

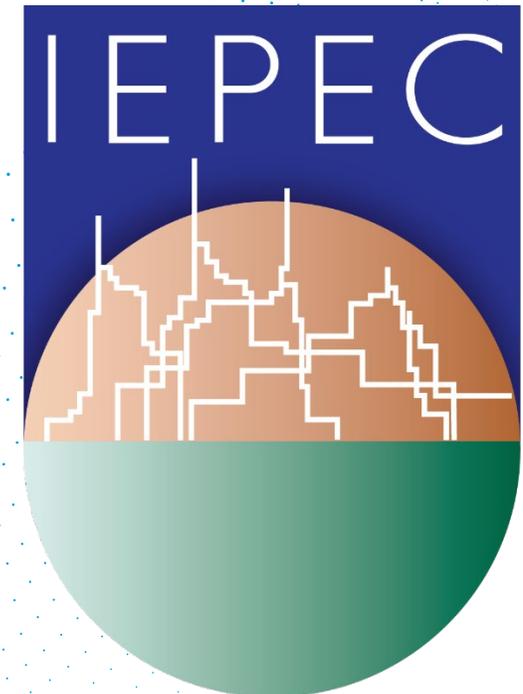


# Differentially Trendy?

## Understanding and Addressing Non-Parallel Trend Bias in Smart Thermostat Impact Estimates

Peter Franzese, CPUC, Lullit Getachew and Ken Agnew, DNV

07 October 2025



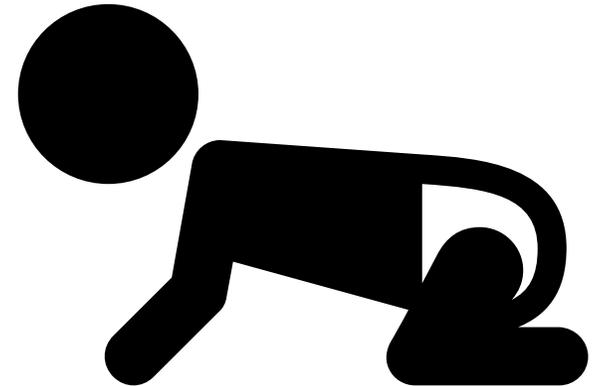
# Agenda

- Lower than expected smart thermostat savings and the “baby hypothesis”
- Difference-in-difference analysis
  - Parallel trends assumption
  - Non-parallel trends bias
- Smart thermostats impact evaluation
- Solutions: Identifying and addressing potential bias

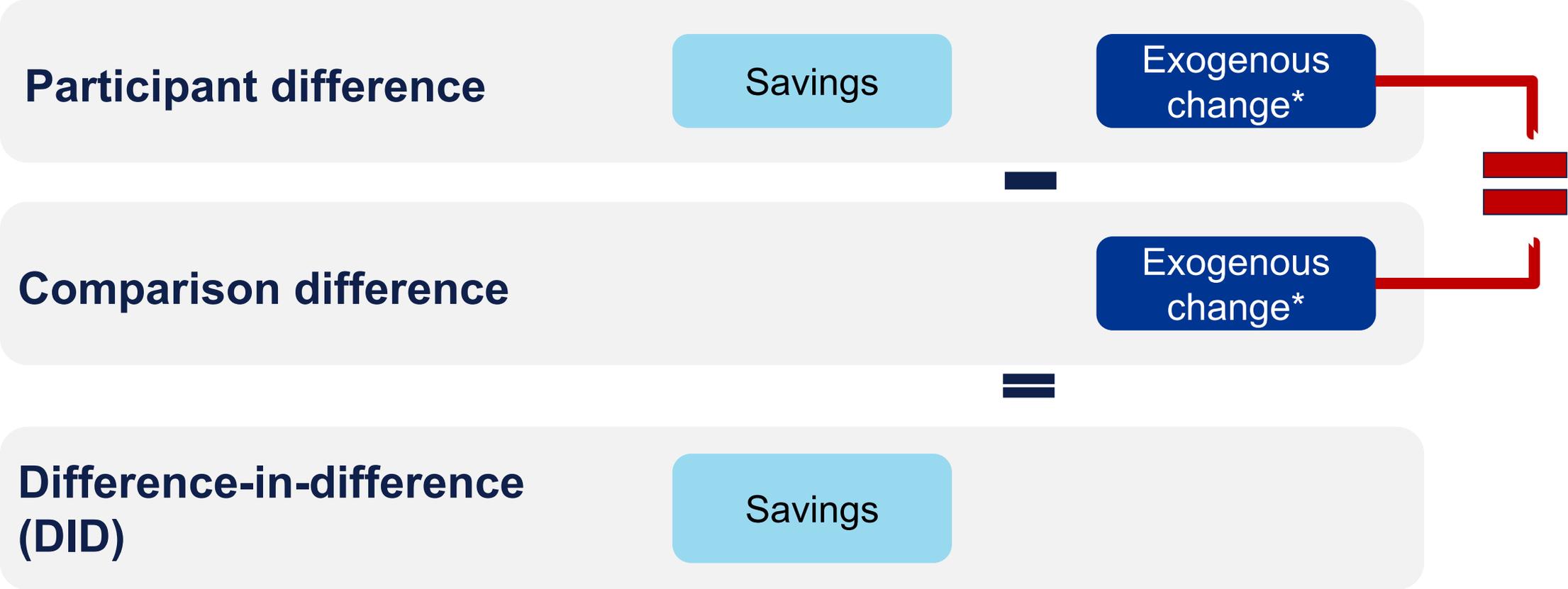


# Non-Parallel Trends in Smart Thermostat Evaluations?

- Why did this become an issue?
  - Savings were not of the expected magnitude
  - A knowledgeable ex-evaluator from a thermostat company floated an hypothesis that leaned on non-parallel trends
- The baby hypothesis:
  - Smart thermostat purchasers were younger, more tech-savvy, better off
  - As a result, more likely to have an upward consumption trend relative to a comparison group
  - Classic self-selection



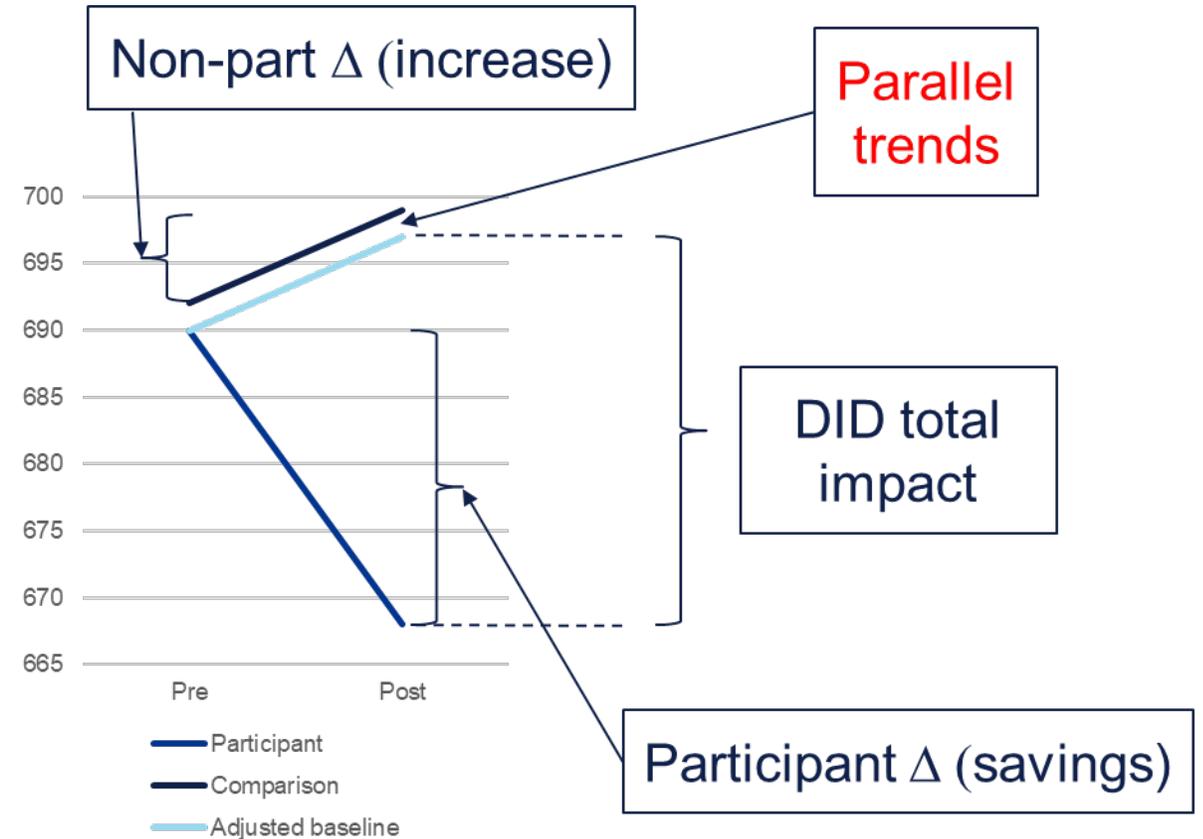
# Billing Analysis Difference-in-Difference



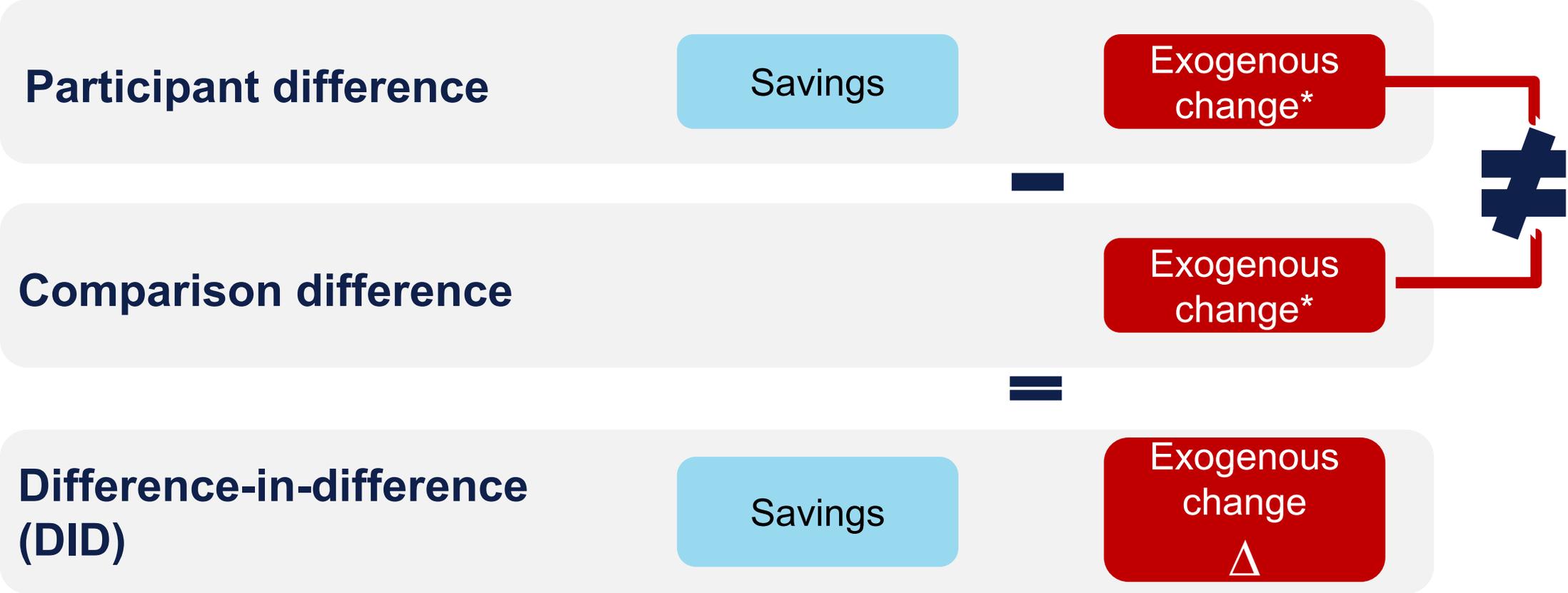
\*Exogenous change – any non-program-related change that is captured.

# Parallel Trends

- Same exogenous change = parallel trends
- Plots pre-post change for both participants and comparison group
- Assuming parallel trends is essential to the DID result



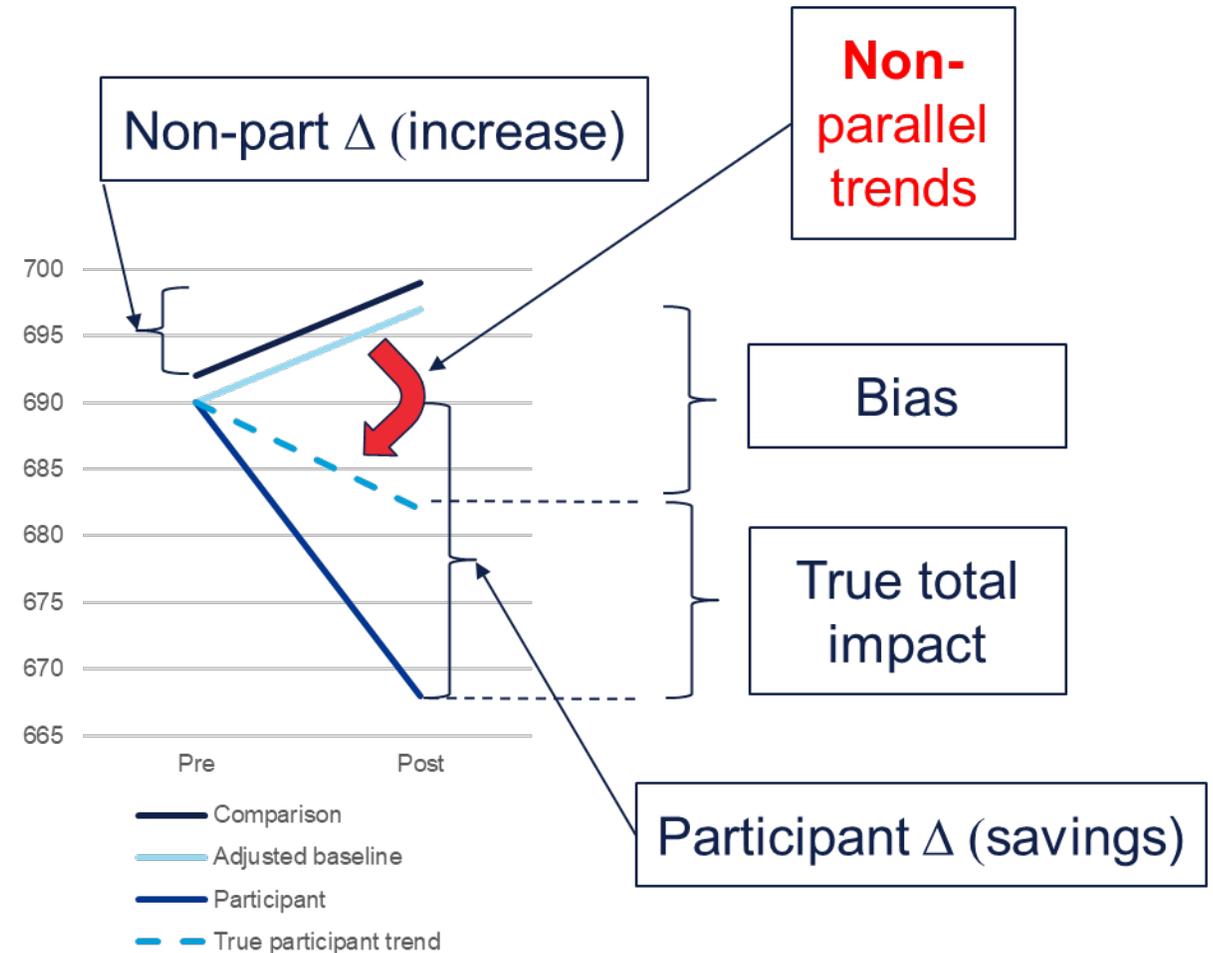
# Billing Analysis Difference-in-Difference w/ Bias



\***Exogenous change** – any non-program-related change that is captured.

# Non-Parallel Trends

- What if the comparison group is a poor proxy for the participant group?
- Biases the DID total impact
  - Bias can be either direction
  - Based on how the true underlying counterfactual trend differs from the comparison group



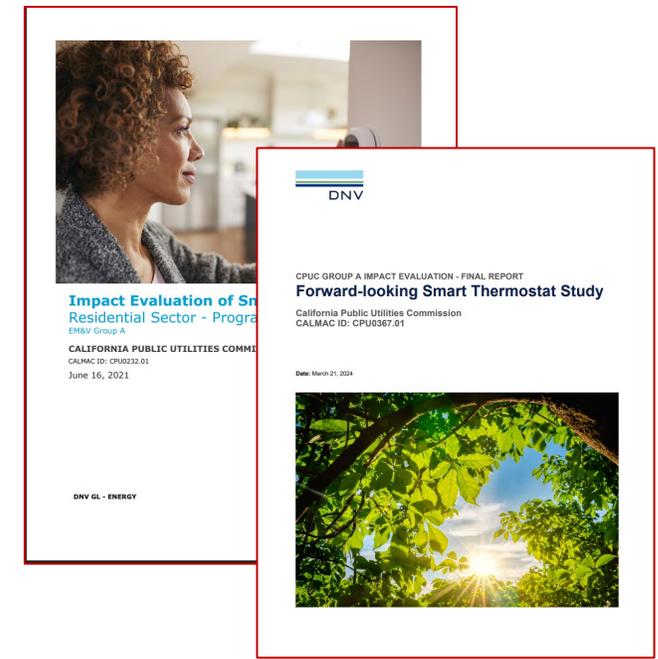
# Non-Parallel Trends in Smart Thermostat Evaluations?

- No evidence of the baby hypothesis put forward, but it was plausible
- Focusing on acknowledged shortcomings of the methods undermines the validity of the evaluation
- Plenty of other reasons smart thermostats might not generate expected savings
  - Remember programmable thermostats?
  - Relative to baseline behavior
  - A substantial behavioral component closely tied to comfort



# California Smart Thermostat Studies

- 2021 – Original impact evaluation
  - Recognized the non-parallel trends issue
  - Developed a test and simple adjustment
- 2023 – Follow-on forward-looking study
  - Opportunity to return to prior data, add more data, pursue further methods related to this issue
  - Developed a more robust version, panel model version of the trend adjustment
- California Public Utilities Commission. 2021. *CPUC Group A Residential PY2019 Smart Thermostat Impact Evaluation – Final Report*. CALMAC ID: CPU0354.01.
- California Public Utilities Commission. 2023. *Forward-Looking Smart Thermostat Study. CPUC Group A Impact Evaluation – Final Report*. CALMAC ID: CPU0367.01.



# 2021 Adjustment Approach

- Smart thermostat impacts → limited to heating and cooling consumption
- If parallel trends → baseload impacts should be zero
- Two-stage, site-level modelling approach to billing analysis makes testing this easy
  - Each site's consumption is disaggregated to baseload / heating & cooling for the weather normalization process
  - Estimated the second stage DID by baseload / heating & cooling
- Did find some evidence of negative savings in the baseload for rebate programs
  - Limit impacts to heating & cool savings estimates
  - Use trend differential apparent in baseload, as % of consumption, to adjust heating & cooling savings upward

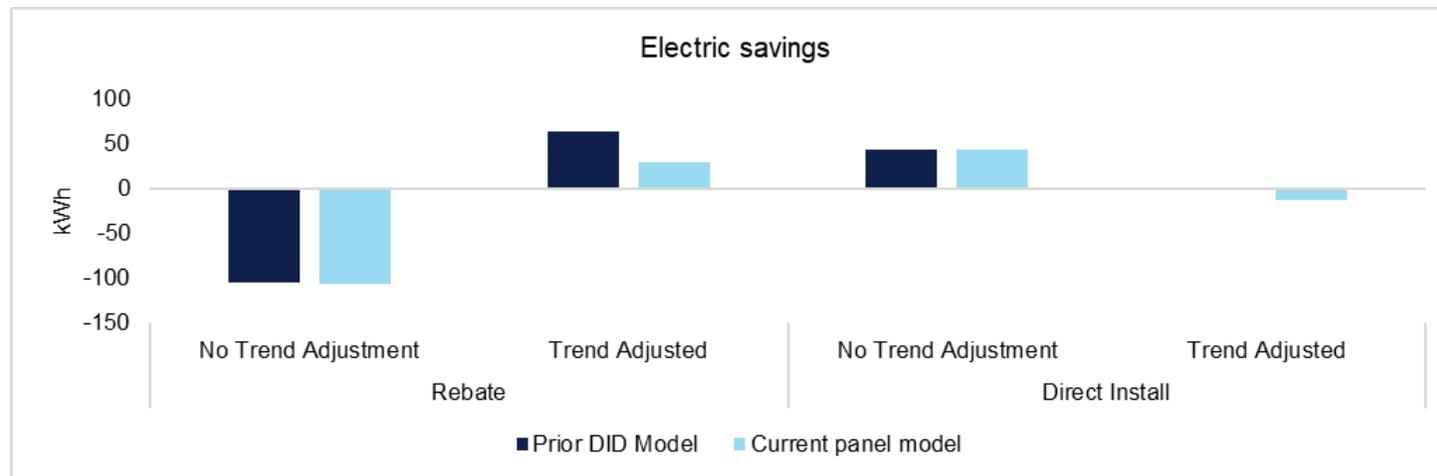
# 2023 Adjustment Approach

$$\bullet Y_{tj} = \alpha_g (1 + \lambda_g t) + \beta_g (1 + \lambda_g t) [H_t(\tau_{Hj})(1 - P_t) + H_t(\tau_{Hj} + \delta_H + \delta_{pH} * T_j)P_t] + \gamma_g (1 + \lambda_g t) [C_t(\tau_{Cj})(1 - P_t) + C_t(\tau_{Cj} + \delta_C + \delta_{pC} * T_j)P_t] + \varepsilon_{tj}$$

- $t$  = time period index, starting at  $t = 1$
- $j$  = customer index
- $g$  = group index, where  $g = p$  for participants and  $g = np$  for non-participants
- $Y_{tj}$  = energy consumption for customer  $j$  at time period  $t$
- $\alpha_g$  = group-specific intercept term that captures baseload consumption of participants, where  $g = p$  for participants and  $g = np$  for non-participants
- $\lambda$  = trend term that increments daily energy consumption, with  $\lambda_{np}t$  capturing trend for non-participants and  $\lambda_p t$  capturing trend for participants
- $T_j$  = 0/1 dummy for customer  $j$ , which equals 1 if a customer is in the participant group, 0 otherwise
- $P_t$  = 0/1 dummy for time  $t$ , which changes from 0 to 1 at  $t$  = participation date for participants and their matches
- $\beta$  = heating use per heating degree-day (HDD)
- $\gamma$  = cooling use per cooling degree-day (CDD)
- $H_t(\tau_{Hj})$  = HDD per day for customer  $j$  at heating reference temperature  $\tau_{Hj}$ , at time period  $t$
- $C_t(\tau_{Cj})$  = CDD per day for customer  $j$  at cooling reference temperature  $\tau_{Cj}$ , at time period  $t$
- $\tau_{Hj}$  = heating reference temperature for customer  $j$  determined by site-level regression models
- $\tau_{Cj}$  = cooling reference temperature for customer  $j$  determined by site-level regression models
- $\delta_H$  = average shift in heating reference temperature for all customers in the post period
- $\delta_C$  = average shift in cooling reference temperature for all customers in the post period
- $\delta_{pH}$  = incremental shift in heating reference temperature for participants in the post period
- $\delta_{pC}$  = incremental shift in cooling reference temperature for participants in the post period

# Smart Thermostat Impact Results, with Adjustments

- Produced results for rebate and direct install, with and without adjustment, with DID (2021) and panel model adjustments (2023)
- Rebate program impacts did show some evidence of non-parallel trend bias
  - Direct install did not show evidence of non-parallel trend bias
- Trend adjustments across both methods were consistent and still resulted in quite modest savings for the rebate program and no savings for direct install



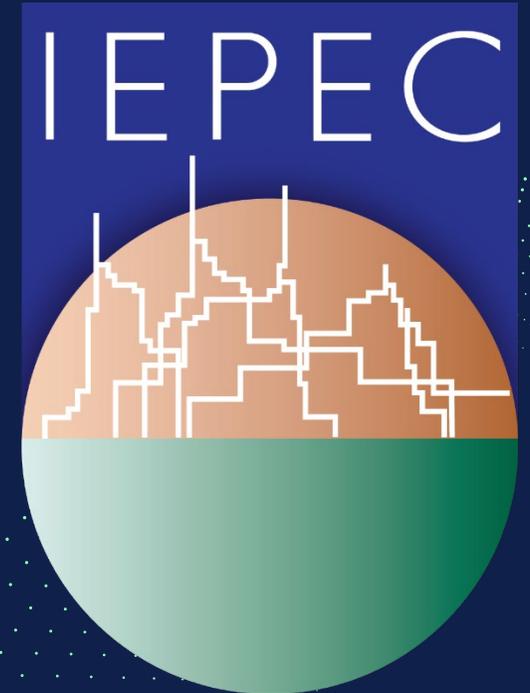
# Takeaways

- The parallel trends assumption is something we need to be thinking about.
  - Realistic way that self-selection can lead to bias in our impact estimates.
- Thermostats offered a useful way to think about, test, and address this issue.
- We offer two different ways to adjust for non-parallel trends . . . for thermostats.

More generally:

- Billing analysis is a rigorous, flexible, empirical approach to developing impacts.
- We need to be more thoughtful about these acknowledged challenges.

# Questions?



Ken.Agnew@dnv.com

Lullit.Getachew@dnv.com

[www.dnv.com](http://www.dnv.com)

