Finding the Perfect Baseline:

Advanced Time Series Control Group Matching Strategies on Energy Consumption

Christopher McBride -- ADM Associates Inc.

2019 IEPEC Conference – Denver, CO

ADM

ENERGY RESEARCH AND EVALUATION

IEPEC

Presentation Focus: Difference-in-Differences and the Selection of a Control Group

Objective: Understand the impacts of an energy efficiency measure through statistical analysis of billing or meter data.

Approach: Compare the kWh consumption across four regimes

- Treatment group pre-installation
- Treatment group post-installation
- Control group pre-installation
- Control group post-installation

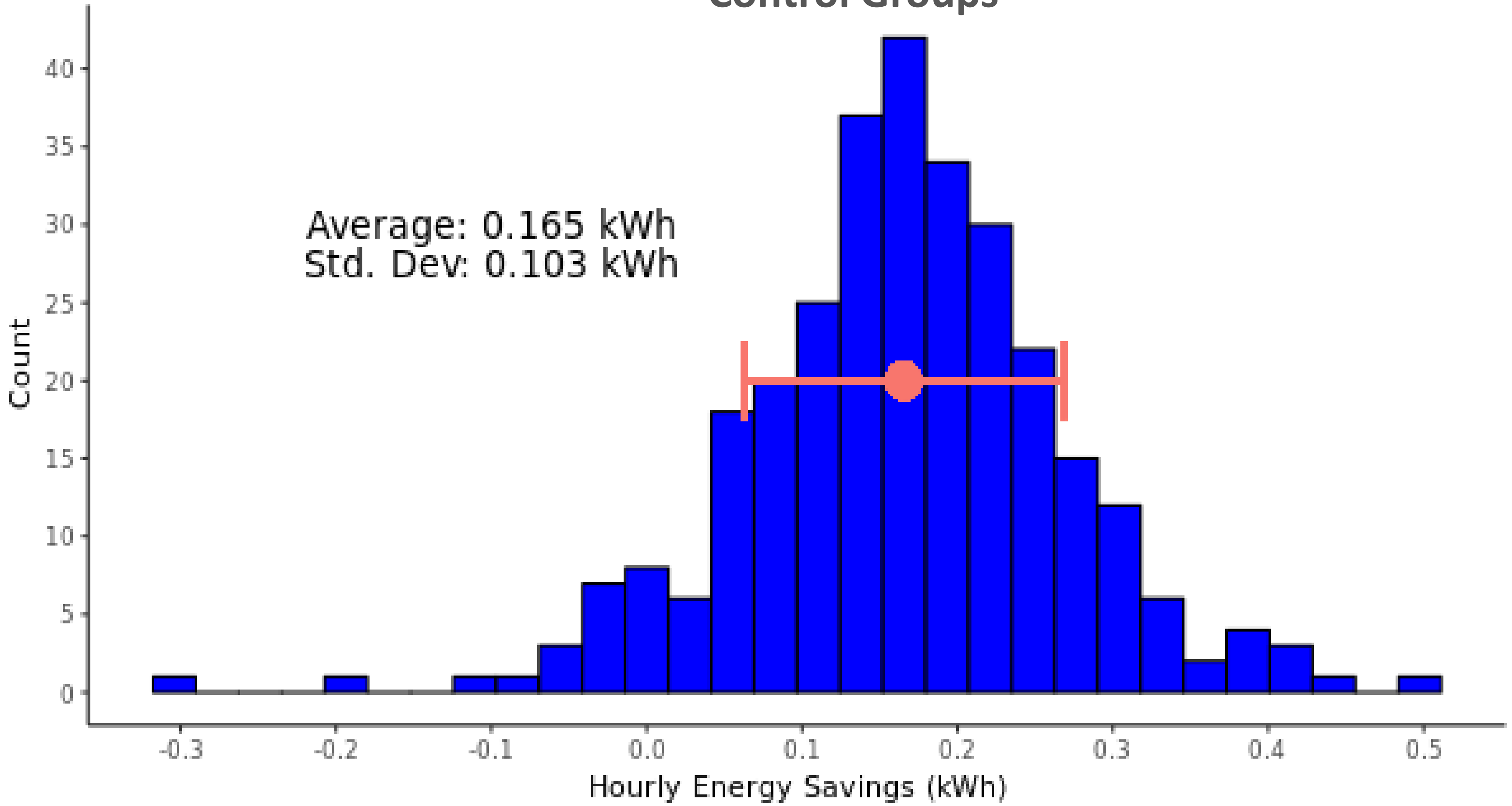
Challenge: *How can we be sure that our control group is an accurate representation of the counterfactual?*

Data Exploration

15-minute AMI Smart Meter

- Abnormal data points / Non-steady state
  - Low occupancy
  - Drastic consumption changes (i.e. Multifamily)
- Random Matching
  - Gives idea of expected range and variation

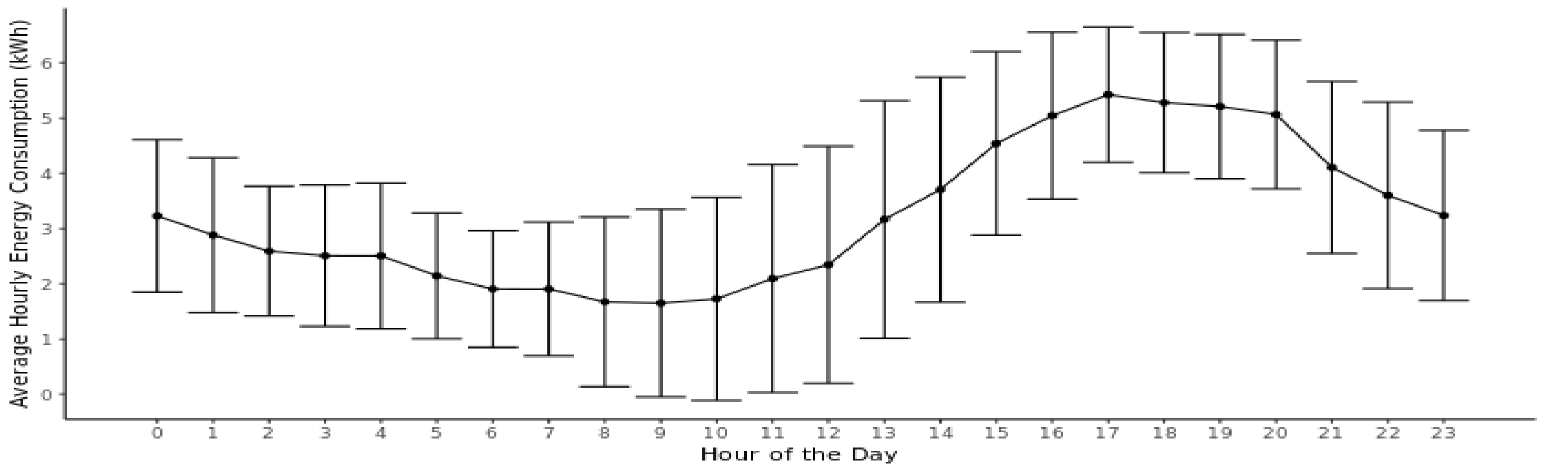
Results of Diff-in-Diff Analysis on 300 Unique Randomly Selected Control Groups

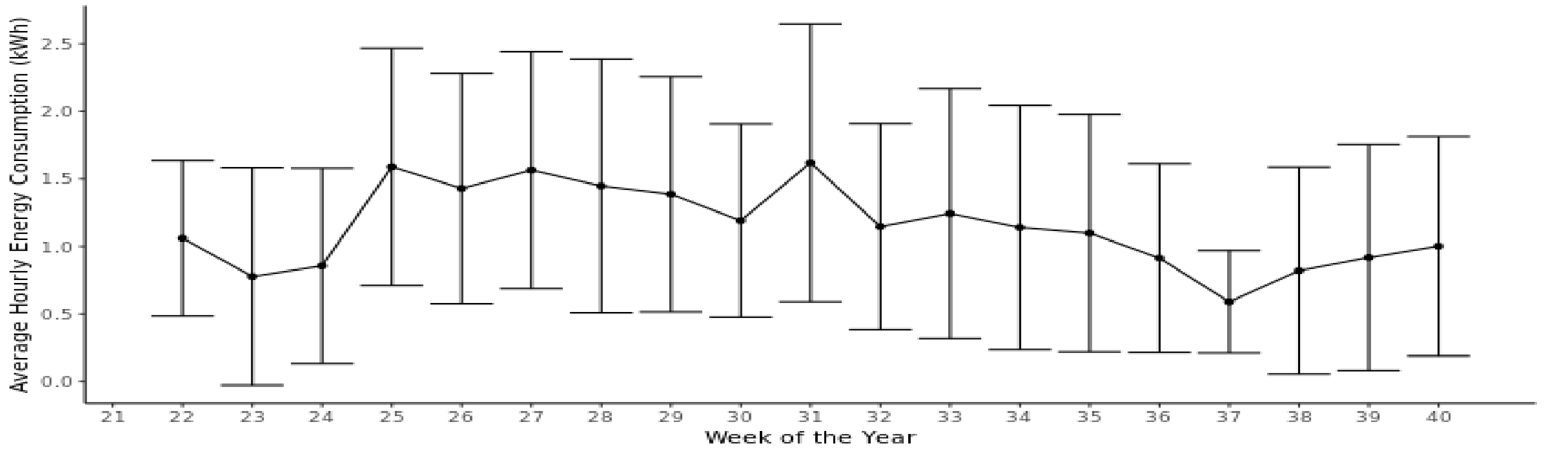


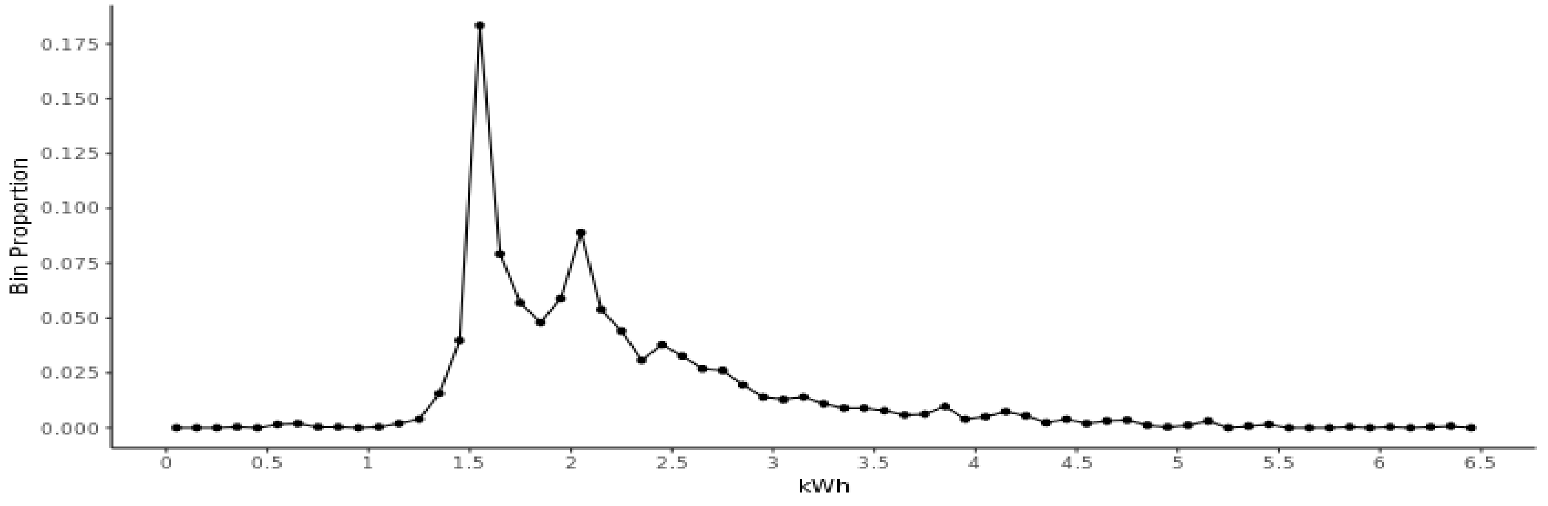
Developing a Covariate Set

3 Covariate Set Options:

- 24-Hour Curve
- Weekly Average
- Normalized Histogram Curve



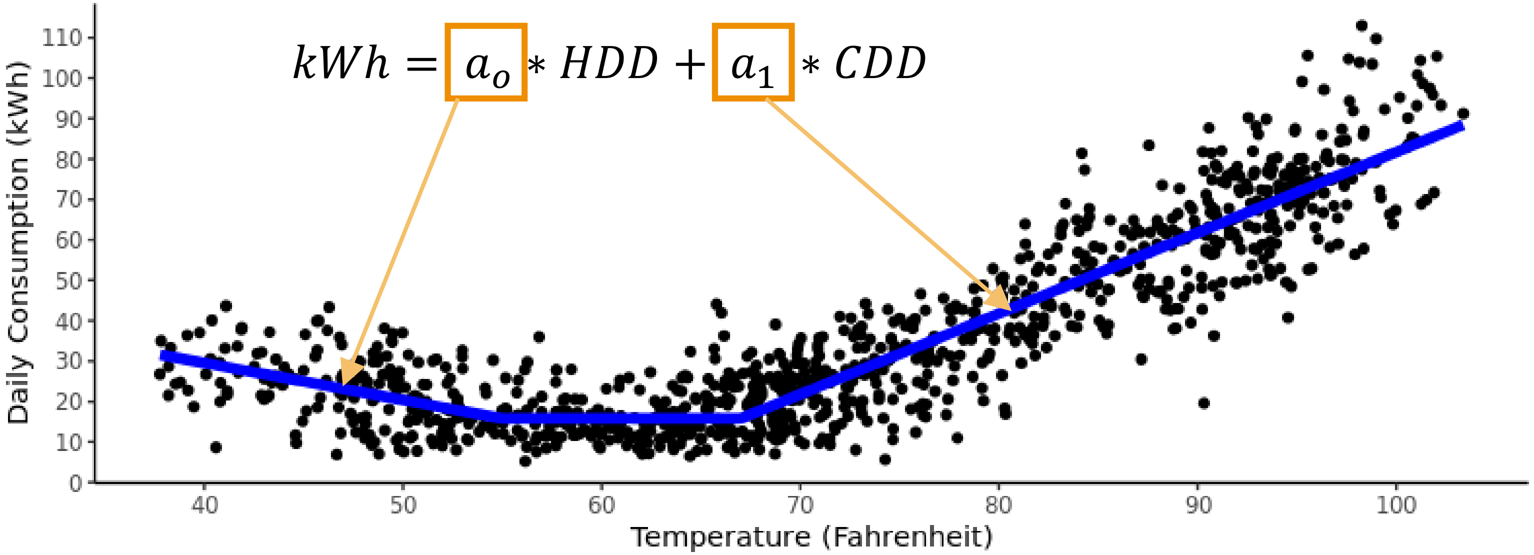


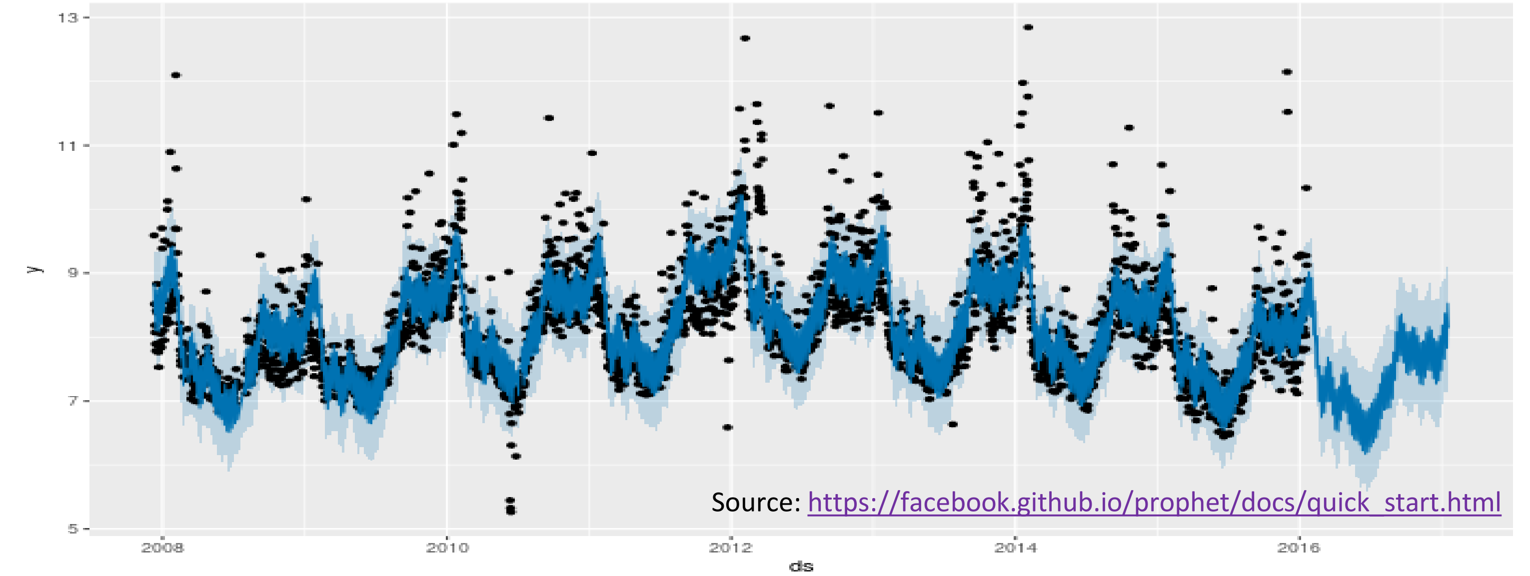


Developing a Covariate Set

3 More Covariate Set Options:

- Temperature Response
- Facebook's Prophet
- Manual Regression Modeling





$$kWh_i(t_i) = \alpha_1 * \sin\frac{2\pi}{\omega t_i} + \alpha_2 * \cos\frac{2\pi}{\omega t_i} + \alpha_3 * Weather + \alpha_4 * Month$$

Matching Algorithms

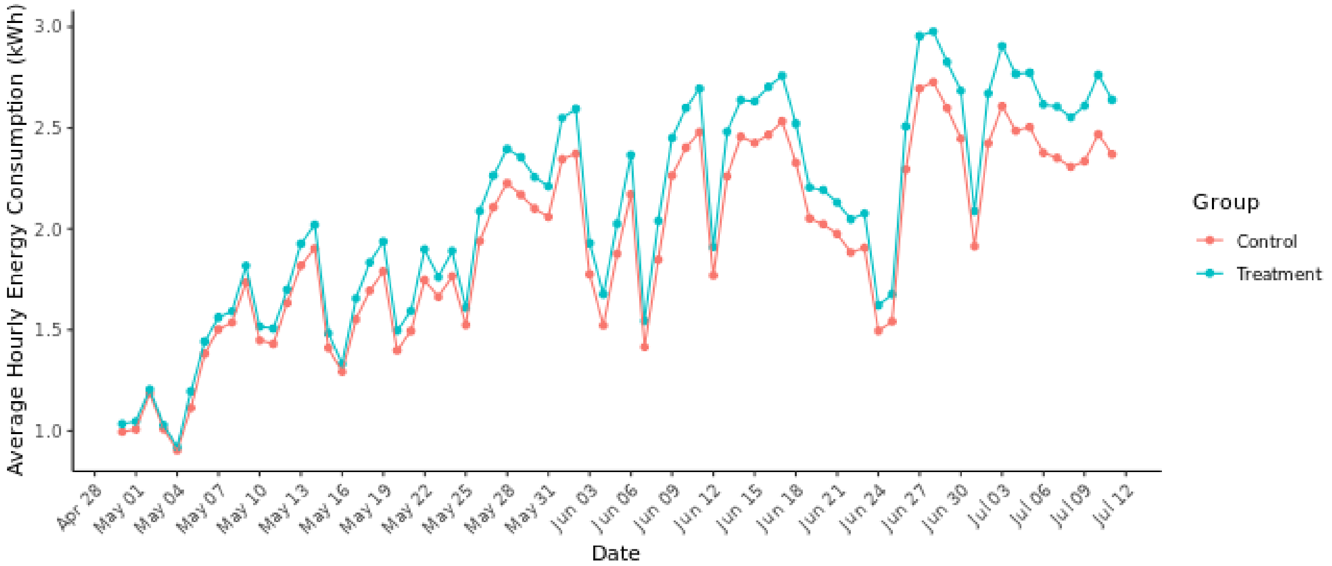
- Propensity Score
- Absolute Difference
  - $d(\boldsymbol{q},\boldsymbol{p}) = \sum_{\boldsymbol{i=1}}^{n}(q_i - p_i)$
    $= (q_1 - p_1) + (q_2 - p_2) + \cdots + (q_n - p_n)$
- Euclidean Distance
  - $d(\boldsymbol{q},\boldsymbol{p}) = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(q_i - p_i)^2}$
    $= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$
- Vector Angle Difference
  - $\theta(\boldsymbol{q},\boldsymbol{p}) = \cos^{-1}\frac{\boldsymbol{q}\cdot\boldsymbol{p}}{\|\boldsymbol{q}\|\,\|\boldsymbol{p}\|}$
    $= \cos^{-1}\frac{q_1p_1+q_2p_2+\cdots+q_np_n}{\sqrt{q_1^2+q_2^2+\cdots+q_n^2}\,\sqrt{p_1^2+p_2^2+\cdots+p_n^2}}$

Reviewing Results

☑ 1. *Are results reasonable?*

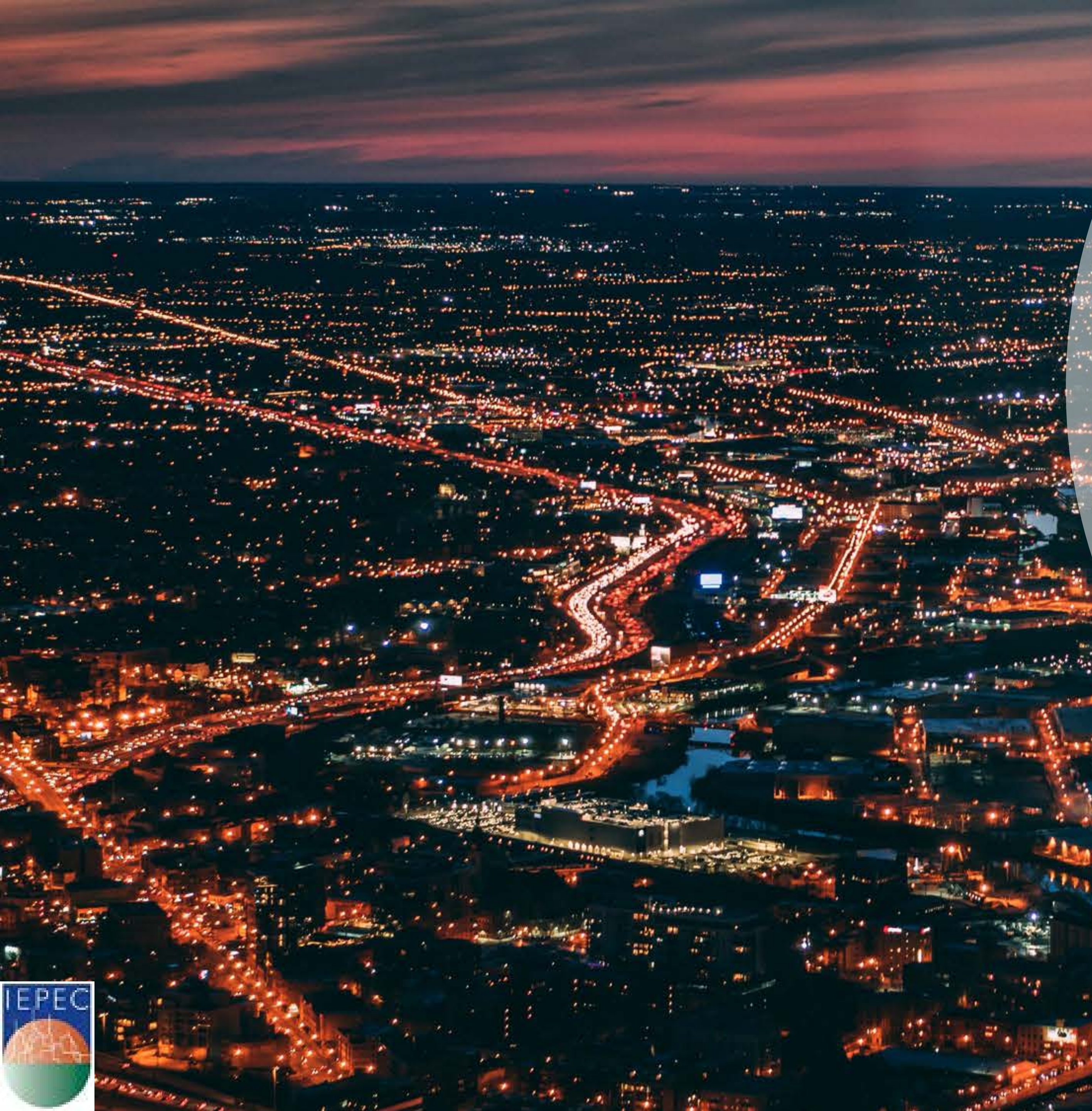
- ☑ Within random matching range?
- ☑ Agree with engineering calculations?

☑ 2. *Has the parallel trends assumption been satisfied?*



☑ 3. *Are results consistent?*

- ☑ Stable under multiple iterations of varying data subsets?
- ☑ Stable under other high-ranking control group matches?



Questions?

For more information, contact:

Christopher McBride

Data Scientist

ADM Associates

christopher.mcbride@admenergy.com

Visit our website, www.admenergy.com