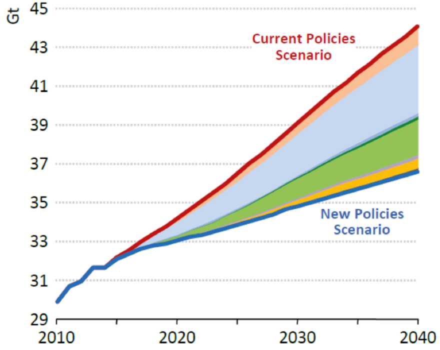
Machine Learning from Schools about Energy Efficiency

Mar Reguant (Northwestern University)

with Fiona Burlig (U Chicago), Chris Knittel (MIT),
David Rapson (UC Davis) and Catherine Wolfram (UC Berkeley)

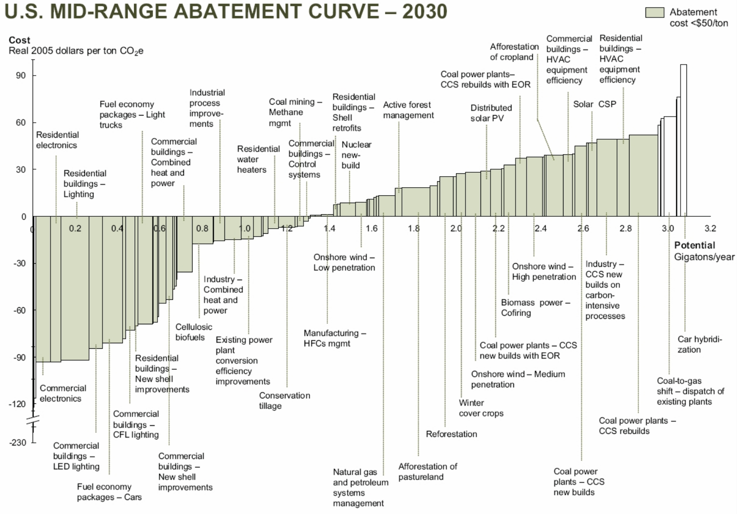
BGSE Summer Forum
June 22, 2018

Energy efficiency is a cornerstone of global climate policy...



CO_2 abatement	2025	2040
Energy service demand	16%	12%
End-use efficiency	53%	48%
Supply efficiency	3%	3%
Fuel and technology switching in end-uses	2%	2%
Renewables	19%	24%
Biofuels	2%	2%
Nuclear	4%	7%
CCS	1%	2%
Total (Gt CO_2)	**2.6**	**7.5**

...and appears to pay for itself



Source: McKinsey analysis

This paper

How effective are energy efficiency upgrades at reducing electricity consumption?

Context: $1 billion EE subsidy program in CA's K-12 schools

Data: 15' interval electricity consumption

Research design: Panel fixed effects meets machine learning

This question is difficult to answer empirically

Central challenge: Energy efficiency upgrades are not randomly assigned.

$\rightarrow$ It is difficult to disentangle energy efficiency from other factors.

$\rightarrow$ We must construct a counterfactual energy consumpton path.

This paper

How effective are energy efficiency upgrades at reducing electricity consumption?

Context: \$1 billion EE subsidy program in CA's K-12 schools

Data: 15' interval electricity consumption

Research design: Panel fixed effects meets machine learning

Energy efficiency at California schools

- In November 2012, CA voters passed Prop 39.
- Closed tax loophole, half the savings devoted to EE at schools.
- Projected to increase spending on EE at CA public K-12 schools by $.5 billion/year for 5 years.

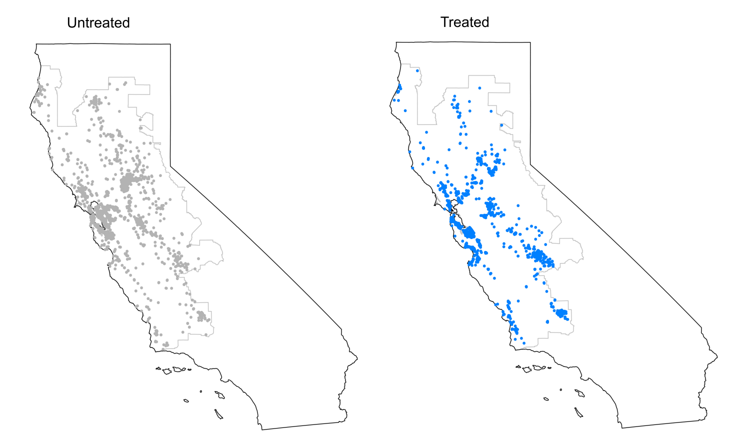
Schools are a useful “lab” for studying EE

- New commercial(ish) context for energy efficiency evaluation
  - Perhaps more optimized than residential? Or less?
- Thousands of public schools in CA.
- All using energy in roughly the same way.
- Availability of interval metering data.

Data

- We consider interventions that happen mostly before Prop 39 (Years 2008-2014).
- Schools in northern California, with 15' interval-metering data (aggregated to 3-hour blocks).
- About 1,000+ schools without treatment, 1000+ with treatment.

Our sample spans the PG&E territory



We consider a very standard panel FE approach

- We compare:
    - Consumption at schools that retrofitted to those that didn't.
    - Consumption before and after retrofits.
- We progressively add a series of control variables (school, hour and month-of-sample fixed effects, plus interactions):

$$Y_{ith} = \beta D_{it} + \alpha_i + \kappa_h + \gamma_t + \varepsilon_{ith}$$

Interpretation of β: Average reduction in KWh at treated schools.

Panel FE are sensitive to specification

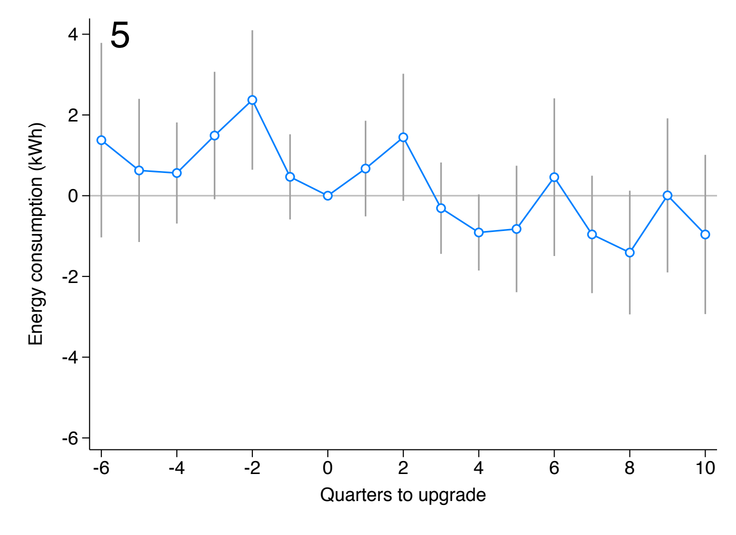
	(1)	(2)	(3)	(4)	(5)
Treat × post	-3.242	-3.243	-3.877	-2.238	-1.304
	(0.450)	(0.449)	(0.450)	(0.480)	(0.466)
Observations	19,193,084	19,193,084	19,192,744	19,192,744	19,193,084
Realization rate	0.728	0.728	0.871	0.503	0.293
School FE, Block FE	Yes	Yes	Yes	Yes	Yes
School-Block FE	No	Yes	Yes	Yes	Yes
School-Block-Month FE	No	No	Yes	Yes	No
Month of Sample Ctrl.	No	No	No	Yes	No
Month of Sample FE	No	No	No	No	Yes

We consider a range of (very saturated) models

Diff-in-Diff: Challenges

- Regression results suggest some percent savings, although magnitude sensitive to specification.
- Potential concerns:
    - Treatment correlated with time, maybe spurious.
    - Differential trends/shocks.
    - Confluence of interventions.
    - Sensitivity to outliers (regression in levels).
    - etc.

Panel FE approach fails an event study test



No sharp effects around interventions

Matching results are highly sensitive to pool

	(1)	(2)	(3)	(4)	(5)
Any district	-2.785	-2.697	-2.992	-0.658	-0.312
	(0.932)	(0.929)	(0.984)	(1.066)	(1.013)
Same district	-0.234	-0.166	-0.400	1.143	0.969
	(0.822)	(0.827)	(0.820)	(0.862)	(0.792)
Opposite district	-3.550	-3.550	-3.642	-0.442	-0.155
	(0.935)	(0.941)	(1.008)	(1.121)	(1.050)
Observations	6,043,052	6,043,046	6,042,653	6,042,653	6,043,046
School FE, Block FE	Yes	Yes	Yes	Yes	Yes
School-Block FE	No	Yes	Yes	Yes	Yes
School-Block-Month FE	No	No	Yes	Yes	No
Month of Sample Ctrl.	No	No	No	Yes	No
Month of Sample FE	No	No	No	No	Yes

A variety of matching choices lead to different results

Outliers appear to matter for regression in levels

	(1)	(2)	(3)
Any intervention	-1.285	-0.377	-0.183
	(0.455)	(0.350)	(0.326)
Observations	19,079,232	18,698,100	18,315,104
Realization rate	0.281	0.087	0.043
Trimming			
Dependent variable (1, 99)		X	
Dependent variable (2, 98)			X

Yet ATE in level easier to compare to expected engineering savings

Panel FE: Recap

- We find realization rates between 28%-83%.
- Results are quite sensitive to controls and outliers.
- Event study displays strong seasonality, even after substantially saturating the model.
- Traditional matching approaches appear sensitive to researcher choices.

Can machine learning help?

- Panel FE models aren't properly specified.
- Schools are very heterogeneous (e.g., climate, size, school calendar).
  - Ideally, introduce school-specific coefficients and trends in a very flexible manner.
- We easily came up with ∼6,000,000 candidate control variables by making them school-hour specific!
- No clear *ex ante* optimal choice.

Machine Learning: Advantages in this application

- Exogenous weather variation and predictable weekly and seasonal patterns drive variation in electricity consumption.
- Schools are relatively stable consumption units:
    - as opposed to single households that move around, unobservably buy a new appliance, expand family size, etc.
    - as opposed to businesses and manufacturing plants, exposed to macroeconomic shocks.

Machine Learning: Advantages in this application

- Exogenous weather variation and predictable weekly and seasonal patterns drive variation in electricity consumption.
- Schools are relatively stable consumption units:
  - as opposed to single households that move around, unobservably buy a new appliance, expand family size, etc.
  - as opposed to businesses and manufacturing plants, exposed to macroeconomic shocks.

Prediction can do well!

Machine Learning: Approach

Step 1

- Use *pre-treatment data* to predict electricity consumption as a function of flexible co-variates, *for each school separately*.

Machine Learning: Approach

Step 1

- Use *pre-treatment data* to predict electricity consumption as a function of flexible co-variates, *for each school separately*.
    - For control schools, determine a "pre-treatment period" randomly.

Machine Learning: Approach

Step 1

- Use *pre-treatment data* to predict electricity consumption as a function of flexible co-variates, *for each school separately*.
    - For control schools, determine a “pre-treatment period” randomly.
    - Use LASSO method (penalized regression).
        - Minimizing the sum of the squared errors plus $\lambda \cdot \sum_{j=1}^{p} |\beta_j|$.
        - Larger “tuning parameters” lead to fewer coefficients.
        - Use bootstrapped cross-validation with training and holdout samples *within pre-treatment*.

Machine Learning: Approach

Step 1

- Use *pre-treatment data* to predict electricity consumption as a function of flexible co-variates, *for each school separately*.
    - For control schools, determine a "pre-treatment period" randomly.
    - Use LASSO method (penalized regression).
        - Minimizing the sum of the squared errors plus $\lambda \cdot \sum_{j=1}^{p} |\beta_j|$.
        - Larger "tuning parameters" lead to fewer coefficients.
        - Use bootstrapped cross-validation with training and holdout samples *within pre-treatment*.
    - Include a wide range of school-specific variables, and also consumption at control schools (a la synthetic control).
    - Also consider other alternatives (random forests).

Machine Learning: Approach

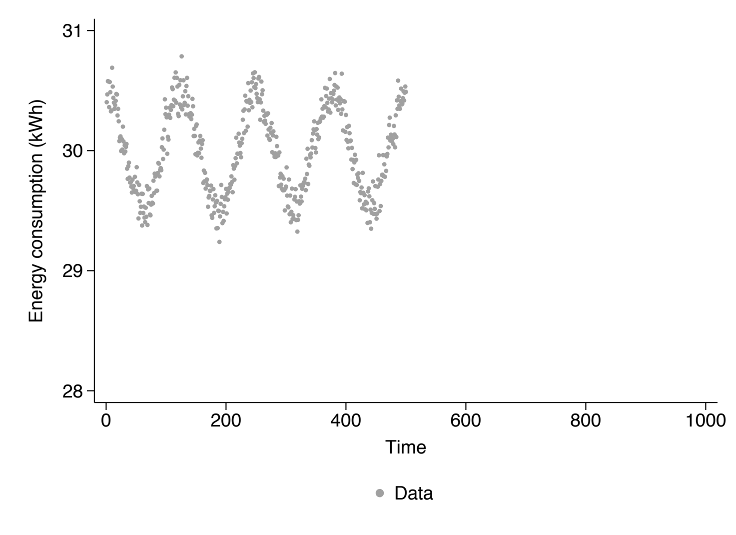
Step 2

- Regress *prediction errors* on treatment and controls.

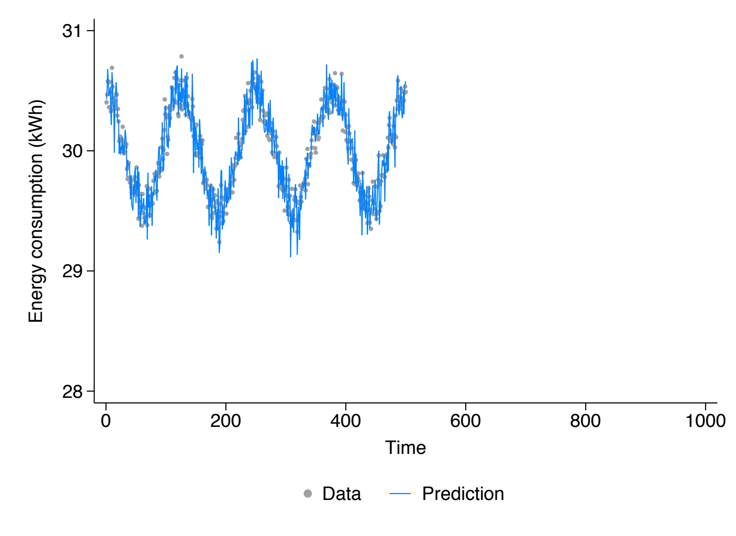
$$Y_{ith} = \beta D_{it} + \alpha_i + \kappa_h + \gamma_t + \varepsilon_{ith}$$

    - Data pooled across schools.
    - Replicates diff-in-diff approach, but Y variable is now the prediction error.

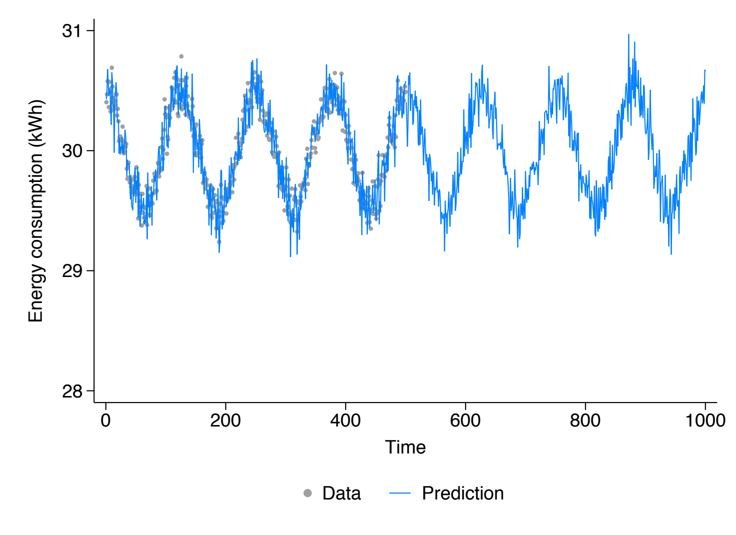
Machine Learning: Graphical intuition



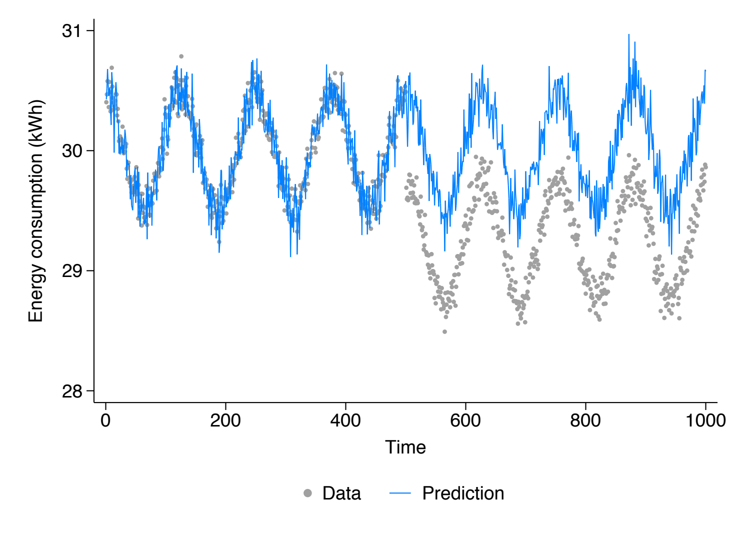
Machine Learning: Graphical intuition



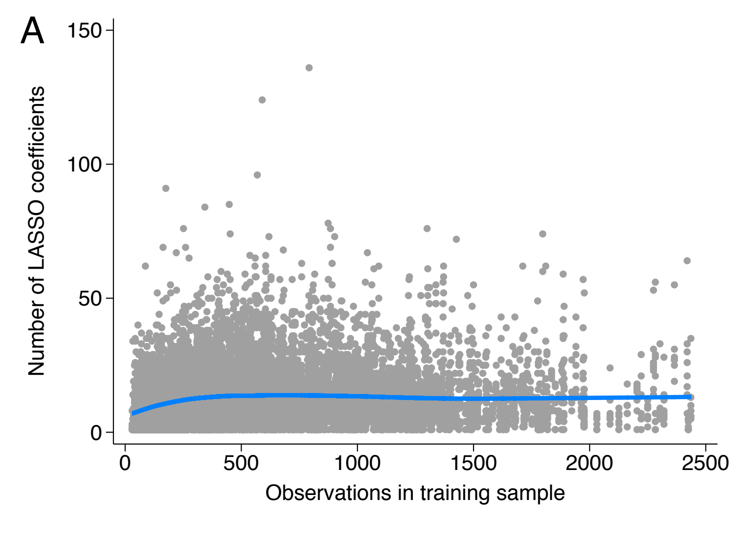
Machine Learning: Graphical intuition



Machine Learning: Graphical intuition

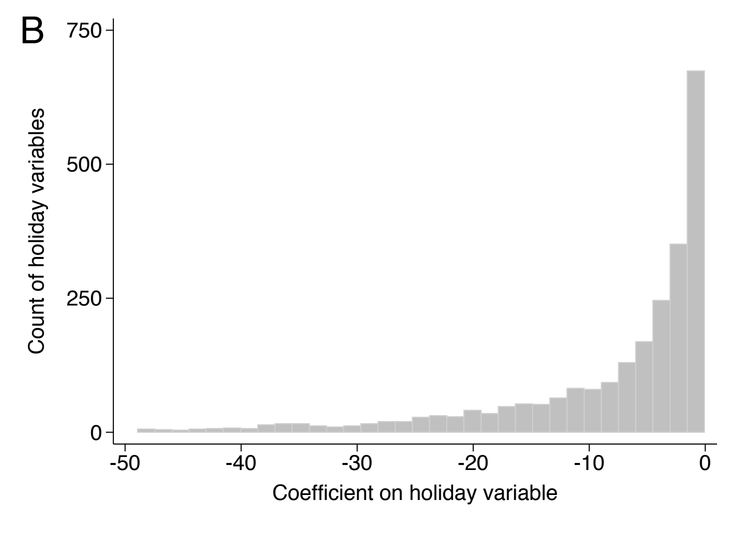


ML check: model complexity scales with observations



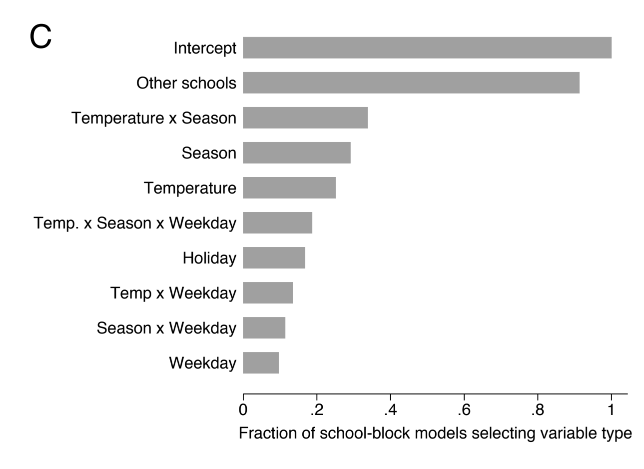
Step 1: Each dot is a school-hour model

ML check: holidays negatively correlated with energy use



Each observation is a school-specific holiday coefficient

ML check: a wide range of variables in model



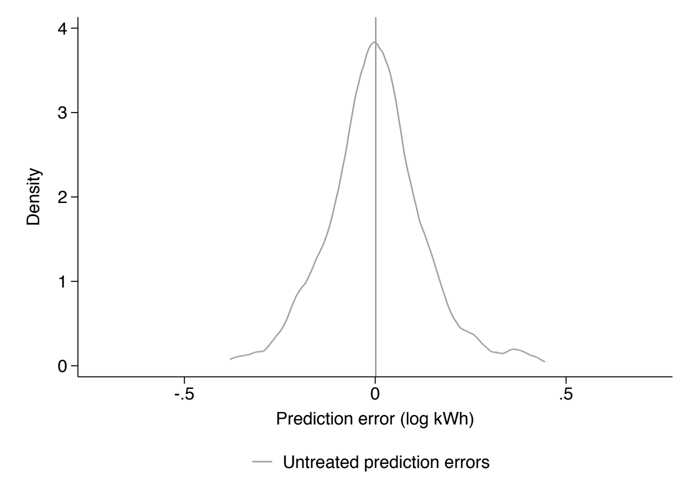
Each school-block model has different predictors

ML check: comparison across methods

	p10	p25	p50	p75	p90
LASSO optimal Lambda	-0.23	0.18	0.44	0.61	0.72
LASSO 1SE Lambda	0.08	0.21	0.43	0.58	0.69
LASSO + Synth optimal Lambda	-0.12	0.30	0.62	0.86	0.93
LASSO + Synth 1SE Lambda	0.13	0.33	0.64	0.86	0.93
LASSO Synth only optimal Lambda	-0.08	0.28	0.61	0.85	0.93
LASSO Synth only 1SE Lambda	0.12	0.32	0.63	0.85	0.93
Forest by school-block	0.09	0.30	0.52	0.67	0.76
Forest by school	-1.70	-0.15	0.42	0.63	0.71

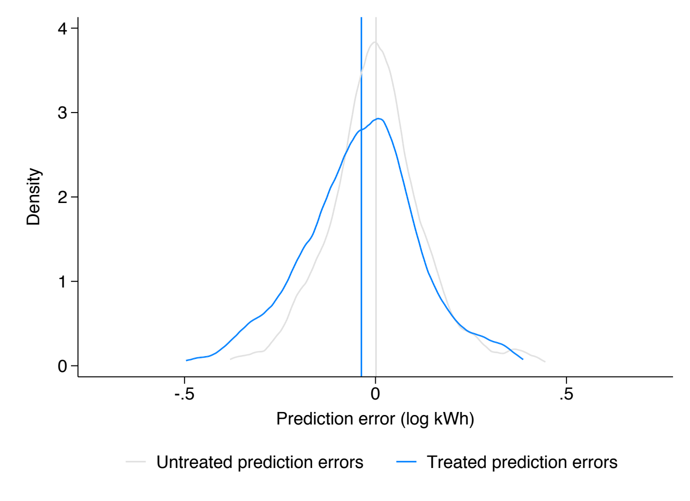
The LASSO with control schools appears to do well out-of-sample

ML check: control school predictions



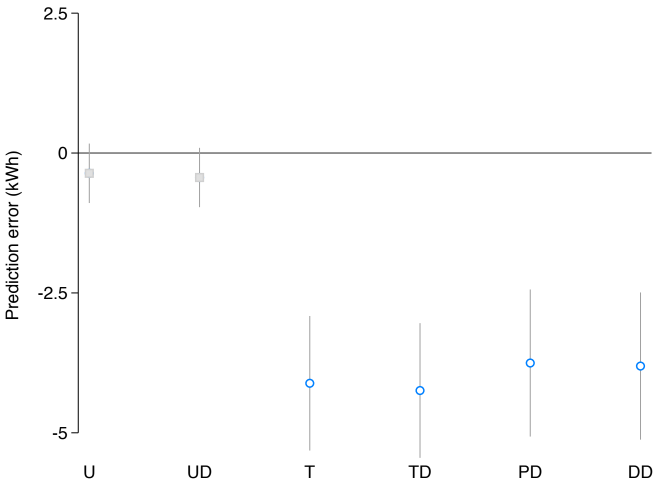
Prediction errors centered around zero well out-of-sample

Preliminary evidence: Treated schools reduce consumption



We see a shift in the distribution for schools with upgrades

ML results are stable across estimators



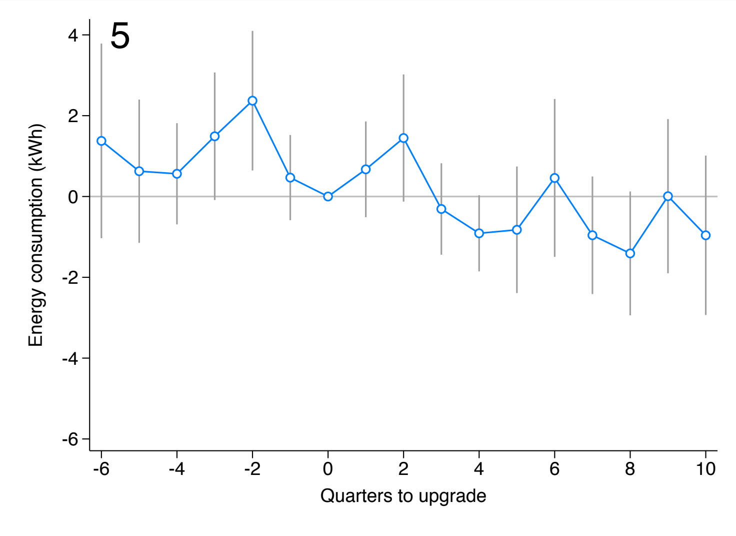
Step 1 unconditioned effects imply savings

ML results suggest larger savings

	(1)	(2)	(3)	(4)	(5)
Treat $\times$ post	-3.774	-3.787	-3.980	-3.134	-2.358
	(0.526)	(0.528)	(0.551)	(0.526)	(0.507)
Observations	19,193,084	19,193,084	19,192,744	19,192,744	19,193,084
Realization rate	0.848	0.850	0.894	0.704	0.530
School FE, Block FE	Yes	Yes	Yes	Yes	Yes
School-Block FE	No	Yes	Yes	Yes	Yes
School-Block-Month FE	No	No	Yes	Yes	No
Month of Sample Ctrl.	No	No	No	Yes	No
Month of Sample FE	No	No	No	No	Yes

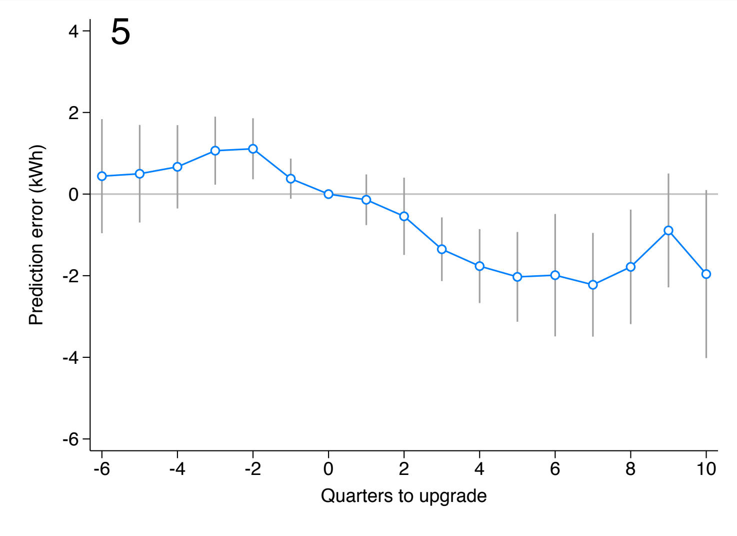
Step 2: Panel results confirm savings, more stable

ML approach passes the event study test



No sharp effects around interventions using diff-in-diff

ML approach passes the event study test



ML cleans up seasonality and noise more effectively

Outliers not an issue even with predictions out of sample!

	(1)	(2)	(3)
Any intervention	-2.281	-2.274	-2.204
	(0.500)	(0.312)	(0.259)
Observations	19,079,232	18,697,640	18,316,058
Realization rate	0.498	0.525	0.522
Trimming			
Dependent variable (1, 99)		X	
Dependent variable (2, 98)			X

Reassuring that our results are not driven by pesky outliers

Machine learning: Recap

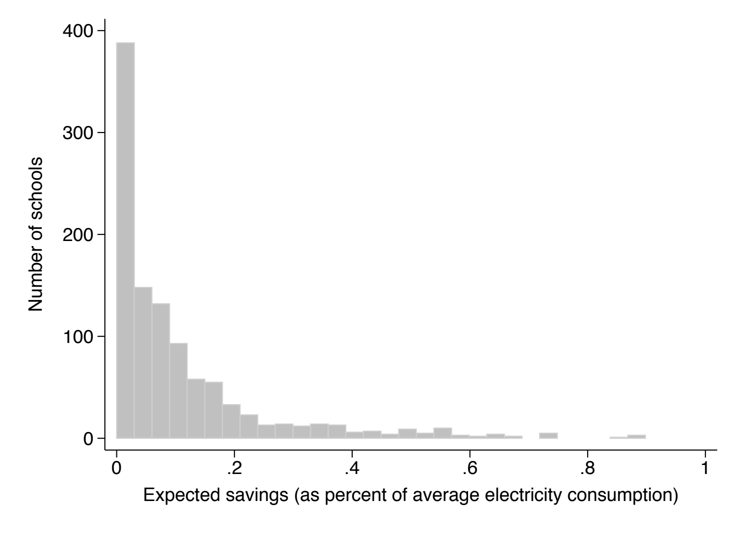
- Out-of-sample predictions appear to do well on average for control schools.
- We see a shift in prediction errors for treated schools.
- Results still sensitive, but more stable, with realization rates between 50-85%.

Discussion: Why such low realized savings?

Several potential explanations:

- General measurement error in expected savings?
  - Errors ins savings engineering model.
  - Timing of savings for which we have additional info.
- Large heterogeneity in realized savings?
  - Average effectiveness vs intervention-by-intervention.
  - Some interventions more effective than others.
  - Some interventions harder to predict.

Data on engineering savings have a long right tail



Measurement exercise

- Our baseline approach considers before-and-after effectiveness of policies on average (potentially less subject to measurement issues).
- Additionally:
    - Consider relationship between actual and expected savings at each project.
    - Consider the role of outliers.
    - Consider exploiting additional installation timing within schools or projects.

Alternative estimator

- We can compare actual v. expected reductions, by estimating:

$$Y_{ith} = -\beta S_{it} + \alpha_i + \kappa_h + \gamma_t + \epsilon_{ith},$$

  where S_{it} is the expected saving times the treatment date.
- $\beta = 1$ implies realizations match engineering estimates

Measurement issues and heterogeneity appear to be important

	(1)	(2)	(3)	(4)	(5)	(6)
Any intervention	0.498	0.525	0.114	0.438	0.628	0.081
	(0.109)	(0.120)	(0.104)	(0.127)	(0.098)	(0.213)
Savings regression				X	X	X
Expected savings trim		X			X	
Time-varying treatment			X			X

Timing seems to really affect measurement, attenuation bias intuitive

Measurement issues and heterogeneity appear to be important

	(1)	(2)	(3)	(4)	(5)	(6)
Any intervention	0.530	0.552	0.286	0.451	0.649	0.081
	(0.114)	(0.126)	(0.257)	(0.130)	(0.101)	(0.216)
HVAC interventions	0.753	0.690	1.039	0.397	0.721	0.286
	(0.359)	(0.396)	(0.326)	(0.222)	(0.139)	(0.170)
Lighting interventions	0.647	0.628	0.128	0.433	0.392	0.398
	(0.261)	(0.267)	(0.236)	(0.186)	(0.123)	(0.167)
Other interventions	0.420	0.459	-0.210	0.212	0.480	-0.251
	(0.233)	(0.281)	(0.517)	(0.133)	(0.307)	(0.343)
Observations	19,193,084	18,934,974	19,193,084	19,193,084	18,934,974	19,193,084
Savings regression				X	X	X
Expected savings trim		X			X	
Time-varying treatment			X			X

Up to 100% savings for some interventions, but only on average

Conclusions

We estimate the causal impact of energy efficiency

- We combine “big data” with machine learning tools.

Methods:

- New ML methods are useful for causal inference.

Findings:

- EE upgrades reduced consumption in schools by 3-5%...
- ...but only deliver $\approx$ 50% of expected savings.
- Targeting with available demographics is hard.

Energy efficiency underdelivers $\rightarrow$ extent depends on method

Thank you.

Questions? Comments?

mar.reguant@northwestern.edu