Online Appendix: Machine Learning from Schools about Energy Efficiency

Fiona Burlig, University of Chicago
Christopher Knittel, MIT
David Rapson, UC Davis
Mar Reguant, Northwestern University
Catherine Wolfram*, UC Berkeley

May 17, 2020

A Alternative machine learning estimators

As a complement to the discussion of our machine learning approach in Section 3.2, in this Appendix, we lay out additional treatment effect estimators that can be constructed using our machine learning methodology, leveraging different time periods and schools as controls.[1] Our ultimate goal is to compare our models' predictions of energy consumption with real energy use. In the absence of other confounding factors, the difference between our predicted counterfactual energy consumption and our data on electricity use would be the causal impact of energy efficiency upgrades, as shown in a graphical stylized representation in Figure A.1. Here, we present a series of estimators based on this idea, but designed to estimate treatment effects in the presence of time-varying changes in energy consumption.

We begin with a test of our method: we compute prediction errors—the average difference between the realized energy consumption and its prediction—at *untreated* schools in the post-"treatment" period:[2]

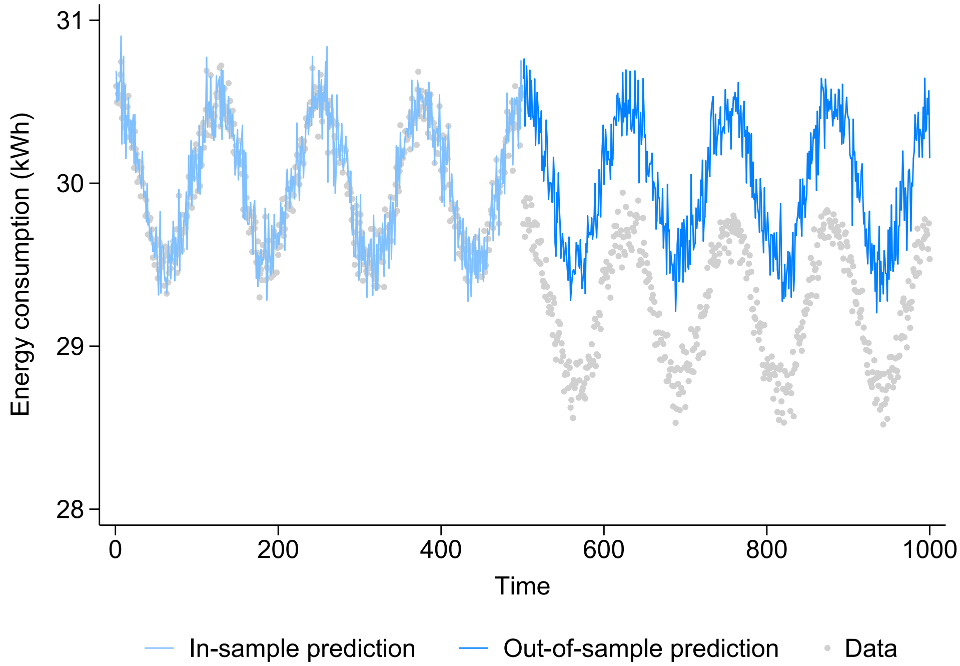
$$\hat{\beta}^U = \frac{1}{(1-P)I}\frac{1}{r}\sum_{i=PI+1}^{I}\sum_{t=1}^{r}(y_{it}-\hat{y}_{it})$$

*Burlig: Harris School of Public Policy and Energy Policy Institute, University of Chicago and NBER, burlig@uchicago.edu. Knittel: Sloan School of Management and Center for Energy and Environmental Policy Research, MIT and NBER, knittel@mit.edu. Rapson: Department of Economics, UC Davis, dsrapson@ucdavis.edu. Reguant: Department of Economics, Northwestern University, CEPR and NBER, mar.reguant@northwestern.edu. Wolfram: Haas School of Business and Energy Institute at Haas, UC Berkeley and NBER, cwolfram@berkeley.edu.

1. This section borrows heavily from a previous version of this paper circulated as NBER WP 23908.

2. Recall that we assigned every untreated school a random "treatment" date. We use only pre-"treatment" data to train untreated schools' models and validate our predictions out of sample.

Figure A.1: Machine learning approach: graphical intuition



Notes: This figure displays a stylized overview of how our machine learning approach works. We use the pre-treatment data only to fit a school-specific machine learning model of energy consumption (light blue line). We then use these models model to create fully out-of-sample predictions of counterfactual energy use in the post-treatment period (dark blue line). We compare the post-treatment counterfactuals to the actual data (gray points) to compute prediction errors. If the method is performing properly, these prediction errors will be close to zero in the untreated group. Non-zero prediction errors in the treatment group correspond to treatment effects.

where there are I total units in the sample, and P is the proportion of treated units, such that the first PI units are treated and the remaining $(1-P)I$ are untreated; there are $m+r$ total time periods, split into $[-m+1, 0]$ pre-treatment periods and $(0, r]$ post-treatment periods; y_{it} is realized energy consumption in school i at time t, and $\hat{y}_{it}$ is predicted energy consumption.[3] If the model has good performance out of sample, $\hat{\beta}^U$ should be zero in expectation. Figure A.2 displays the results of this estimator: our "treatment effect," a reduction in electricity consumption by about 0.23 kWh, which is not significant and relatively small.[4]

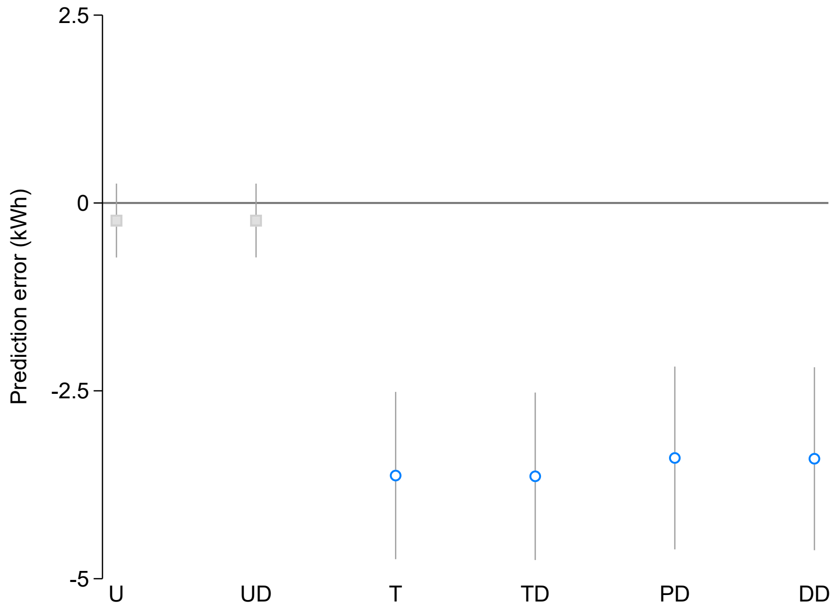
We may be concerned about systematic biases in the prediction model. To correct for potential biases in the predictions, we can extend our estimator to compare prediction errors in the post-treatment period with prediction errors in the pre-treatment period:

$$\hat{\beta}^{UD} = \hat{\beta}^U - \frac{1}{(1-P)I}\frac{1}{m}\sum_{i=PI+1}^{I}\sum_{t=-m+1}^{0}(y_{it} - \hat{y}_{it})$$

3. To avoid clutter, we do not include the 'i' subscripts on m and r (and generally abstract from issues associated with unbalanced panels) although those parameters differ by school, as described above.

4. For this and all estimators below, we cluster our standard errors at the school level.

Figure A.2: Comparing machine learning estimators



Notes: This figure shows average treatment effects, in the form of prediction errors based on electricity consumption in kWh per hour from a variety of different machine learning estimators. The effect marked U shows prediction errors (real energy consumption minus predicted energy consumption) in untreated schools in the post-treatment period only; UD presents prediction errors in the untreated group in the post-treatment period minus pre-period prediction errors for the untreated group. We expect these effects to be close to zero, as they use untreated schools only. The effect marked T presents prediction errors in treated schools in the post-treatment period only. TD presents prediction errors in the treated group in the post-treatment period minus pre-period prediction errors for the treated group. PD presents post-treatment-period prediction errors in the treated group minus post-treatment-period prediction errors in the untreated group. Finally, DD presents the prediction errors in the post- minus the pre-period for the treated group minus prediction errors in the post-minus pre-period for the untreated group. For all estimators, we cluster our standard errors at the school level.

Figure A.2 shows that after controlling for changes over time, $\hat{\beta}^{UD}$ yields a similar small decline in energy consumption, consistent with pre-treatment errors being close to zero as expected. These two tests provide suggestive evidence that our machine learning models are performing as expected.

We can now leverage predicted energy consumption to estimate treatment effects at treated schools. We begin with the simplest estimator:

$$\hat{\beta}^T = \frac{1}{PI}\frac{1}{r}\sum_{i=1}^{PI}\sum_{t=1}^{r}\left(y_{it} - \hat{y}_{it}\right),$$

which is analogous to $\hat{\beta}^U$, but with treated rather than untreated schools. If energy efficiency upgrades deliver savings, $\hat{\beta}^T$ should be negative, as predicted energy use, generated without any knowledge of the upgrade, will overestimate actual energy consumption. This is exactly what we see in Figure A.2: a treatment effect of 3.62 kWh, statistically significant at the 1 percent level.

As with the untreated schools, we can also compare treated schools to themselves over time:

$$\hat{\beta}^{TD} = \hat{\beta}^{T} - \frac{1}{PI}\frac{1}{m}\sum_{i=1}^{PI}\sum_{t=-m+1}^{0}(y_{it} - \hat{y}_{it})$$

We again expect this to be negative and similar to the previous result, and it is: Figure A.2 shows the treatment effect estimate of 3.63 kWh, statistically significant at the 1 percent level, confirming negligible errors in the pre-treatment.

To the extent that there are systematic differences between the prediction and the observed outcomes for untreated schools during the post period, e.g. due to underlying common trends, and to the extent that these differences reflect trends and biases in the predictive model that are common across schools, we can use these differences as a bias correction for the treated schools by estimating:

$$\hat{\beta}^{PD} = \hat{\beta}^{T} - \hat{\beta}^{U},$$

which yields a post-differenced corrected treatment effect estimate of 3.39 kWh (statistically significant at the 1 percent level and shown in Figure A.2) under the assumption of common trends and shocks between treated and untreated schools.

We can also estimate a "triple difference" that exploits the differences in predictions between treated and untreated schools during the pre- and post-period, by taking the differences of the before and after estimators at treated and untreated schools:

$$\hat{\beta}^{DD} = \hat{\beta}^{TD} - \hat{\beta}^{UD}.$$

This difference will tend to provide very similar results to those only using post data, as the corrections using pre-treatment data are relatively small. Using this estimator, we find that energy efficiency upgrades caused a 3.40 kWh reduction in energy consumption, significant at the 1 percent level, and shown in Figure A.2. Note that this triple difference relies on the same identifying assumptions as the panel fixed effects estimator described in the corresponding section above, namely, that conditional on covariates, treated and untreated schools are trending similarly. The key difference is that for this estimator to be identified, we need treated and untreated schools to be trending similarly in *prediction errors*, rather than in energy consumption.

Taken together, these results suggest that our machine learning method is delivering causal estimates of the impact of energy efficiency on electricity use. Estimates for untreated schools are close to zero, as expected, while the estimates for treated schools find treatment effects between 3.40 and 3.63 kWh, consistent with specifications (1) and (2) of Table 4, which are the closest analogue to this exercise.

B Heterogeneous realization rates across schools

To explore whether heterogeneity in realization rates is related to school characteristics, we proceed in two steps. First, we compute school-specific realization rate estimates, as described in Section 4. Next, we project these estimates onto covariates that are readily available to policymakers.

We do this by regressing our school-specific treatment effects onto a variety of covariates via quantile regression, in order to remove the undue influence of outliers in these noisy estimates.[5] We include one observation per treated school in our sample, and weight the observations by the length of the time series of energy data for each school.[6] We center all variables (except for dummy variables) around their mean and normalize by their standard deviation.

Appendix Table B.1 presents the results of this exercise. Column (1) shows that the median realization rate for treated schools using this approach is close to 74 percent. Column (2) shows that median realization rates are larger for HVAC and lighting interventions (the most prevalent types of upgrades in our sample), although the estimates are very noisy. We add latitude, longitude, a coastal climate zone indicator, and temperature in Column (3). Columns (4)-(5) control for standardized values of yet more covariates, including school enrollment, the Academic Performance Index and the poverty rate. Except for coastal schools having lower realization rates, we find no other statistically significant correlations between observable characteristics and realization rates.[7] These descriptive regressions should be interpreted with caution. These are cross-sectional estimates, and school size is likely correlated with a variety of other important factors.

5. Note that we could also have used a quantile regression approach in our high-frequency data, which would assuage potential concerns about outliers. Because we rely on a large set of high-dimensional fixed effects for identification, however, this is computationally intractable.

6. Note that untreated schools are not included in these regressions.

7. We explored a variety of other potential demographic variable, the size of the interventions, and interactions between coastal schools and HVAC interventions, but we did not find any clear correlation with realization rates.

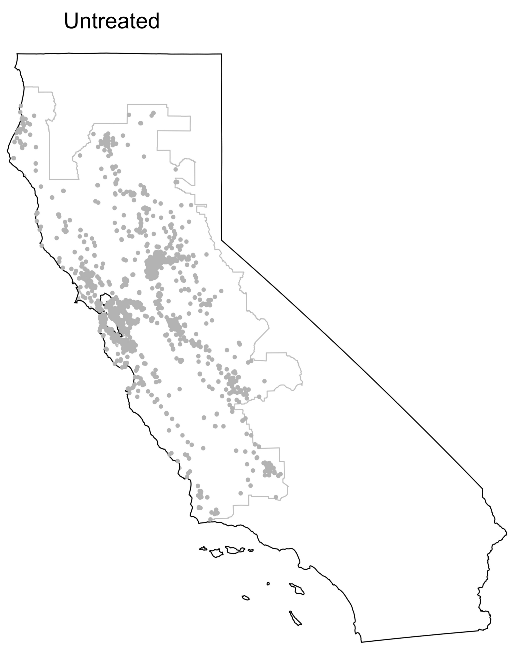
Table B.1: Predicting heterogeneous effects

Variable	(1)	(2)	(3)	(4)	(5)
Constant	0.74	0.48	0.53	0.66	0.57
	(0.10)	(0.29)	(0.31)	(0.32)	(0.31)
HVAC only (0/1)		0.30	0.32	0.15	0.37
		(0.35)	(0.38)	(0.39)	(0.39)
Lighting only (0/1)		0.32	0.16	0.13	0.31
		(0.38)	(0.41)	(0.42)	(0.41)
HVAC and Lighting (0/1)		0.33	0.19	0.01	0.06
		(0.37)	(0.41)	(0.42)	(0.41)
Coastal (0/1)			-0.31	-0.35	-0.48
			(0.17)	(0.18)	(0.18)
Longitude			0.08	0.10	-0.16
			(0.29)	(0.30)	(0.30)
Latitude			0.04	0.04	-0.09
			(0.22)	(0.23)	(0.22)
Average temperature (° F)			-0.32	-0.38	-0.27
			(0.26)	(0.27)	(0.26)
Total enrollment				0.18	0.14
				(0.12)	(0.11)
Academic perf. index (200-1000)					-0.20
					(0.15)
Poverty rate					-0.05
					(0.15)
Number of schools	838	838	818	783	765

Notes: This table presents results from median regressions of school-specific realization rates on a variety of covariates. The school-specific realization rates are estimated from a regression of prediction errors (in kWh) on school-specific treatment indicators and school-by-hour-by-month fixed effects. This table presents results for treated schools only. All estimates are weighted by the number of observations at each school. All variables (except dummy variables) are normalized to have mean zero and a standard deviation of one. Standard errors, robust to heteroskedasticity, are in parentheses.

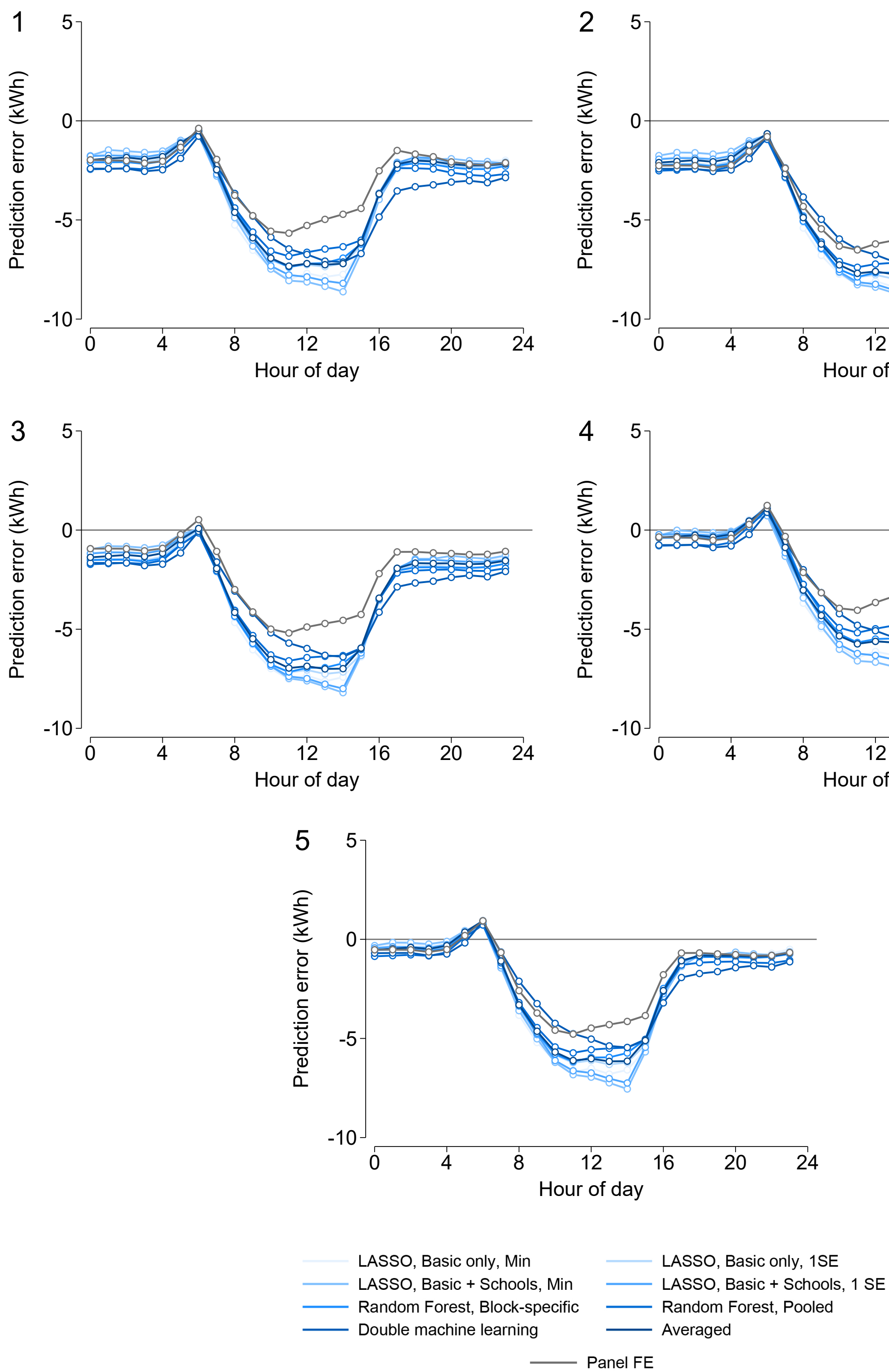
C Supplemental tables and figures

Figure C.1: Locations of untreated and treated schools



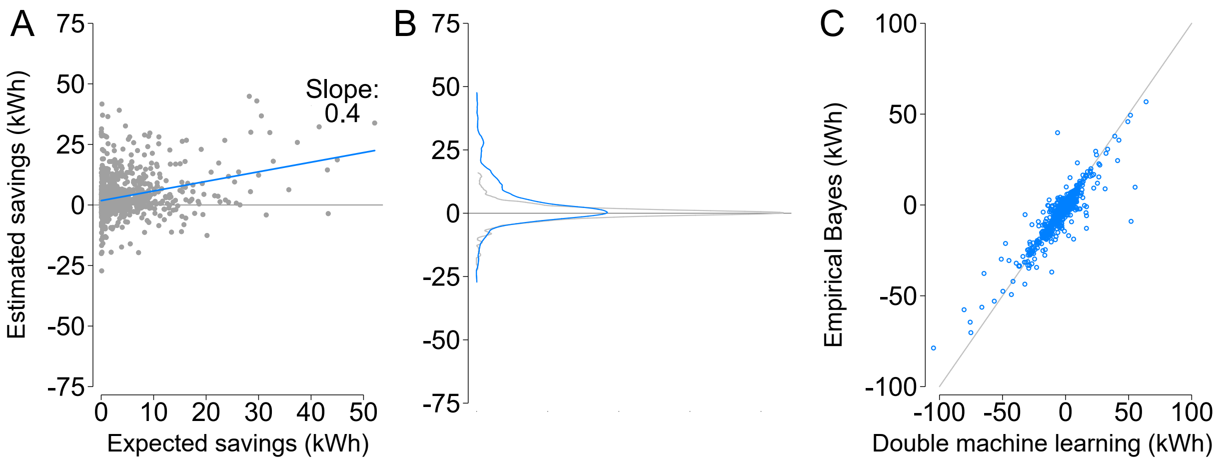
Notes: This figure displays the locations of schools in our sample. "Untreated" schools, in gray on the left, did not undertake any energy efficiency upgrades during our sample period. "Treated" schools, in blue on the right, installed at least one upgrade during our sample. The light gray outline shows the PG&E service territory.

Figure C.2: Machine learning results by hour (alternative prediction methods)



Notes: This figure presents treatment effects for each hour of the day estimated using prediction errors based on electricity consumption in kWh as the dependent variable. Here, we present results from 9 different estimation procedures: LASSOs with, without, and exclusively using other schools' consumption as candidate variables using a larger and smaller tuning parameter; random forests with and without imposing hour-specific branches; our double machine learning procedure; the average over all non-double machine learning predictions; and the panel fixed effects analogue. Each panel corresponds to one column of Table 4.

Figure C.3: School-specific effects with double machine learning



Notes: This figure displays school-specific savings estimates. We generate these estimates by regressing prediction errors in kWh onto prediction errors for treatment date, following a double machine learning procedure with two sample splits as in Chernozhukov et al. (2018). The coefficients are the average across the two samples. Panel A compares estimated savings with expected savings among treated schools only. Panel B displays kernel densities of estimated savings in the untreated group (gray line) and estimated savings in the treated group (blue line). Panel C plots the correlation between estimated savings generated using the empirical Bayes method described in the main text and estimated savings generated using the double machine learning approach.

Table C.1: Effects of bond measures on energy use in untreated schools

	(1)	(2)	(3)	(4)	(5)
Bond × post	-0.35	0.42	0.63	0.69	0.57
	(0.64)	(0.70)	(0.74)	(0.67)	(0.76)
Observations	20,860,080	20,858,880	20,858,880	20,860,080	20,858,880
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating a variation of Equation (3.1), with hourly energy consumption in kWh as the dependent variable. Rather than using energy efficiency upgrades as a treatment variable, we instead use a treatment indicator for school district bond measures, set equal to 1 for schools in districts with bonds after the passage of a bond, and 0 otherwise: $Y_{ith} = \beta\ Bond\_Passed_{dt} + \alpha_{ith} + \varepsilon_{ith}$, where $Bond\_Passed_{dt}$ is an indicator equal to 1 after district d (to which school i belongs) passed a bond and zero otherwise, and α_{ith} are a range of fixed effects to control for confounders. We estimate these effects only on schools that did not undergo an energy efficiency upgrade during our sample period. Standard errors, clustered at the school level, are in parentheses.

Table C.2: Panel fixed effects results (all hourly)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Average program estimates						
Realization rate	0.68	0.81	0.52	0.31	0.43	0.41
Point estimate	-2.88	-3.47	-2.15	-1.26	-1.74	-1.65
	(0.44)	(0.44)	(0.46)	(0.45)	(0.47)	(0.46)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360	57,480,360
Panel B: Average school-specific estimates						
Realization rate	0.51	0.57	0.50	0.39	0.45	0.46
	(0.12)	(0.12)	(0.12)	(0.11)	(0.12)	(0.12)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp Ctrl	No	No	No	No	No	Yes

Notes: This table presents an analog to Table 2 using hourly data instead of month-hour weighted collapsed data. Panel A in this table reports results from estimating Equation (3.1), with hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses. Realization rates are calculated by dividing the regression results on a complementary regression of *ex ante* engineering energy savings where expected (and zero otherwise) on our treatment variable, where we include the same set of controls and fixed effects. Panel B reports results from estimating Equation (3.2), in which the independent variable equals (the negative of) average expected savings for treated schools after their first upgrade, and 0 otherwise.

Table C.3: Panel fixed effects results (alternative standard errors)

Clustering	(1)	(2)	(3)	(4)	(5)	(6)
	-2.88	-3.47	-2.15	-1.26	-1.74	-1.60
School	(0.44)	(0.44)	(0.46)	(0.45)	(0.47)	(0.45)
School, month of sample	[1.56]	[0.71]	[0.51]	[0.51]	[0.48]	[0.47]
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp. Ctrl	No	No	No	No	No	Yes

Notes: This table reports results from estimating Equation (3.1), with hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. This table shows two variations on clustered standard errors: errors clustered at the school level, as in the main text, in parentheses; and errors clustered at the school and month-of-sample level, in brackets.

Table C.4: Sensitivity of panel fixed effects results to outliers (average school-specific estimates)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Trim outlier observations						
Realization rate	0.35	0.39	0.31	0.24	0.27	0.24
	(0.07)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)
Observations	56,323,212	56,321,525	56,321,525	56,323,212	56,321,525	56,321,525
Panel B: Trim outlier schools						
Realization rate	0.69	0.76	0.64	0.53	0.58	0.58
	(0.08)	(0.09)	(0.09)	(0.08)	(0.09)	(0.10)
Observations	56,737,632	56,736,096	56,736,096	56,737,632	56,736,096	56,736,096
Panel C: Trim observations and schools						
Realization rate	0.50	0.57	0.47	0.37	0.42	0.40
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Observations	55,689,089	55,687,427	55,687,427	55,689,089	55,687,427	55,687,427
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp. Ctrl	No	No	No	No	No	Yes

Notes: This table reports results from estimating Equation (3.2), with hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to individual expected savings for treated schools after their first upgrade, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses. In Panel A, we drop observations below the 1st or above the 99th percentile of the dependent variable: energy consumption. In Panel B, we drop schools below the 1st or above the 99th percentile of expected savings. In Panel C, we drop both.

Table C.5: Panel fixed effects estimates (donuts, average program estimates)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Drop ± 1 month						
Realization rate	0.68	0.82	0.53	0.33	0.44	0.40
Point estimate	-2.93	-3.53	-2.19	-1.32	-1.78	-1.67
	(0.45)	(0.45)	(0.48)	(0.47)	(0.49)	(0.47)
Observations	56,170,128	56,168,424	56,168,424	56,170,128	56,168,424	56,168,424
Panel B: Drop ± 2 months						
Realization rate	0.72	0.85	0.57	0.37	0.47	0.43
Point estimate	-3.14	-3.73	-2.40	-1.53	-1.94	-1.84
	(0.47)	(0.47)	(0.52)	(0.51)	(0.53)	(0.52)
Observations	53,593,008	53,591,088	53,591,088	53,593,008	53,591,088	53,591,088
Panel C: Drop ± 3 months						
Realization rate	0.77	0.88	0.62	0.46	0.51	0.46
Point estimate	-3.44	-3.96	-2.66	-1.91	-2.13	-1.99
	(0.49)	(0.48)	(0.55)	(0.55)	(0.57)	(0.56)
Observations	51,056,424	51,054,408	51,054,408	51,056,424	51,054,408	51,054,408
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp. Ctrl	No	No	No	No	No	Yes

Notes: This table reports results from estimating Equation (3.1), with hourly energy consumption in kWh as the dependent variable, and an indicator equal to 1 for schools after their first energy efficiency upgrade and 0 otherwise. In each panel, we drop a number of months immediately before and after treatment: Panel A drops 1 month before and after, Panel B drops 2 months before and after, and Panel C drops three months before and after.Standard errors, clustered at the school level, are in parentheses.

Table C.6: Panel fixed effects estimates (donuts, school-specific estimates)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Drop ± 1 month						
Realization rate	0.53	0.59	0.51	0.41	0.46	0.46
	(0.12)	(0.13)	(0.12)	(0.12)	(0.12)	(0.13)
Observations	56,170,128	56,168,424	56,168,424	56,170,128	56,168,424	56,168,424
Panel B: Drop ± 2 months						
Realization rate	0.55	0.61	0.53	0.43	0.48	0.49
	(0.12)	(0.13)	(0.13)	(0.12)	(0.13)	(0.14)
Observations	53,593,008	53,591,088	53,591,088	53,593,008	53,591,088	53,591,088
Panel C: Drop ± 3 months						
Realization rate	0.59	0.64	0.56	0.47	0.50	0.50
	(0.13)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)
Observations	51,056,424	51,054,408	51,054,408	51,056,424	51,054,408	51,054,408
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp. Ctrl	No	No	No	No	No	Yes

Notes: This table reports results from estimating Equation (3.2), with hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to individual expected savings for treated school after their first upgrade, and 0 otherwise. In each panel, we drop a number of months immediately before and after treatment: Panel A drops 1 month before and after, Panel B drops 2 months before and after, and Panel C drops three months before and after.Standard errors, clustered at the school level, are in parentheses.

Table C.7: Panel fixed effects results (continuous treatment timing)

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Average program estimates						
Realization rate	0.35	0.39	0.17	0.13	0.16	0.27
Point estimate	-0.78	-0.89	-0.40	-0.33	-0.39	-0.60
	(0.35)	(0.35)	(0.46)	(0.47)	(0.47)	(0.38)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360	57,480,360
Panel B: Average school-specific estimates						
Realization rate	0.12	0.15	0.06	0.03	0.05	0.17
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.18)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes	Yes
Time trend	No	No	Yes	No	No	No
Month of Sample FE	No	No	No	Yes	Yes	Yes
Temp Ctrl	No	No	No	No	No	Yes

Notes: This table reports results from estimating Equation (3.1) and Equation (3.2), with hourly energy consumption in kWh as the dependent variable. The independent variable in Panel A is a treatment indicator, set equal to the percentages of performed upgrades for treatment schools, and 0 otherwise. The independent variable in Panel B is a treatment indicator, set equal to individual cumulative expected savings for treated schools, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses.

Table C.8: Matching results

	(1)	(2)	(3)	(4)	(5)
Any district	-2.70	-2.99	-0.66	-0.31	-0.47
	(0.93)	(0.98)	(1.07)	(1.01)	(1.13)
Same district	-0.17	-0.40	1.14	0.97	0.92
	(0.83)	(0.82)	(0.86)	(0.79)	(0.84)
Opposite district	-3.55	-3.64	-0.44	-0.16	-0.03
	(0.94)	(1.01)	(1.12)	(1.05)	(1.18)
Observations	6,043,046	6,042,653	6,042,653	6,043,046	6,042,653
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.1), with hourly energy consumption in kWh as the dependent variable. As above, the independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. The untreated group in these regressions is chosen via nearest-neighbor matching. In particular, we match one untreated school to each treated school. Each row in the table employs a different restriction on which schools are allowed to be matched to any given treatment school. "Any district" matches allow any untreated school to be matched to a treatment school; "same district" matches are restricted to untreated schools in the same school district, and "opposite district" matches are restricted to untreated schools from different districts. In each case, the matching variables are the mean, maximum, and standard deviation of electricity consumption in three hour blocks (e.g., 9 AM-Noon) from the pre-treatment period; demographic variables measured at the census block level, including the poverty rate, log of per capita income, school-level variables (enrollment; age of the school; grades taught; an academic performance index; and climate). These estimates are relatively sensitive to which schools are included. Standard errors, clustered at the school level, are in parentheses.

Table C.9: R^2s of prediction models across machine learning methods

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
10th percentile	-0.04	0.08	0.00	0.12	0.10	-1.81	-1.27	0.09
25th percentile	0.21	0.22	0.29	0.30	0.27	-0.19	0.05	0.29
50th percentile	0.47	0.44	0.59	0.58	0.47	0.39	0.52	0.56
75th percentile	0.67	0.62	0.83	0.82	0.63	0.61	0.66	0.75
90th percentile	0.76	0.73	0.92	0.91	0.72	0.70	0.73	0.83
Method	LASSO	LASSO	LASSO	LASSO	RF	RF	DML	AVG
Hour-specific model	X	X	X	X	X			
Untreated schools $-i$			X	X				
Tuning parameter	Min	1SE	Min	1SE				

Notes: This table reports the R^2 of the prediction models for untreated schools during the post-treatment period. As that these predictions are completely out-of-sample, and therefore extreme outliers may be a concern, we present the distribution of the R^2. Columns 1 through 4 display predictions generated via LASSO, while Columns 5 and 6 show predictions generated using a random forest algorithm. Column 7 uses an alternative double machine learning approach with forests, and Column 8 averages across all models except for Column 7. In all but Column 6 and 7, we generate prediction models for each school-hour separately. All models include as basic variables day of the week, a holiday dummy, a seasonal spline, a temperature spline, and all of their their multi-way interactions. In Columns 3 and 4, we include energy consumption at all (other) untreated schools as candidate variables. For the LASSO estimates, we report results for two tuning parameters: "Min," which minimizes the root mean squared error, or "1SE," which chooses a more parsimonious model than Min, but which has a root mean squared error that remains within one standard error of Min. Overall, we find that the LASSO model where we allow for both basic variables and untreated school consumption, with a 1SE tuning parameter, provides the best overall fit. Note that some of the R^2s are negative. This is entirely possible here, as these are fully-out-of-sample predictions.

Table C.10: Machine learning results (all hourly)

	(1)	(2)	(3)	(4)	(5)
Panel A: Average program estimates					
Realization rate	0.86	0.92	0.75	0.53	0.60
Point estimate	-3.64	-3.92	-3.17	-2.10	-2.42
	(0.50)	(0.52)	(0.49)	(0.47)	(0.49)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360
Panel B: Average school-specific estimates					
Realization rate	0.57	0.58	0.55	0.50	0.50
	(0.13)	(0.14)	(0.14)	(0.13)	(0.13)
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table presents an analog to Table 4 using hourly data instead of month-hour weighted collapsed data. Panel A in this table reports results from estimating Equation (3.3), with prediction errors in hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses. Realization rates are calculated by dividing the regression results on a complementary regression of ex-ante engineering energy savings where expected (and zero otherwise) on our treatment variable, also including the same set of controls. Panel B reports results from estimating Equation (3.4), in which the independent variable equals (the negative of) average expected savings for treated schools after their first upgrade, and 0 otherwise. All regressions include a control for being in the post-training period for the machine learning.

Table C.11: Machine learning results (alternative standard errors)

Clustering	(1)	(2)	(3)	(4)	(5)
	-3.92	-4.23	-2.22	-1.68	-1.87
School	(0.44)	(0.44)	(0.45)	(0.45)	(0.47)
School, month of sample	[0.71]	[0.61]	[0.47]	[0.50]	[0.48]
Observations	57,481,920	57,480,360	57,480,360	57,481,920	57,480,360
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.3), with prediction errors in hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. This table shows two variations on clustered standard errors: errors clustered at the school level, as in the main text, in parentheses; and errors clustered at the school and month-of-sample level, in brackets. All regressions include a control for being in the post-training period for the machine learning.

Table C.12: Machine learning results (bootstrap)

	(1)	(2)	(3)	(4)	(5)
Average program estimates					
Realization rate	0.84	0.89	0.73	0.55	0.61
Point estimate	-3.60	-3.85	-3.12	-2.21	-2.46
	(0.50)	(0.53)	(0.51)	(0.48)	(0.50)
Observations	57,580,749	57,580,749	57,580,749	57,580,749	57,580,749
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.3), with prediction errors in hourly energy consumption in kWh as the dependent variable. We generate these estimates using 50 bootstrap runs, created by first bootstrapping weeks of data to input into the school-specific LASSO models, and then bootstrapping schools to include into the final regression. The independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. All regressions include a control for being in the post-training period for the machine learning.

Table C.13: Sensitivity of machine learning results to outliers (average school-specific estimates)

	(1)	(2)	(3)	(4)	(5)
Panel A: Trim outlier observations					
Realization rate	0.50	0.51	0.39	0.35	0.36
	(0.09)	(0.10)	(0.09)	(0.08)	(0.08)
Observations	56,332,278	56,330,677	56,330,677	56,332,278	56,330,677
Panel B: Trim outlier schools					
Realization rate	0.83	0.86	0.69	0.62	0.64
	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
Observations	56,737,632	56,736,096	56,736,096	56,737,632	56,736,096
Panel C: Trim observations and schools					
Realization rate	0.68	0.70	0.54	0.49	0.50
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Observations	55,673,654	55,672,077	55,672,077	55,673,654	55,672,077
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.4), with hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to individual expected savings for treated schools after their first upgrade, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses. In Panel A, we drop observations below the 1st or above the 99th percentile of the dependent variable: energy consumption. In Panel B, we drop schools below the 1st or above the 99th percentile of expected savings. In Panel C, we drop both.

Table C.14: Machine learning estimates (donuts, average program estimates)

	(1)	(2)	(3)	(4)	(5)
Panel A: Drop ± 1 month					
Realization rate	0.87	0.93	0.77	0.55	0.62
Point estimate	-3.72	-4.00	-3.26	-2.19	-2.51
	(0.51)	(0.53)	(0.51)	(0.49)	(0.51)
Observations	56,170,128	56,168,424	56,168,424	56,170,128	56,168,424
Panel B: Drop ± 2 months					
Realization rate	0.89	0.95	0.80	0.58	0.65
Point estimate	-3.90	-4.18	-3.45	-2.36	-2.67
	(0.53)	(0.56)	(0.54)	(0.52)	(0.55)
Observations	53,593,008	53,591,088	53,591,088	53,593,008	53,591,088
Panel C: Drop ± 3 months					
Realization rate	0.92	0.98	0.83	0.63	0.68
Point estimate	-4.11	-4.37	-3.68	-2.60	-2.86
	(0.56)	(0.59)	(0.57)	(0.56)	(0.58)
Observations	51,056,424	51,054,408	51,054,408	51,056,424	51,054,408
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.3), with prediction errors in hourly energy consumption in kWh as the dependent variable, and an indicator equal to 1 for schools after their first energy efficiency upgrade and 0 otherwise. In each panel, we drop a number of months immediately before and after treatment: Panel A drops 1 month before and after, Panel B drops 2 months before and after, and Panel C drops three months before and after treatment. Standard errors, clustered at the school level, are in parentheses.

Table C.15: Machine learning estimates (donuts, school-specific estimates)

	(1)	(2)	(3)	(4)	(5)
Panel A: Drop ± 1 month					
Realization rate	0.58	0.59	0.56	0.51	0.51
	(0.14)	(0.14)	(0.14)	(0.13)	(0.14)
Observations	56,170,128	56,168,424	56,168,424	56,170,128	56,168,424
Panel B: Drop ± 2 months					
Realization rate	0.60	0.61	0.59	0.52	0.54
	(0.14)	(0.15)	(0.14)	(0.14)	(0.14)
Observations	53,593,008	53,591,088	53,591,088	53,593,008	53,591,088
Panel C: Drop ± 3 months					
Realization rate	0.62	0.63	0.61	0.55	0.56
	(0.15)	(0.15)	(0.15)	(0.14)	(0.15)
Observations	51,056,424	51,054,408	51,054,408	51,056,424	51,054,408
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.4), with prediction errors in hourly energy consumption in kWh as the dependent variable. The independent variable is a treatment indicator, set equal to individual expected savings for treated school after their first upgrade, and 0 otherwise. In each panel, we drop a number of months immediately before and after treatment: Panel A drops 1 month before and after, Panel B drops 2 months before and after, and Panel C drops three months before and after.Standard errors, clustered at the school level, are in parentheses.

Table C.16: Machine learning results (continuous treatment timing)

	(1)	(2)	(3)	(4)	(5)
Panel A: Average program estimates					
Realization rate	0.38	0.39	0.25	0.23	0.24
Point estimate	-0.91	-0.94	-0.61	-0.55	-0.58
	(0.44)	(0.47)	(0.49)	(0.47)	(0.49)
Observations	55,822,576	55,821,180	55,821,180	55,822,576	55,821,180
Panel B: Average school-specific estimates					
Realization rate	0.15	0.15	0.10	0.09	0.09
	(0.23)	(0.23)	(0.22)	(0.22)	(0.22)
Observations	55,822,576	55,821,180	55,821,180	55,822,576	55,821,180
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports results from estimating Equation (3.1) and Equation (3.2), with prediction errors in hourly energy consumption in kWh as the dependent variable. The independent variable in Panel A is a treatment indicator, set equal to the percentages of performed upgrades for treatment schools, and 0 otherwise. The independent variable in Panel B is a treatment indicator, set equal to individual cumulative expected savings for treated schools, and 0 otherwise. Standard errors, clustered at the school level, are in parentheses.

Table C.17: Machine learning results (alternative prediction methods)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat × post	-2.37	-2.31	-2.46	-2.42	-2.25	-2.31	-2.20	-2.24
	(0.50)	(0.50)	(0.49)	(0.49)	(0.50)	(0.50)	(0.50)	(0.49)
Realization rate	0.59	0.58	0.61	0.60	0.56	0.58	0.54	0.56
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Method	LASSO	LASSO	LASSO	LASSO	RF	RF	DML	AVG
Hour-specific model	X	X	X	X	X			
Untreated schools $-i$			X	X				
Tuning parameter	Min	1SE	Min	1SE				

Notes: This table reports results from estimating Equation (3.3), with prediction errors in hourly energy consumption in kWh as the dependent variable. All regressions include school-by-hour and month-of-sample fixed effects. Each column displays results from a different prediction approach. Columns 1 through 4 display predictions generated via LASSO, while Columns 5 and 6 show predictions generated using a random forest algorithm. Column 7 uses an alternative double machine learning approach with forests as in (6) and column 8 averages across all models except for Column 7. In all but Column 6 and 7, we generate prediction models for each school-hour separately. All models include as variables day of the week, a holiday dummy, a seasonal spline, a temperature spline, and all of their their multi-way interactions. In Columns 3 and 4, we include energy consumption at all (other) untreated schools as candidate variables. For the LASSO estimates, we report results for two tuning parameters: "Min," which minimizes the root mean squared error, or "1SE," which chooses a slightly more parsimonious model than Min, but which has a root mean squared error that remains within one standard error of Min. In all cases, the independent variable is a treatment indicator, set equal to 1 for treated schools after their first upgrade, and 0 otherwise. Regressions include a control for being in the post-training period (or the prediction error of the post variable in the case of double machine learning). Standard errors are clustered at the school level.

Table C.18: Double machine learning results

	(1)	(2)	(3)	(4)	(5)
Panel A: No trim					
Realization rate	0.88	0.90	0.73	0.54	0.54
Point estimate	-3.87	-3.93	-3.16	-2.21	-2.20
	(0.52)	(0.53)	(0.50)	(0.49)	(0.50)
Observations	57,481,920	57,477,350	57,477,350	57,481,920	57,477,350
Panel B: Trim outlier observations					
Realization rate	0.81	0.83	0.69	0.55	0.55
Point estimate	-3.44	-3.49	-2.87	-2.19	-2.17
	(0.35)	(0.35)	(0.33)	(0.34)	(0.34)
Observations	56,332,278	56,327,608	56,327,608	56,332,278	56,327,608
Panel C: Trim outlier schools					
Realization rate	0.94	0.96	0.76	0.54	0.54
Point estimate	-3.64	-3.70	-2.92	-1.99	-1.98
	(0.50)	(0.51)	(0.48)	(0.47)	(0.48)
Observations	56,737,632	56,733,086	56,733,086	56,737,632	56,733,086
Panel D: Trim observations and schools					
Realization rate	0.88	0.89	0.74	0.59	0.58
Point estimate	-3.35	-3.40	-2.79	-2.13	-2.11
	(0.35)	(0.35)	(0.33)	(0.34)	(0.34)
Observations	55,683,522	55,678,868	55,678,868	55,683,522	55,678,868
School-Hour FE	Yes	Yes	Yes	Yes	Yes
School-Hour-Month FE	No	Yes	Yes	No	Yes
Time trend	No	No	Yes	No	No
Month of Sample FE	No	No	No	Yes	Yes

Notes: This table reports alternative results for Tables 4 and 5 using a double machine learning estimator (Chernozhukov et al. (2018)). Note that, for computational tractability, we predict electricity consumption and treatment timing school-by-school in a first step. In the second step, the dependent variable is the prediction error of electricity consumption in kWh. The independent variable is the prediction error in treatment timing interacted with a treatment dummy, to allow treated schools to have different treatment effect from the treatment timing: $\tilde{Y}_{ith} = \beta_0\ \tilde{T}_{it} + \beta_1\ Treated_i \times \tilde{T}_{it} + \alpha_{ith} + \epsilon_{it}$. Standard errors are clustered at the school level.