthly savings estimates vary from as low as ***negative 36.4 kWh*** to as high as 73.1 kWh
- In 2014, the BSTS percent savings estimate is nearly **four times higher** than the RCT estimate

RE-EM Tree models combine the flexibility of tree-based methods with the structure of random effects models

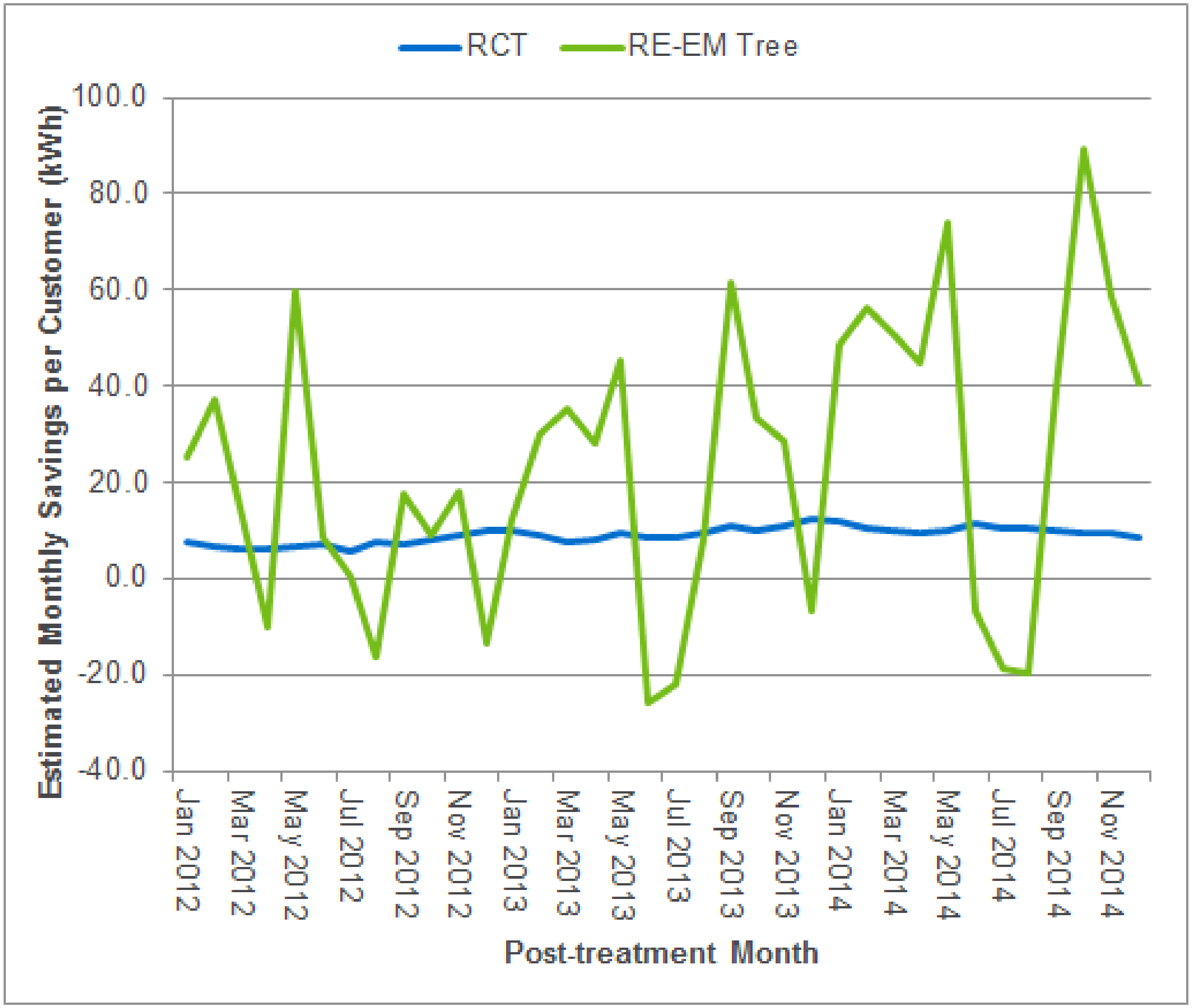
- Tree-based models are rule-based models that partition data based on one or more nested **if-then statements** applied to the independent variables
- RE-EM Tree model in this case used **similar variables** to those of the BSTS model
- **Disadvantages** of tree-based models
  - Prone to model instability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

Simple example of a tree-based model



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

RE-EM Tree also produced results that were extremely noisy from month to month and not accurate in any given year

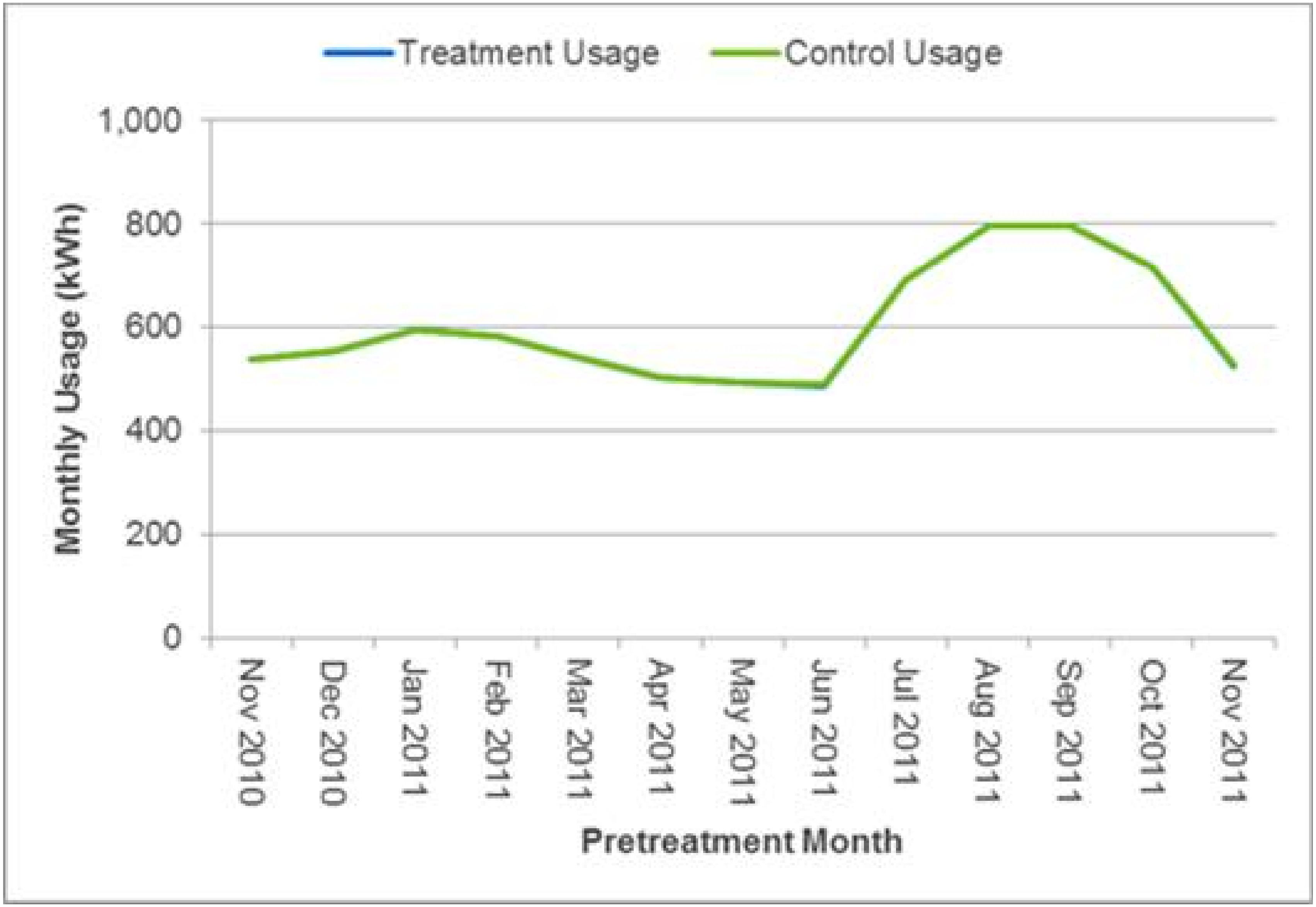


- RE-EM Tree monthly savings estimates vary from as low as ***negative 25.8 kWh*** to as high as 89.3 kWh
- In 2014, the RE-EM Tree percent savings estimate is nearly **four times higher** than the RCT estimate

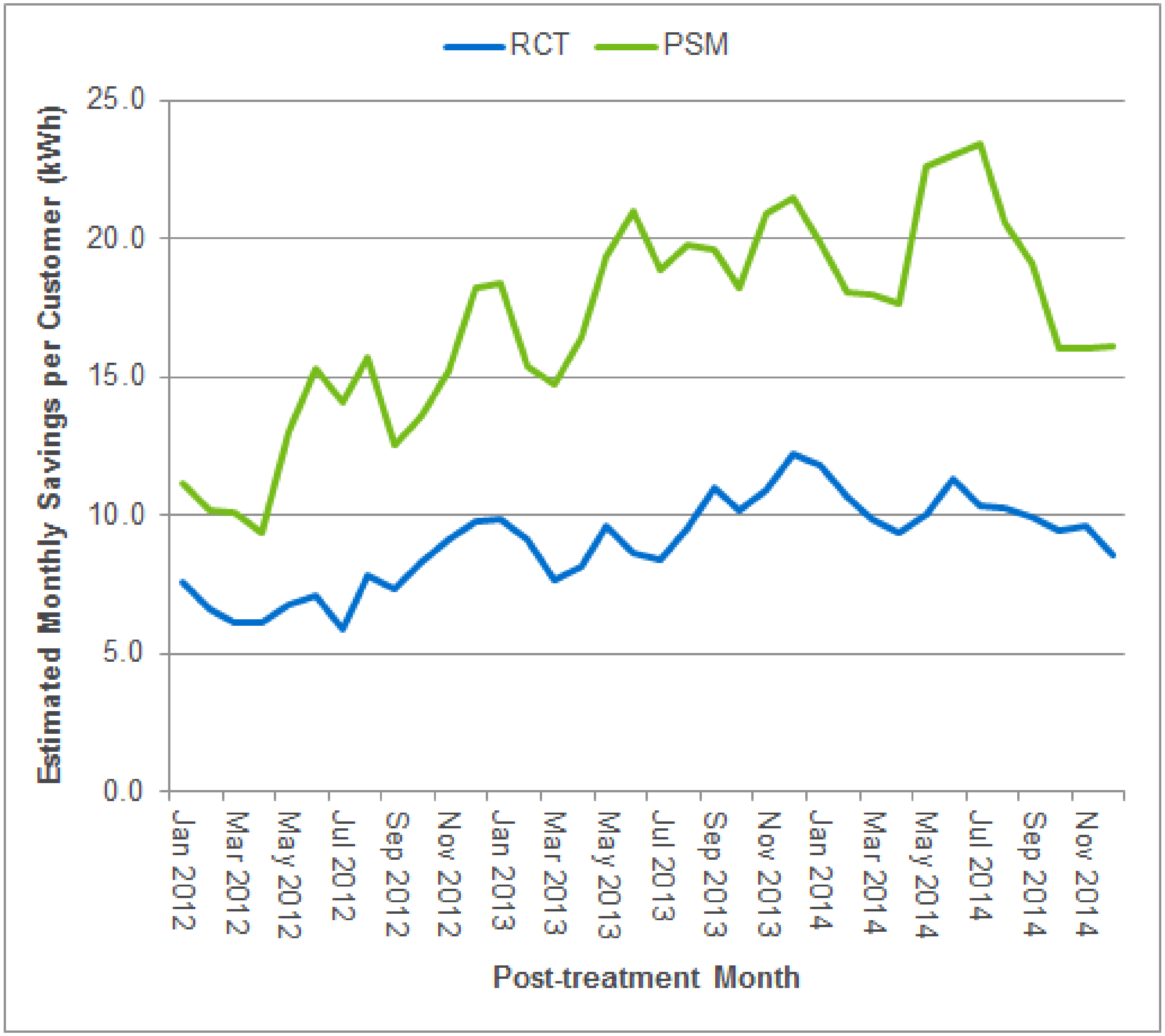
PSM creates a matched control group in the absence of a randomized experiment

- 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

During the pretreatment period, the difference between treatment usage and matched control group usage is very small (<0.3%)



PSM produced results that are much more similar to the RCT results in both magnitude and consistency



- 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**

Conclusion – Estimates from the three alternative methods tested are quite different from the RCT results

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

What's next?

- 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**

For comments or questions, contact:

Josh Schellenberg
Principal
jschellenberg@nexant.com

Nexant, Inc.
101 Montgomery St., 15th Floor
San Francisco, CA 94104
415-777-0707