Appendix to "Using Machine Learning to Target Treatment: The Case of Household Energy Use" (2025)

Christopher R. Knittel and Samuel Stolper

A By-wave summary statistics

Table A1: Treatment-Control Balance – CT 4/2014, EDelivery

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	624.908	626.590	1.682
	(161.105)	(161.356)	(1.479)
Home value ($)	277,420.992	280,210.844	2,789.852
	(231,731.282)	(246,351.486)	(2,149.374)
Home square footage	17.036	17.211	0.174**
	(8.518)	(8.521)	(0.078)
Number of rooms in home	6.715	6.752	0.037**
	(1.967)	(1.988)	(0.018)
Year home built (1-5)	1,968.422	1,968.311	-0.111
	(22.322)	(22.469)	(0.205)
Single-family occupancy (=1)	0.836	0.836	0.000
	(0.371)	(0.370)	(0.003)
Renter (=1)	0.161	0.160	-0.001
	(0.368)	(0.367)	(0.003)
Annual income	94,718.798	95,834.531	1,115.733**
	(61,535.744)	(62,282.609)	(565.807)
Education (1-5)	3.173	3.174	0.001
	(1.205)	(1.201)	(0.011)
GreenAware score (1-4)	2.193	2.212	0.018*
	(1.155)	(1.155)	(0.011)
Number of adults	2.278	2.292	0.015
	(1.288)	(1.290)	(0.012)
Child in home (=1)	0.578	0.583	0.005
	(0.494)	(0.493)	(0.005)
Participated in EA (=1)	0.315	0.320	0.005
	(0.465)	(0.466)	(0.004)
Age	52.288	52.229	-0.059
	(13.953)	(13.798)	(0.128)
N			85,360

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A2: Treatment-Control Balance – CT 4/2014 , HEA

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	749.164	751.321	2.157
	(384.918)	(392.121)	(7.521)
Home value ($)	344,217.602	354,250.865	10,033.262*
	(308,261.525)	(316,415.308)	(6,038.665)
Home square footage	19.059	19.076	0.017
	(8.677)	(8.640)	(0.168)
Number of rooms in home	7.042	7.030	-0.012
	(1.835)	(1.801)	(0.035)
Year home built (1-5)	1,968.777	1,968.634	-0.143
	(21.701)	(21.006)	(0.417)
Single-family occupancy (=1)	0.953	0.951	-0.002
	(0.211)	(0.216)	(0.004)
Renter (=1)	0.039	0.038	-0.001
	(0.193)	(0.190)	(0.004)
Annual income	108,461.636	109,783.954	1,322.318
	(67,451.158)	(68,656.181)	(1,317.492)
Education (1-5)	3.436	3.439	0.003
	(1.196)	(1.210)	(0.023)
GreenAware score (1-4)	1.997	1.974	-0.024
	(1.139)	(1.130)	(0.022)
Number of adults	2.466	2.454	-0.013
	(1.287)	(1.304)	(0.025)
Child in home (=1)	0.506	0.501	-0.005
	(0.500)	(0.500)	(0.010)
Participated in EA (=1)	0.998	0.998	0.000
	(0.049)	(0.045)	(0.001)
Age	57.529	57.296	-0.233
	(14.981)	(15.132)	(0.292)
N			11,883

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A3: Treatment-Control Balance – CT 4/2014

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	1,269.834	1,269.769	-0.064
	(358.778)	(357.906)	(2.914)
Home value ($)	455,372.587	460,166.524	4,793.937
	(525,747.650)	(532,752.340)	(4,275.916)
Home square footage	22.133	22.043	-0.090
	(11.756)	(11.340)	(0.095)
Number of rooms in home	7.456	7.459	0.004
	(2.118)	(2.114)	(0.017)
Year home built (1-5)	1,973.040	1,972.769	-0.270
	(21.252)	(21.431)	(0.173)
Single-family occupancy (=1)	0.909	0.911	0.002
	(0.288)	(0.284)	(0.002)
Renter (=1)	0.069	0.068	-0.001
	(0.253)	(0.251)	(0.002)
Annual income	119,517.238	119,713.659	196.420
	(74,771.881)	(74,765.329)	(607.444)
Education (1-5)	3.375	3.378	0.002
	(1.223)	(1.223)	(0.010)
GreenAware score (1-4)	2.211	2.211	-0.000
	(1.181)	(1.184)	(0.010)
Number of adults	2.778	2.767	-0.011
	(1.382)	(1.361)	(0.011)
Child in home (=1)	0.531	0.536	0.005
	(0.499)	(0.499)	(0.004)
Participated in EA (=1)	0.272	0.268	-0.004
	(0.445)	(0.443)	(0.004)
Age	57.229	57.207	-0.022
	(13.450)	(13.567)	(0.109)
N			199,802

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A4: Treatment-Control Balance – CT 2/2016

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	826.375	825.667	-0.708
	(263.422)	(262.723)	(2.188)
Home value ($)	329,114.650	327,052.713	-2,061.937
	(353,410.505)	(360,702.968)	(2,943.767)
Home square footage	18.501	18.523	0.022
	(9.522)	(9.891)	(0.079)
Number of rooms in home	6.909	6.925	0.016
	(1.995)	(1.994)	(0.017)
Year home built (1-5)	1,970.027	1,970.351	0.324*
	(21.951)	(22.030)	(0.182)
Single-family occupancy (=1)	0.868	0.869	0.001
	(0.338)	(0.338)	(0.003)
Renter (=1)	0.119	0.115	-0.004
	(0.324)	(0.319)	(0.003)
Annual income	97,970.099	98,789.843	819.744
	(66,083.544)	(66,608.057)	(549.607)
Education (1-5)	3.148	3.161	0.013
	(1.233)	(1.230)	(0.010)
GreenAware score (1-4)	2.158	2.172	0.014
	(1.151)	(1.157)	(0.010)
Number of adults	2.390	2.383	-0.007
	(1.306)	(1.295)	(0.011)
Child in home (=1)	0.504	0.508	0.004
	(0.500)	(0.500)	(0.004)
Participated in EA (=1)	0.255	0.256	0.001
	(0.436)	(0.436)	(0.004)
Age	56.457	56.399	-0.057
	(15.074)	(15.299)	(0.125)
N			137,896

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A5: Treatment-Control Balance – CT 2/2016, LowIncome

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	599.101	597.337	-1.764
	(328.851)	(331.891)	(5.079)
Home value ($)	198,801.086	198,997.287	196.201
	(140,341.332)	(164,668.160)	(2,341.044)
Home square footage	19.301	19.273	-0.028
	(13.212)	(14.483)	(0.213)
Number of rooms in home	7.557	7.538	-0.019
	(2.633)	(2.704)	(0.041)
Year home built (1-5)	1,964.336	1,964.282	-0.054
	(20.718)	(20.475)	(0.317)
Single-family occupancy (=1)	0.523	0.534	0.011
	(0.500)	(0.499)	(0.008)
Renter (=1)	0.586	0.577	-0.009
	(0.493)	(0.494)	(0.008)
Annual income	50,289.331	50,806.214	516.883
	(48,131.584)	(48,166.971)	(740.344)
Education (1-5)	2.292	2.311	0.019
	(1.023)	(1.018)	(0.016)
GreenAware score (1-4)	2.489	2.476	-0.014
	(0.942)	(0.943)	(0.014)
Number of adults	1.750	1.754	0.004
	(1.150)	(1.135)	(0.018)
Child in home (=1)	0.521	0.514	-0.007
	(0.500)	(0.500)	(0.008)
Participated in EA (=1)	0.464	0.457	-0.007
	(0.499)	(0.498)	(0.008)
Age	57.745	57.545	-0.200
	(16.635)	(16.620)	(0.256)
N			16,981

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A6: Treatment-Control Balance – CT 3/2016

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	467.986	473.962	5.976
	(253.364)	(257.055)	(4.822)
Home value ($)	280,558.442	284,357.853	3,799.411
	(308,964.173)	(313,255.935)	(5,879.759)
Home square footage	20.264	20.444	0.180
	(14.551)	(15.874)	(0.281)
Number of rooms in home	7.476	7.465	-0.011
	(2.466)	(2.449)	(0.047)
Year home built (1-5)	1,967.569	1,967.822	0.253
	(20.404)	(20.442)	(0.387)
Single-family occupancy (=1)	0.564	0.576	0.012
	(0.496)	(0.494)	(0.009)
Renter (=1)	0.465	0.442	-0.023**
	(0.499)	(0.497)	(0.009)
Annual income	80,994.130	83,083.766	2,089.636
	(69,703.829)	(71,062.177)	(1,327.998)
Education (1-5)	2.755	2.795	0.040*
	(1.221)	(1.224)	(0.023)
GreenAware score (1-4)	2.430	2.399	-0.032
	(1.089)	(1.093)	(0.021)
Number of adults	1.646	1.625	-0.021
	(1.060)	(1.026)	(0.020)
Child in home (=1)	0.605	0.608	0.002
	(0.489)	(0.488)	(0.009)
Participated in EA (=1)	0.293	0.302	0.009
	(0.455)	(0.459)	(0.009)
Age	45.861	45.824	-0.038
	(12.857)	(13.099)	(0.245)
N			17,395

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A7: Treatment-Control Balance – CT 1/2017

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	733.469	730.406	-3.063
	(274.802)	(268.116)	(2.425)
Home value ($)	312,766.965	313,738.457	971.492
	(333,034.112)	(346,042.747)	(2,984.370)
Home square footage	18.354	18.444	0.090
	(10.519)	(10.394)	(0.093)
Number of rooms in home	6.916	6.930	0.014
	(2.076)	(2.108)	(0.018)
Year home built (1-5)	1,969.314	1,969.392	0.078
	(21.260)	(21.503)	(0.189)
Single-family occupancy (=1)	0.843	0.844	0.001
	(0.364)	(0.363)	(0.003)
Renter (=1)	0.175	0.172	-0.003
	(0.380)	(0.378)	(0.003)
Annual income	91,492.890	91,995.989	503.100
	(65,316.370)	(65,069.355)	(579.262)
Education (1-5)	3.035	3.042	0.007
	(1.235)	(1.228)	(0.011)
GreenAware score (1-4)	2.121	2.126	0.006
	(1.108)	(1.113)	(0.010)
Number of adults	2.204	2.221	0.017
	(1.266)	(1.271)	(0.011)
Child in home (=1)	0.476	0.472	-0.004
	(0.499)	(0.499)	(0.004)
Participated in EA (=1)	0.262	0.264	0.001
	(0.440)	(0.441)	(0.004)
Age	56.612	56.713	0.101
	(15.873)	(15.876)	(0.141)
N			69,517

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A8: Treatment-Control Balance – EMA 4/2014

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	630.265	626.852	-3.413
	(296.179)	(291.622)	(4.139)
Home value ($)	633,066.601	629,949.589	-3,117.012
	(450,001.040)	(429,762.562)	(6,268.217)
Home square footage	20.467	20.289	-0.178
	(13.680)	(11.869)	(0.189)
Number of rooms in home	7.472	7.493	0.021
	(2.571)	(2.636)	(0.036)
Year home built (1-5)	1,957.055	1,956.932	-0.123
	(26.754)	(26.612)	(0.374)
Single-family occupancy (=1)	0.716	0.706	-0.011*
	(0.451)	(0.456)	(0.006)
Renter (=1)	0.154	0.159	0.004
	(0.361)	(0.366)	(0.005)
Annual income	112,438.060	111,245.929	-1,192.131
	(73,750.526)	(73,077.377)	(1,031.499)
Education (1-5)	3.598	3.602	0.004
	(1.245)	(1.251)	(0.017)
GreenAware score (1-4)	1.945	1.923	-0.022
	(1.065)	(1.053)	(0.015)
Number of adults	2.433	2.435	0.002
	(1.381)	(1.397)	(0.019)
Child in home (=1)	0.451	0.450	-0.001
	(0.498)	(0.498)	(0.007)
Participated in EA (=1)	0.415	0.406	-0.009
	(0.493)	(0.491)	(0.007)
Age	57.076	57.031	-0.045
	(15.438)	(15.556)	(0.216)
N			49,610

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A9: Treatment-Control Balance – EMA 2/2016

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	571.617	575.180	3.564
	(283.500)	(285.178)	(2.734)
Home value ($)	595,170.323	599,902.105	4,731.782
	(432,377.807)	(441,666.748)	(4,185.243)
Home square footage	19.753	19.802	0.049
	(15.017)	(14.043)	(0.142)
Number of rooms in home	7.213	7.247	0.034
	(2.753)	(2.798)	(0.027)
Year home built (1-5)	1,961.592	1,961.642	0.050
	(26.782)	(26.829)	(0.258)
Single-family occupancy (=1)	0.529	0.532	0.003
	(0.499)	(0.499)	(0.005)
Renter (=1)	0.257	0.251	-0.005
	(0.437)	(0.434)	(0.004)
Annual income	102,283.108	101,844.466	-438.643
	(72,966.299)	(72,937.368)	(702.635)
Education (1-5)	3.527	3.523	-0.004
	(1.264)	(1.262)	(0.012)
GreenAware score (1-4)	2.035	2.049	0.014
	(1.127)	(1.125)	(0.011)
Number of adults	2.000	1.997	-0.003
	(1.277)	(1.266)	(0.012)
Child in home (=1)	0.467	0.467	0.000
	(0.499)	(0.499)	(0.005)
Participated in EA (=1)	0.357	0.353	-0.004
	(0.479)	(0.478)	(0.005)
Age	50.998	51.108	0.111
	(15.045)	(15.038)	(0.145)
N			59,892

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A10: Treatment-Control Balance – EMA 1/2017

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	545.194	546.630	1.437
	(237.995)	(240.419)	(2.270)
Home value ($)	623,797.117	619,312.014	-4,485.103
	(416,153.005)	(411,110.190)	(3,936.992)
Home square footage	22.627	22.657	0.030
	(16.386)	(17.233)	(0.159)
Number of rooms in home	7.967	7.961	-0.006
	(2.698)	(2.718)	(0.026)
Year home built (1-5)	1,962.793	1,962.589	-0.204
	(24.825)	(24.919)	(0.236)
Single-family occupancy (=1)	0.423	0.418	-0.005
	(0.494)	(0.493)	(0.005)
Renter (=1)	0.400	0.402	0.002
	(0.490)	(0.490)	(0.005)
Annual income	98,107.398	97,907.337	-200.062
	(71,352.630)	(70,929.453)	(676.579)
Education (1-5)	3.481	3.484	0.003
	(1.259)	(1.258)	(0.012)
GreenAware score (1-4)	2.104	2.104	0.000
	(1.122)	(1.123)	(0.011)
Number of adults	1.725	1.735	0.010
	(1.149)	(1.164)	(0.011)
Child in home (=1)	0.405	0.405	0.000
	(0.491)	(0.491)	(0.005)
Participated in EA (=1)	0.299	0.300	0.001
	(0.458)	(0.458)	(0.004)
Age	46.785	46.884	0.099
	(13.070)	(13.210)	(0.125)
N			47,401

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A11: Treatment-Control Balance – NH 2/2014

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	633.387	633.962	0.574
	(340.881)	(339.372)	(3.292)
Home value ($)	253,800.756	254,720.858	920.102
	(167,772.409)	(160,748.651)	(1,590.017)
Home square footage	18.562	18.593	0.032
	(10.208)	(10.170)	(0.099)
Number of rooms in home	6.615	6.597	-0.018
	(1.860)	(1.883)	(0.018)
Year home built (1-5)	1,977.931	1,977.671	-0.260
	(20.008)	(20.327)	(0.195)
Single-family occupancy (=1)	0.867	0.873	0.006*
	(0.340)	(0.333)	(0.003)
Renter (=1)	0.107	0.106	-0.001
	(0.309)	(0.307)	(0.003)
Annual income	87,679.281	88,092.698	413.417
	(56,621.379)	(56,647.211)	(548.087)
Education (1-5)	3.034	3.061	0.027**
	(1.153)	(1.161)	(0.011)
GreenAware score (1-4)	2.171	2.153	-0.018
	(1.111)	(1.110)	(0.011)
Number of adults	2.415	2.423	0.009
	(1.281)	(1.285)	(0.012)
Child in home (=1)	0.356	0.349	-0.008*
	(0.479)	(0.477)	(0.005)
Participated in EA (=1)	0.445	0.446	0.001
	(0.497)	(0.497)	(0.005)
Age	60.519	60.494	-0.025
	(13.506)	(13.370)	(0.130)
N			42,709

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A12: Treatment-Control Balance – NH 4/2015

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	1,006.966	1,007.972	1.006
	(261.429)	(261.039)	(3.206)
Home value ($)	303,305.418	299,718.922	-3,586.496
	(207,309.823)	(217,554.293)	(2,580.139)
Home square footage	21.863	21.776	-0.088
	(9.475)	(9.207)	(0.115)
Number of rooms in home	7.191	7.193	0.002
	(1.846)	(1.868)	(0.023)
Year home built (1-5)	1,981.566	1,981.854	0.288
	(19.693)	(19.294)	(0.240)
Single-family occupancy (=1)	0.943	0.939	-0.004
	(0.232)	(0.240)	(0.003)
Renter (=1)	0.063	0.066	0.003
	(0.243)	(0.248)	(0.003)
Annual income	106,192.935	104,771.302	-1,421.633*
	(60,897.988)	(59,520.977)	(742.423)
Education (1-5)	3.250	3.234	-0.016
	(1.168)	(1.172)	(0.014)
GreenAware score (1-4)	2.219	2.249	0.031**
	(1.157)	(1.155)	(0.014)
Number of adults	2.755	2.750	-0.006
	(1.324)	(1.331)	(0.016)
Child in home (=1)	0.495	0.495	-0.000
	(0.500)	(0.500)	(0.006)
Participated in EA (=1)	0.522	0.509	-0.014**
	(0.500)	(0.500)	(0.006)
Age	56.428	56.201	-0.227
	(12.130)	(12.042)	(0.149)
N			32,571

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A13: Treatment-Control Balance – WMA 2/2014

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	648.841	648.066	-0.774
	(332.561)	(329.828)	(3.940)
Home value ($)	234,587.032	236,885.349	2,298.317
	(146,498.370)	(166,456.981)	(1,757.374)
Home square footage	17.132	17.136	0.004
	(9.373)	(10.318)	(0.112)
Number of rooms in home	6.606	6.608	0.002
	(1.967)	(1.959)	(0.023)
Year home built (1-5)	1,962.022	1,961.976	-0.046
	(23.937)	(23.627)	(0.284)
Single-family occupancy (=1)	0.904	0.905	0.000
	(0.294)	(0.294)	(0.003)
Renter (=1)	0.102	0.097	-0.004
	(0.302)	(0.296)	(0.004)
Annual income	76,915.441	76,696.648	-218.793
	(52,144.371)	(51,334.224)	(617.490)
Education (1-5)	3.007	2.991	-0.017
	(1.195)	(1.195)	(0.014)
GreenAware score (1-4)	2.082	2.106	0.024*
	(1.089)	(1.095)	(0.013)
Number of adults	2.473	2.491	0.018
	(1.347)	(1.348)	(0.016)
Child in home (=1)	0.364	0.379	0.015***
	(0.481)	(0.485)	(0.006)
Participated in EA (=1)	0.415	0.411	-0.004
	(0.493)	(0.492)	(0.006)
Age	60.568	60.376	-0.192
	(14.981)	(15.046)	(0.178)
N			95,455

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A14: Treatment-Control Balance – WMA 1/2015

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	585.276	578.991	-6.285
	(334.700)	(334.875)	(4.684)
Home value ($)	195,001.443	196,685.032	1,683.589
	(143,407.985)	(160,307.845)	(2,077.791)
Home square footage	20.409	20.245	-0.165
	(17.800)	(16.845)	(0.245)
Number of rooms in home	7.592	7.563	-0.029
	(2.521)	(2.529)	(0.035)
Year home built (1-5)	1,955.603	1,955.520	-0.082
	(24.865)	(25.067)	(0.349)
Single-family occupancy (=1)	0.721	0.725	0.005
	(0.449)	(0.446)	(0.006)
Renter (=1)	0.432	0.433	0.000
	(0.495)	(0.495)	(0.007)
Annual income	51,806.448	51,548.310	-258.138
	(46,632.552)	(44,819.977)	(645.254)
Education (1-5)	2.456	2.463	0.007
	(1.128)	(1.128)	(0.016)
GreenAware score (1-4)	2.433	2.429	-0.004
	(0.994)	(0.991)	(0.014)
Number of adults	2.060	2.047	-0.013
	(1.313)	(1.301)	(0.018)
Child in home (=1)	0.550	0.552	0.002
	(0.497)	(0.497)	(0.007)
Participated in EA (=1)	0.481	0.476	-0.005
	(0.500)	(0.499)	(0.007)
Age	55.423	55.456	0.033
	(16.482)	(16.471)	(0.231)
N			24,837

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Table A15: Treatment-Control Balance – WMA 12/2015

	Treatment (1)	**Control** (2)	**Balance** (3)
Baseline Consumption (kWh)	546.054	547.550	1.495
	(315.114)	(335.962)	(8.776)
Home value ($)	222,638.506	226,450.466	3,811.960
	(171,756.942)	(189,450.955)	(4,808.905)
Home square footage	21.228	21.827	0.599
	(15.217)	(18.349)	(0.433)
Number of rooms in home	7.721	7.716	-0.005
	(2.498)	(2.392)	(0.069)
Year home built (1-5)	1,958.683	1,957.925	-0.758
	(25.367)	(24.669)	(0.698)
Single-family occupancy (=1)	0.726	0.714	-0.012
	(0.446)	(0.452)	(0.012)
Renter (=1)	0.419	0.409	-0.010
	(0.493)	(0.492)	(0.014)
Annual income	57,713.711	57,123.270	-590.441
	(52,061.798)	(52,575.447)	(1,438.771)
Education (1-5)	2.680	2.625	-0.055
	(1.234)	(1.196)	(0.034)
GreenAware score (1-4)	2.393	2.417	0.025
	(1.016)	(1.023)	(0.028)
Number of adults	1.888	1.871	-0.017
	(1.233)	(1.264)	(0.034)
Child in home (=1)	0.568	0.574	0.006
	(0.495)	(0.495)	(0.014)
Participated in EA (=1)	0.409	0.396	-0.014
	(0.492)	(0.489)	(0.014)
Age	50.539	50.130	-0.409
	(15.808)	(15.923)	(0.437)
N			11,272

Note: Columns (1) and (2) display the mean and (in parentheses) standard deviation of each listed household characteristic for the treatment and control groups, respectively. Column (3) displays differences in means and (in parentheses) standard errors from a t-test. The wave sample size ("N") is listed at the bottom of the table. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

B Additional results

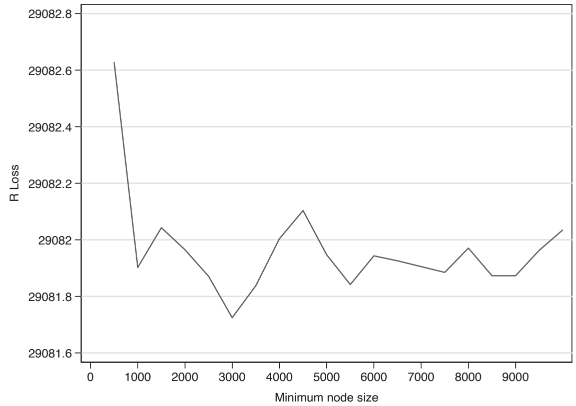
Table B1: Difference-in-means estimates of average treatment effect

	Mean usage year 1 (1)	Mean usage year 2 (2)	Mean usage year 3 (3)
Treatment	-6.43***	-11.46***	-14.39***
	(1.48)	(2.76)	(3.27)
N	809,349	540,728	477,529

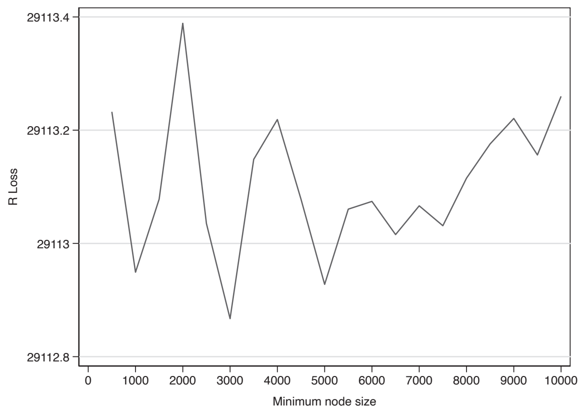
Notes: We differences in means via cross-sectional regression of the relevant post-treatment mean (year 1, 2, or 3) on treatment status. We run a separate regression for each wave and then calculate a pooled average treatment effect (ATE) as the average of wave-specific ATEs weighted by the fraction of the full sample in each wave. We calculate standard errors (shown in parentheses) according to the formula provided by Athey and Imbens (2017) for stratified randomized experiments. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Figure B1: Tuning minimum node size in forest growth

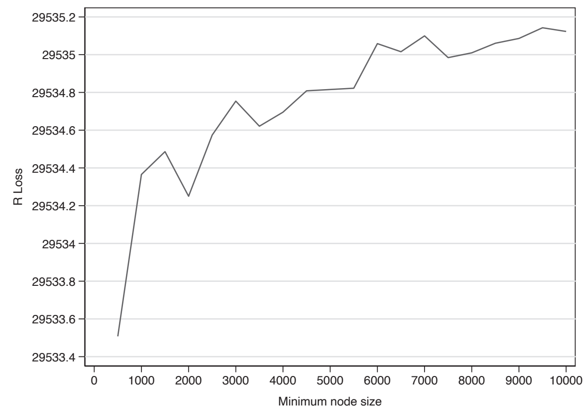
Panel A. Full-sample forest



Panel B. "Internal" targeting forest

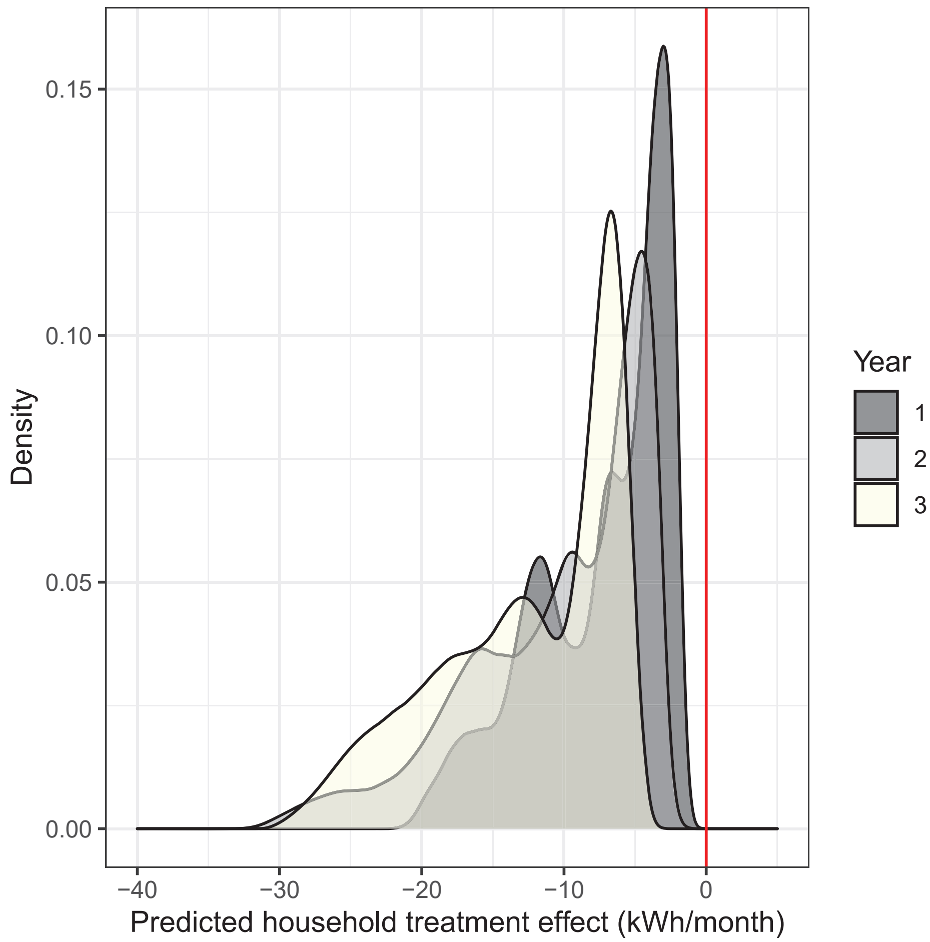


Panel C. "External" targeting forest



Notes: Each panel plots the average out-of-bag prediction error ("R loss") from forests grown with a range of minimum node sizes (x-values). Panel A uses the full sample of households as forest input data; Panel B uses a random 50% subsample; Panel C uses 2014 waves. We use these results to tune the minimum node size parameter, choosing the size that produces the minimum prediction error as our preferred parameter value (Nie and Wager, 2021).

Figure B2: Distribution of predicted treatment effects from a clustered forest



Notes: The figure is based on an identical calculation to that of Figure 5, except that in this case the forest employs clustering at the zip code level (see Athey and Wager (2019) for implementation details). Each plotted distribution is a kernel density of household treatment effects in a specific year (1, 2, or 3) of Home Energy Report programming. The sample is fixed across years: only households with non-missing consumption in all three post-years are included. Treatment effect predictions come from our causal forest (Section 2.2).

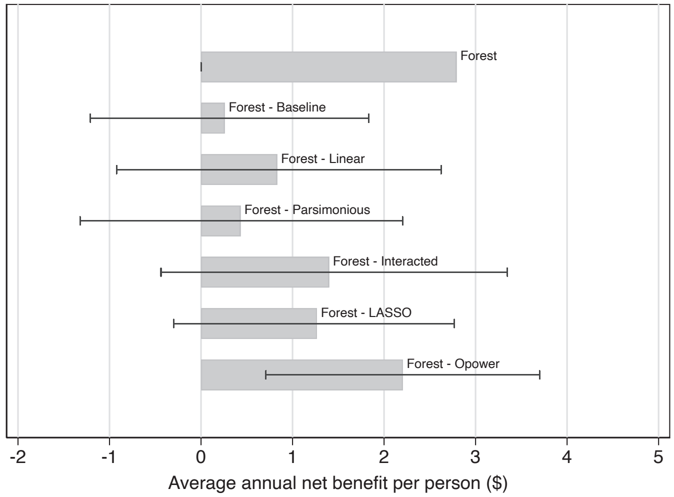
Table B2: Balance between training and hold-out samples in the external horserace

	Sample-wide mean (1)	**Difference in means** (2)
Residual consumption (kWh)	-0.363	54.204***
	(377.002)	(1.682)
Baseline Consumption (kWh)	868.381	-189.800***
	(445.090)	(2.027)
Home value ($)	371,924.560	9,487.963***
	(382,843.118)	(1,811.991)
Home square footage	19.658	0.795***
	(10.996)	(0.065)
Number of rooms in home	7.081	0.236***
	(2.159)	(0.013)
Year home built (1-5)	1968	-2.148***
	(23.917)	(0.137)
Single-family occupancy (=1)	0.858	-0.148***
	(0.349)	(0.002)
Renter (=1)	0.111	0.121***
	(0.314)	(0.002)
Annual income	101,049.129	-13,943.123***
	(67,776.523)	(345.198)
Education (1-5)	3.241	-0.115***
	(1.232)	(0.007)
GreenAware score (1-4)	2.147	0.037***
	(1.135)	(0.006)
Number of adults	2.556	-0.261***
	(1.363)	(0.007)
Child in home (=1)	0.474	0.020***
	(0.499)	(0.003)
Participated in EA (=1)	0.367	0.101***
	(0.482)	(0.003)
Age	57.442	-2.806***
	(14.339)	(0.078)

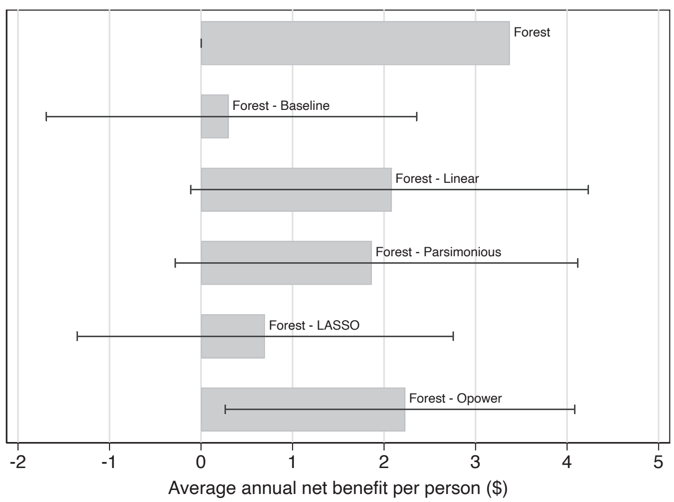
Notes: Column 1 displays the full-sample mean and (in parentheses) standard deviation of each listed household characteristic. Column 2 displays differences in means between the training group (waves beginning in 2014) and the hold-out group (those beginning after 2014), as well as (in parentheses) the corresponding standard errors. Column 2 estimates come from linear regression of each characteristic on a binary variable equaling one if the household is in the hold-out group, with robust standard errors.. * $p < 0.01$, $^{**}p < 0.05$, *** $p < 0.01$.

Figure B3: Social net benefits of targeting, by predictive method, with alternative bootstrapping

Panel A. Training on a random sample



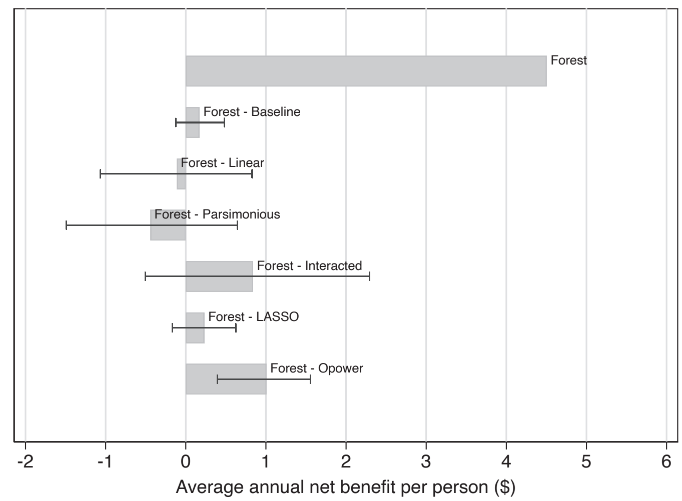
Panel B. Training on 2014 waves



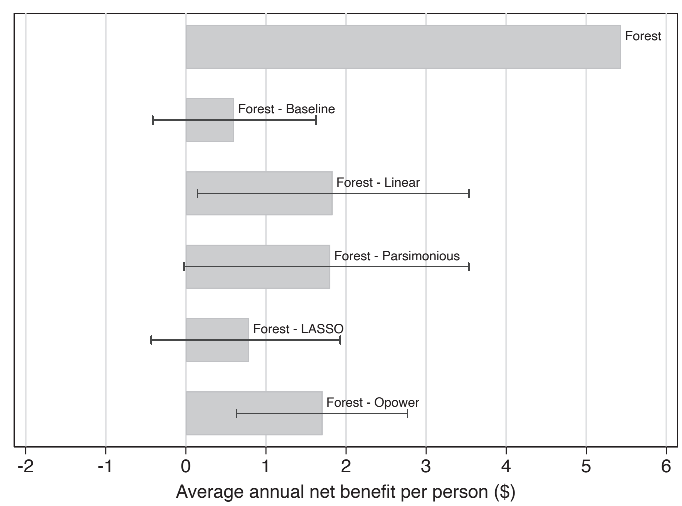
Notes: This figure is based on an identical calculation to that of Figure 9, except that an alternative, and more conservative, bootstrapping procedure (the "non fixed-rule" version) is used to generate confidence intervals; see Appendix C.5 for details. The top bar is the estimated annual social net benefits (SNB) produced from the treatment assignment chosen by the forest, relative to a no-action counterfactual. Each other bar depicts the estimated annual gain in social net benefits (SNB) produced from targeting using the forest instead of the listed alternative method. The targeting rule is to treat if predicted $SNB > 0$. Net benefits are expressed as an average per household in the full test sample, so that Panels A and B are more comparable. Panel A depicts results from building all predictive models with a 50% random sample of households and targeting in the other 50% "test" sample. Panel B depicts results from building all predictive models exclusively with households in HER waves beginning in 2014 and targeting among waves beginning in 2015 or later; the "Interacted" model is not included in Panel B because it does not identify households that satisfy the targeting criterion. Confidence intervals are generated via bootstrapping, which we describe in greater detail in Appendix C.5.

Figure B4: Social net benefits of targeting by predictive method, MC=3.5

Panel A. Training on a random sample



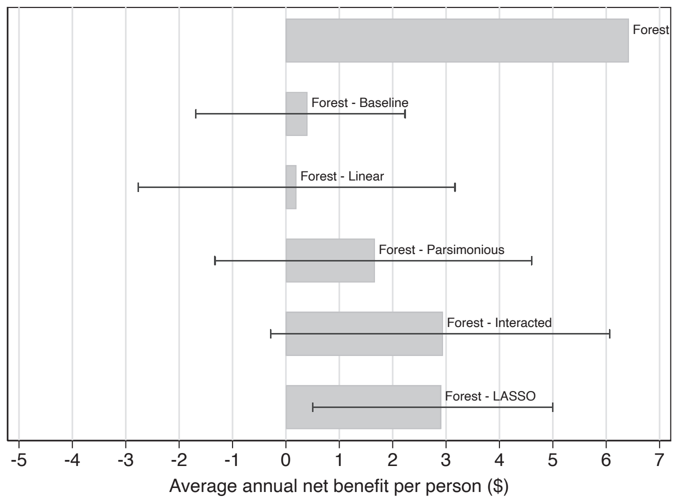
Panel B. Training on 2014 waves



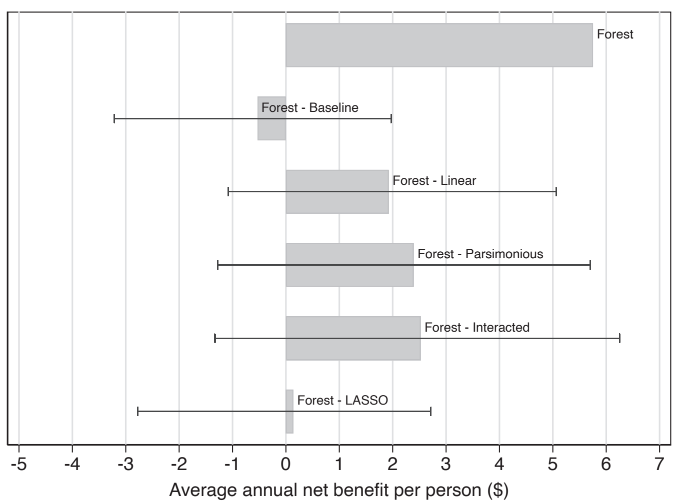
Notes: This figure is based on an identical calculation to that of Figure 9, except that instead of using an annual HER marginal cost of $7, we here use a marginal cost of $3.50. The top bar is the estimated annual social net benefits (SNB) produced from the treatment assignment chosen by the forest, relative to a no-action counterfactual. Each other bar depicts the estimated annual gain in social net benefits (SNB) produced from targeting using the forest instead of the listed alternative method. The targeting rule is to treat if predicted $SNB > 0$. Net benefits are expressed as an average per household in the full test sample, so that Panels A and B are more comparable. Panel A depicts results from building all predictive models with a 50% random sample of households and targeting in the other 50% "test" sample. Panel B depicts results from building all predictive models exclusively with households in HER waves beginning in 2014 and targeting among waves beginning in 2015 or later; the "Interacted" model is not included in Panel B because it does not identify households that satisfy the targeting criterion. Confidence intervals are generated via bootstrapping, which we describe in greater detail in Appendix C.5.

Figure B5: Social net benefits produced by the top half of households, by method

Panel A. Training on a random sample



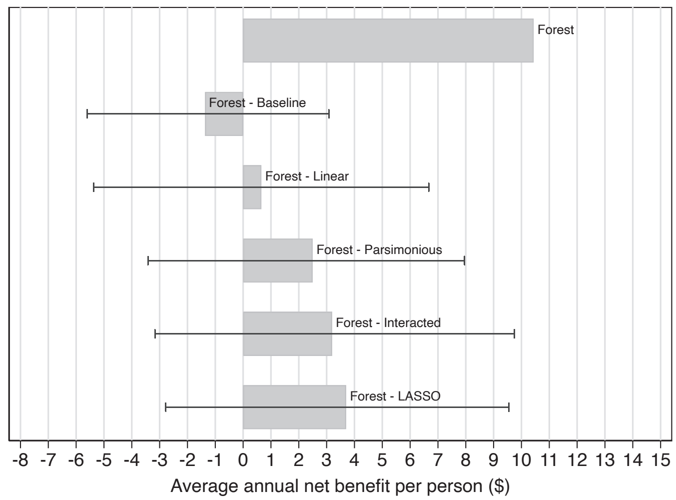
Panel B. Training on 2014 waves



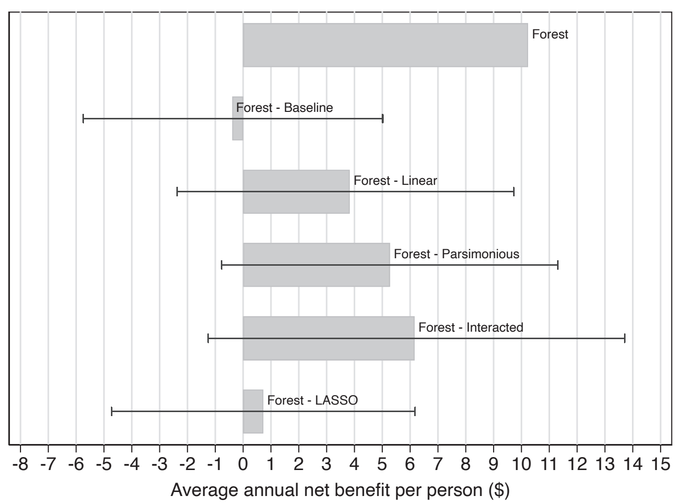
Notes: The targeting rule is to treat if predicted SNB are above the sample median. The top bar is the estimated annual social net benefits (SNB) produced from the treatment assignment chosen by the forest, relative to a no-action counterfactual. Each other bar depicts the estimated annual gain in social net benefits (SNB) produced from targeting using the forest instead of the listed alternative method. The targeting rule is to treat if predicted $SNB > 0$. Net benefits are expressed as an average per household in the full test sample, so that Panels A and B are more comparable. Panel A depicts results from building all predictive models with a 50% random sample of households and targeting in the other 50% "test" sample. Panel B depicts results from building all predictive models exclusively with households in HER waves beginning in 2014 and targeting among waves beginning in 2015 or later; the "Interacted" model is not included in Panel B because it does not identify households that satisfy the targeting criterion. Confidence intervals are generated via bootstrapping, which we describe in greater detail in Appendix C.5.

Figure B6: Social net benefits produced by the top quartile of households, by method

Panel A. Training on a random sample



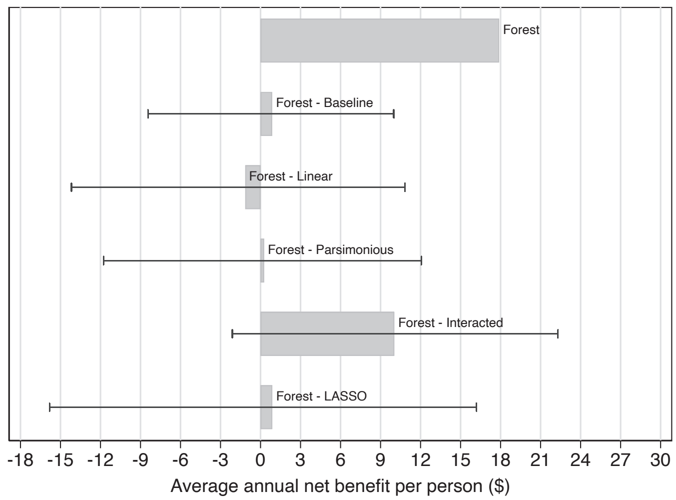
Panel B. Training on 2014 waves



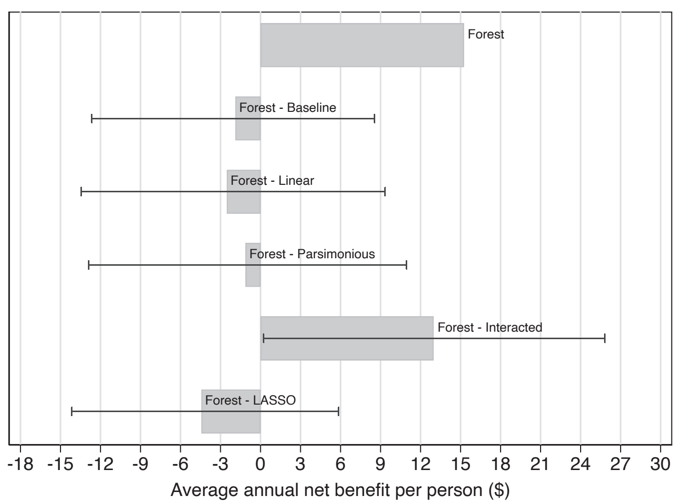
Notes: The targeting rule is to treat if predicted SNB are abovein the top quartile of the sample. The top bar is the estimated annual social net benefits (SNB) produced from the treatment assignment chosen by the forest, relative to a no-action counterfactual. Each other bar depicts the estimated annual gain in social net benefits (SNB) produced from targeting using the forest instead of the listed alternative method. The targeting rule is to treat if predicted $SNB > 0$. Net benefits are expressed as an average per household in the full test sample, so that Panels A and B are more comparable. Panel A depicts results from building all predictive models with a 50% random sample of households and targeting in the other 50% "test" sample. Panel B depicts results from building all predictive models exclusively with households in HER waves beginning in 2014 and targeting among waves beginning in 2015 or later; the "Interacted" model is not included in Panel B because it does not identify households that satisfy the targeting criterion. Confidence intervals are generated via bootstrapping, which we describe in greater detail in Appendix C.5.

Figure B7: Social net benefits produced among by top decile of households, by method

Panel A. Training on a random sample



Panel B. Training on 2014 waves



Notes: The targeting rule is to treat if predicted SNB are abovein the top decile of the sample. The top bar is the estimated annual social net benefits (SNB) produced from the treatment assignment chosen by the forest, relative to a no-action counterfactual. Each other bar depicts the estimated annual gain in social net benefits (SNB) produced from targeting using the forest instead of the listed alternative method. The targeting rule is to treat if predicted $SNB > 0$. Net benefits are expressed as an average per household in the full test sample, so that Panels A and B are more comparable. Panel A depicts results from building all predictive models with a 50% random sample of households and targeting in the other 50% "test" sample. Panel B depicts results from building all predictive models exclusively with households in HER waves beginning in 2014 and targeting among waves beginning in 2015 or later; the "Interacted" model is not included in Panel B because it does not identify households that satisfy the targeting criterion. Confidence intervals are generated via bootstrapping, which we describe in greater detail in Appendix C.5.

C Technical details

C.1 Multiple imputation

We use multiple imputation (MI) to fill in missing values of household characteristics. We implement MI through the multivariate imputation by chained equations (MICE) approach. The process can be broken down into the following steps:

1. We define a set of variables $X_1, \ldots, X_n$ to be used in the imputation model. Every missing value is filed in at random to act as a placeholder.

2. The placeholder values for the first variable with at least one missing value, X_1, are returned to missing and the observed values of X_1 are regressed on $X_2, \ldots, X_n$ using a regression model (e.g., linear, logistic) based on the data type of X_1. Predictive mean matching (e.g., known-nearest neighbor) can also be performed.

3. The missing values of X_1 are replaced by simulated draws from the posterior predictive distribution of X_1. In the remaining steps, X_1 consists of the observed and imputed values.

4. Repeat Steps 2-3 for the remaining $n-1$ variables where the value of each variable is updated. For example, the next step would be to regress X_2 is regressed on the newly imputed values of X_1 and $X_3, \ldots, X_n$ and estimate missing values of X_2 with draws from its posterior predictive distribution. A "cycle" is said to have passed when all variables have been imputed.

5. Repeat Steps 2-4 for 20 cycles to stabilize the results. The placeholder values at the start of each cycle are the imputed values from the previous cycle. A single imputed dataset is produced at the end of all 10 cycles.

6. Repeat Steps 1-5 M number of times. (White et al., 2011) suggests that a rule of thumb for deciding M is that M should be a least equal to the percentage of incomplete cases in the dataset.

C.2 Classification analysis

In Section 3.2 of the paper, we present a classification analysis (CLAN), which is a comparison of average attributes in the "most affected" and "least affected" subpopulations proposed by Chernozhukov et al. (2022). It is designed to produce valid inference on a *feature* of conditional average treatment effects (CATEs), even when valid inference on the CATEs themselves is infeasible. The calculation requires treatment effect predictions as input but is *generic* with respect to the machine learning method used to generate them; we use regression forests. Deryugina et al. (2019) use

CLAN to compare the attributes of Medicare beneficiaires most and least affected by pollution. We follow the algorithm described in their appendix and numbered below:

1. Starting with the full sample of accounts observed in post year 2, split the sample in two equal parts randomly

2. Using one subsample, grow two regression forests to predict Y using X: one using only control group observations and one using only treatment group observations. These forests consist of 10,000 trees each and use by-wave treatment fractions as sample weights

3. Using the other subsample, predict Y according to each prediction (treatment and control) model from Step 2

4. Subtract the control prediction from the treatment prediction to obtain a treatment effect prediction $S(X)$

We iterate over this process 100 times to obtain 100 treatment effect predictions for each household. We then calculate the mean of these 100 predictions for each household. Sorting the households from lowest mean treatment effect to highest, we focus on the bottom quintile (with the largest reductions) and the top quintile (with treatment effects near zero). In Table 3, we compare mean characteristics in these two groups, estimating a difference in means by regressing a characteristic (from X) on a binary variable equaling one if a household is in the top quintile; standard errors are clustered at the zip code level.

C.3 Heterogeneity curves

In Section 3.2 of the paper, we describe the construction of "heterogeneity curves" depicting non-parametric relationships between predicted treatment effect and individual attributes. We present those curves in Figure 7. Here, we provide further detail on our procedure for generating them.

We follow Knaus (2022), who reviews and extends "double machine learning" methods – which rely on the doubly-robust scores of Robins et al. (1994) – and develops an R-package ('cDML') that facilitates use of the scores to estimate parameters of interest. Heterogeneity curves are one such estimation procedure integrated into cDML. We first run the main cDML function that computes nuisance parameters, scores, and average treatment effects. Like Knaus (2022), we compute the nuisance parameters (the conditional mean and propensity score as introduced in Section 2.2) via random forest, with the same forest parameterization choices as in our main causal forest. Next, we run spline regressions through the cDML package. The regressions are based on the 'crs' R-package (Racine and Nie, 2022), and they use B-splines with cross-validated degree and number of knots.

C.4 Targeting to raise social welfare

In Section 4 of the paper, we describe a pair of targeting exercises to test the forest's ability to generate welfare improvements. We reprint the targeting algorithm here, now with additional details where appropriate. We describe our procedure for generating confidence intervals on these welfare improvements in the next subsection.

1. *Split the full sample of available households into two: a training set for estimating the model, and a test set for targeting and its evaluation.*

The splitting rule here depends on which of the two versions of the targeting exercise (described in Section 4) is being carried out. In the first version, the training and test sets are randomly drawn, mutually exclusive 50-percent subsamples; in the second, the training set is all households whose program wave started in 2014 and the test set is households from waves beginning later.

2. *Estimate a predictive model with the training sample.*

The forest is our main predictive model of interest, but we also estimate a lasso model and four regression models, as described in the text. We use the "glmnet" package in R (Hastie et al., 2021) to implement the lasso and predict household treatment effects from it. The lasso-based glmnet seeks to find the coefficient values (zero and otherwise) that minimize mean squared error plus a penalty for the sum of coefficient absolute values exceeding an arbitrary threshold. We use 10-fold cross-validated lasso to find the threshold value (λ_{min}) that yields the minimum mean cross-validated error (Hastie et al., 2021). Our lasso regressions use cross-sectional data, to mimic the cross-sectional nature of the forest. Our dependent variable is the difference in electricity consumption between the second year of the program and the year prior to its start; this takes advantage of the information contained in our panel data, as in the forest. The set of potential predictor variables consists of treatment status, the X vector, and all interactions between elements of the two. We additionally include inverse probability weights by wave.

We estimate all four regression models with the same implementation choices as with the lasso: we use weighted least squares with inverse probability weights; we use cross-sectional regressions, and we use the pre-post (year 2) difference in electricity consumption as our dependent variable. The first model, which we call the "Baseline", has the following form:

$$Y_{Di} = \alpha + \beta_0 T_i + \left(\beta_1 T_i * Y_{0i}\right) + \epsilon_i \tag{1}$$

where Y_{Di} is the difference between electricity consumption in the second year of the program and the year prior to program start, T_i is a treatment assignment binary variable, Y_{0i} is electricity consumption in the year prior to program start, and ϵ_i is an error term. Each successive model beyond the "Baseline" has more explanatory terms than the last. The "Linear" model interacts treatment with all fifteen household characteristics:

$$Y_{Di} = \alpha_0 + \beta_0 T_i + \Sigma_{j=1}^{15}\big(\beta_k T_i * X_{ji}\big) + \epsilon_i \tag{2}$$

where j indexes the characteristics in the vector X_i. The "Parsimonious" model adds interactions between treatment and the square of each each characteristic:

$$Y_{Di} = \alpha_0 + \beta_0 T_i + \Sigma_{j=1}^{15}\big(\beta_k T_i * X_{ji}\big) + \Sigma_{j=1}^{15}\big(\gamma_k T_i * (X_{ji})^2\big) + \epsilon_i \tag{3}$$

Finally, the "Interacted" model includes interactions between treatment and the product of each pair of characteristics:

$$Y_{Di} = \alpha_0 + \beta_0 T_i + \Sigma_{j=1}^{15}\big(\beta_j T_i * X_{ji}\big) + \Sigma_{j=1}^{15}\big(T_i * X_{ji} * \Sigma_{k=j}^{15}\big(\delta_{jk} X_{ki}\big)\big) + \epsilon_i \tag{4}$$

3. *In the test sample, predict household-level treatment effects (using the model estimated in Step 2) and willingness to pay (using the model with parameters taken from Allcott and Kessler (2019) described above).*

We generate household treatment-effect predictions in the test sample using the now-estimated forest and regression models. We also estimate willigness to pay (WTP) for HERs as a function of household attributes. We use coefficients from Allcott and Kessler (2019), who elicit WTP for HERs experimentally, to calibrate a model of WTP. Allcott and Kessler (2019) report results from a regression of household-specific WTP on the logarithm of income, indicators for retirement, marriage, homeownership, and single-family occupancy, and homebuyer's credit worthiness score. Our data do not match up perfectly to theirs, but we do have measures of income, age, number of adults in the household, homeownership, and single-family occupancy. We define households with a head-of-household that is older than 65 as "retired." We define households with at least two adults living in the household as "married." Allcott and Kessler (2019) do not report a constant term for the regression but do report an average WTP. We thus use, as our own constant term, the difference between their reported mean WTP and the fitted mean value in our data using their regression coefficients. The exact equation is

$$\begin{aligned} WTP_i = 0.0603 * log(Income_i) - 1.588 * Retired_i \\ + 0.683 * Married_i - 0.780 * Rent_i + 0.322 * Single_i \end{aligned} \tag{5}$$

4. Calculate predicted social benefits for each household according to Equation 6.

For convenience, we reprint and describe Equation 6 below:

$$SNB_i = -TE_i * 12 * SMC_e - MC_{HER} + WTP_i, \tag{6}$$

TE_i is the predicted monthly treatment effect for household i, which we multiply by 12 to convert to an annual number; SMC_e is the social marginal cost of electricity (which includes both generation costs and environmental externalities); MC_{HER} is the marginal cost of sending a household HERs for one year; and WTP_i is a household's annual willingness to pay for HERs. We set $SMC_e =$ \$0.065 per kWh, which is the short-run estimate of Borenstein and Bushnell (2022) for the trio of states in our sample in 2016. We set $MC =$ \$7.00 per year, based on consultation with Eversource.

5. Identify all test-sample households whose induced social benefits exceed marginal cost; this is the group "targeted" for treatment.

6. Estimate an average treatment effect (ATE) in the targeted group.

For each targeting method (the forest and four regressions), we estimate an ATE in the targeted sample using cross-sectional regression of Y_i on T_i with inverse probability weights. This specification is analogous to the four predictive regression models used in Step 2, with the latter additionally including interaction terms to predict conditional ATEs.

7. Calculate "actual" aggregate social net benefits in the targeted group according to Equation 6, but replacing each targeted household's predicted TE with the estimated ATE from Step 7.

C.5 Bootstrapping to generate confidence intervals

To conduct inference, we use bootstrapping to generate confidence intervals on our estimates of the forest's performance relative to other models. Following Gerarden and Yang (2022), we carry out two versions of a 1,000-iteration bootstrap. In the first, which we feature in Figure 9 as well as Appendix Figures B4-B7, we fix each model's targeting rule (that is, treatment assignment) as the rule estimated from the original sample. Every bootstrap is a re-drawing (with replacement, and keeping the same N and treatment-control ratio as its parent sample) of the test set used to estimate actual social net benefits produced by each model. As Gerarden and Yang (2022) note, this approach does not account for the effect of sampling variation on the specific targeting rule estimated. Our second version of the bootstrap, featured in Appendix Figure B3, does account for this effect: every bootstrap is a re-drawing with replacement of the training and test sets, and the training set is used to develop a new targeting rule, which is then populated by the test set to estimate benefits.

From the 1,000 iterations of either version, we identify the 2.5 and 97.5 quantiles of estimated social net benefits. Then, we compute 95 percent confidence intervals as

$$[CI_L, CL_H] = [\hat{W} - (\bar{W} - W_{2.5}), \hat{W} + (W_{97.5} - \bar{W})] \tag{7}$$

where $\hat{W}$ is our main point estimate of net social welfare (bar length, in Figure 9 and all Appendix targeting figures) for a given method, $\bar{W}$ is the mean of our bootstrapped estimates of the same measure, and $W_{2.5}$ and $W_{97.5}$ are the required quantiles of the bootstrapped distribution.