GRF for Saga/Match Treatment Heterogeneity

This document presents the current `GRF` (Generalized Random Forests) analysis for treatment heterogeneity. We present results for the following outcomes:

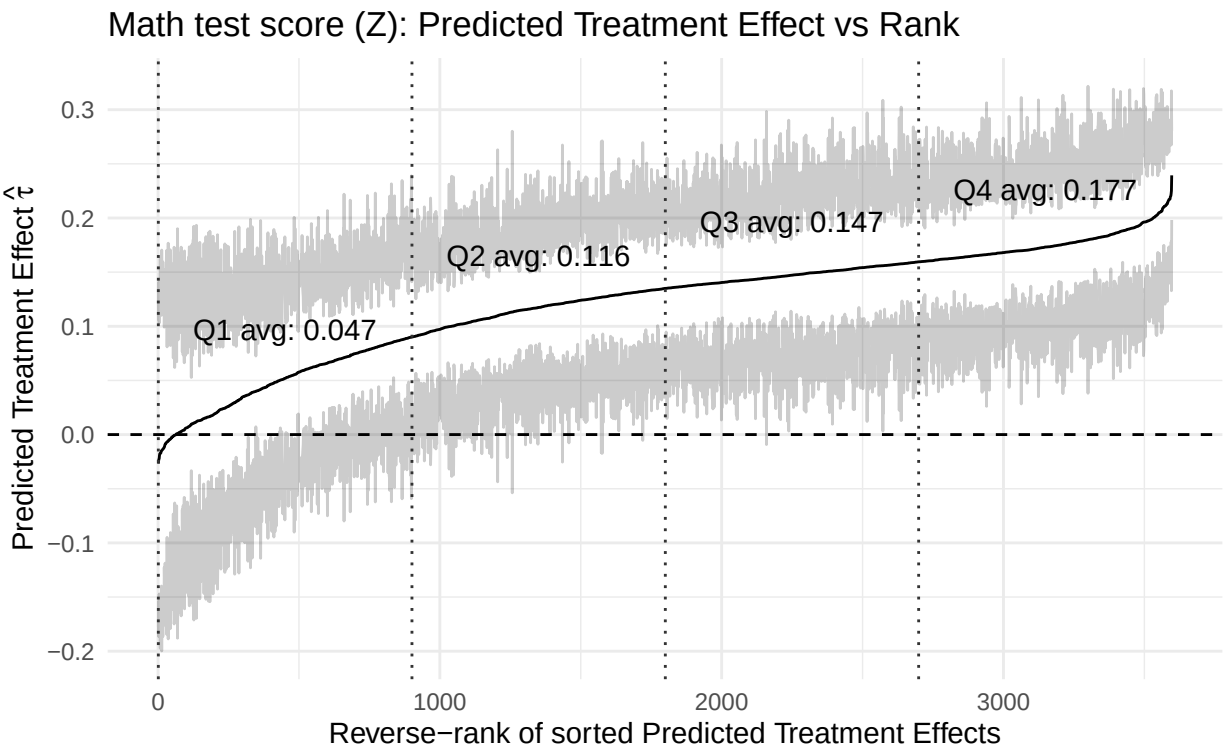
- Math Test Scores
- Math Class Failures
- Math GPA
- Reading Test Scores
- Non-math GPA
- Graduated on-time
- Ever graduated
- Participation in Saga tutoring in year 2

We use the most recent release of the `GRF` package by Tibshirani, Athey, et. al. We try to follow the example of Jon Davis and Sara Heller wherever possible, but in the interceding years there has been some updates in the underlying package, so there are some differences. Some implementation details:

- We use a training sample split (`sample.fraction`) of 0.50 when building each tree in each forest (as per Davis+Heller)
  - Davis & Heller used 0.8, but the documentation for the package notes "when variance estimates are requested, `sample.fraction` cannot be greater than 0.5"
- For each outcome of interest, we grow 20,000 trees to make each random forest
- Following Davis + Heller, we "adjust for differences in treatment probabilities [by] using inverse probability weights throughout the procedure" (following their calculation)
- Jon Davis + Sara Heller dealt with missingness in covariates by imputing block means and including missingness dummies
  - Since then, the underlying code has been updated with its own methods to deal with missingness
- The package now supports 'clustering', so we cluster observations at the individual level to account for multiple observations of students randomized multiple times in study 2 (ensuring the same student can't be in both test/train splits when fitting each individual tree)

We use all covariates from our main analyses (except for randomization block). These include gender, age, learning disability, free lunch recipient, race, baseline grade level, GPA, baseline test performance (and within-baseline-school math test decile), days absent from school, disciplinary incidents, including suspensions, and arrests.

Outcome: Math test score (Z)



Our causal forest estimates an overall average treatment effect of 0.122 (0.026).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.001 and a 'differential forest prediction' (DFP) coefficient of 0.684, with a corresponding DFP p-value of 0.103.[1]

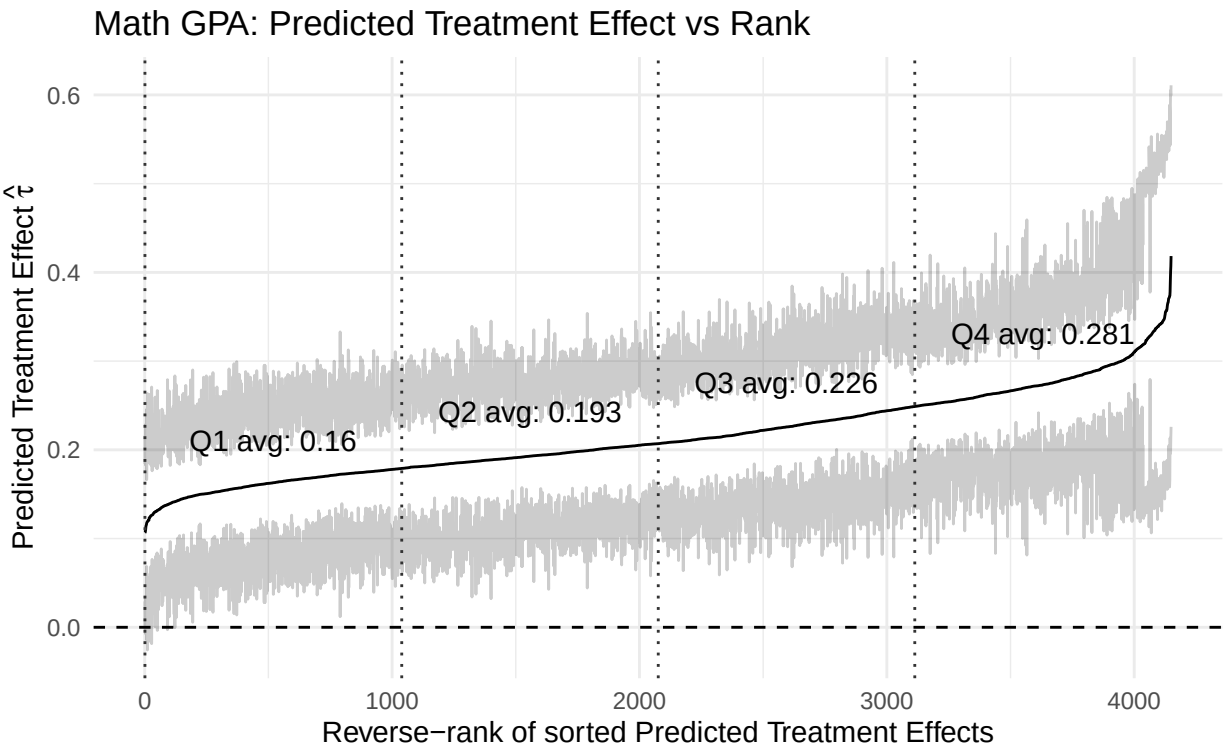
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.063, 0.154]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.076, 0.125]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.013, 0.274].

[1] A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 1: Summary table by Quartile of Predicted Treatment Effects on Math test score (Z)

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.047	0.116	0.147	0.177
N	900.000	899.000	899.000	899.000
Age	14.798	14.616	14.491	14.284
Female	0.147	0.148	0.140	0.201
Has IEP	0.338	0.231	0.087	0.023
Has Free/Reduced Lunch	0.892	0.908	0.875	0.897
Black	0.602	0.590	0.513	0.335
Hispanic	0.363	0.379	0.447	0.595
Other Race	0.034	0.031	0.040	0.070
In 9th Grade	0.707	0.724	0.675	0.809
In 10th Grade	0.288	0.275	0.325	0.191
Baseline GPA	1.679	2.209	2.646	2.896
Num. A's	2.739	5.255	7.705	13.692
Num. B's	3.507	7.440	8.977	9.901
Num. C's	7.062	9.830	8.199	7.854
Num. D's	4.642	3.769	2.365	2.311
Num. F's	4.484	2.646	1.280	1.078
Missing Baseline GPA/Grades	0.176	0.044	0.050	0.006
Days Absent	23.141	15.534	9.699	7.703
Missing Attendance Data	0.133	0.000	0.000	0.000
Math Test Score (Z)	-1.243	-0.143	0.484	0.562
Reading Test Score (Z)	-0.915	-0.034	0.433	0.316
Missing Math Test	0.309	0.017	0.003	0.000
Missing Reading Test	0.310	0.022	0.007	0.001
Out-of-School Suspensions	2.449	0.992	0.242	0.082
Disciplinary Incidents	1.860	0.829	0.239	0.090
Any Arrests at Baseline	0.191	0.125	0.067	0.029
Arrests: Violent Crime	0.131	0.063	0.029	0.009
Arrests: Property Crime	0.053	0.034	0.021	0.009
Arrests: Drug Crime	0.046	0.020	0.008	0.003
Math Score - Decile in Previous School	2.443	5.336	7.241	7.340
In Study 2	0.442	0.478	0.471	0.566
Participated in Year 1 of Study	0.296	0.281	0.230	0.263

Outcome: Math GPA



Our causal forest estimates an overall average treatment effect of 0.216 (0.029).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.007 and a 'differential forest prediction' (DFP) coefficient of 0.68, with a corresponding DFP p-value of 0.155.[2]

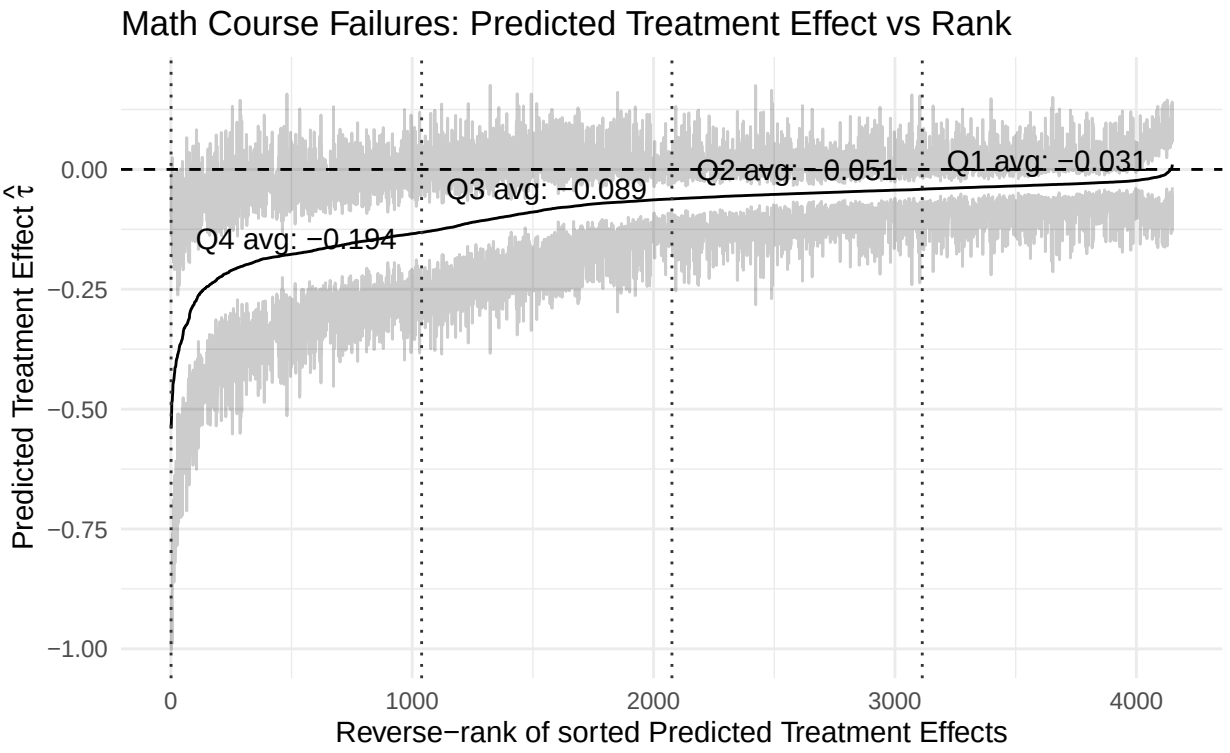
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.062, 0.206]. The 95% confidence interval for the difference between the above-median and below-median group is [0.019, 0.249]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.058, 0.273].

[2]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 2: Summary table by Quartile of Predicted Treatment Effects on Math GPA

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.160	0.193	0.226	0.281
N	1038.000	1037.000	1037.000	1037.000
Age	14.342	14.314	14.797	15.136
Female	0.187	0.234	0.134	0.068
Has IEP	0.250	0.162	0.179	0.098
Has Free/Reduced Lunch	0.986	0.936	0.848	0.784
Black	0.581	0.580	0.526	0.390
Hispanic	0.398	0.379	0.442	0.527
Other Race	0.021	0.041	0.033	0.083
In 9th Grade	1.000	0.982	0.591	0.228
In 10th Grade	0.000	0.016	0.395	0.767
Baseline GPA	2.180	2.476	2.251	2.064
Num. A's	8.405	9.905	6.342	1.398
Num. B's	7.785	9.786	7.030	2.275
Num. C's	12.040	9.983	5.843	3.019
Num. D's	6.182	2.672	2.278	2.208
Num. F's	4.455	2.981	2.604	1.025
Missing Baseline GPA/Grades	0.000	0.002	0.031	0.286
Days Absent	15.729	15.670	20.478	17.969
Missing Attendance Data	0.000	0.000	0.002	0.176
Math Test Score (Z)	-0.186	0.046	-0.105	0.117
Reading Test Score (Z)	-0.157	0.017	-0.057	0.089
Missing Math Test	0.060	0.068	0.156	0.230
Missing Reading Test	0.064	0.074	0.159	0.231
Out-of-School Suspensions	1.339	1.392	1.795	0.747
Disciplinary Incidents	0.808	0.891	1.559	0.763
Any Arrests at Baseline	0.199	0.116	0.206	0.097
Arrests: Violent Crime	0.163	0.100	0.141	0.032
Arrests: Property Crime	0.071	0.062	0.095	0.031
Arrests: Drug Crime	0.039	0.048	0.095	0.024
Math Score - Decile in Previous School	5.263	5.906	5.551	6.487
In Study 2	0.615	0.600	0.439	0.227
Participated in Year 1 of Study	0.294	0.270	0.237	0.247

Outcome: Math Course Failures



Our causal forest estimates an overall average treatment effect of -0.092 (0.02).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.021 and a 'differential forest prediction' (DFP) coefficient of 0.823, with a corresponding DFP p-value of 0.031.[3]

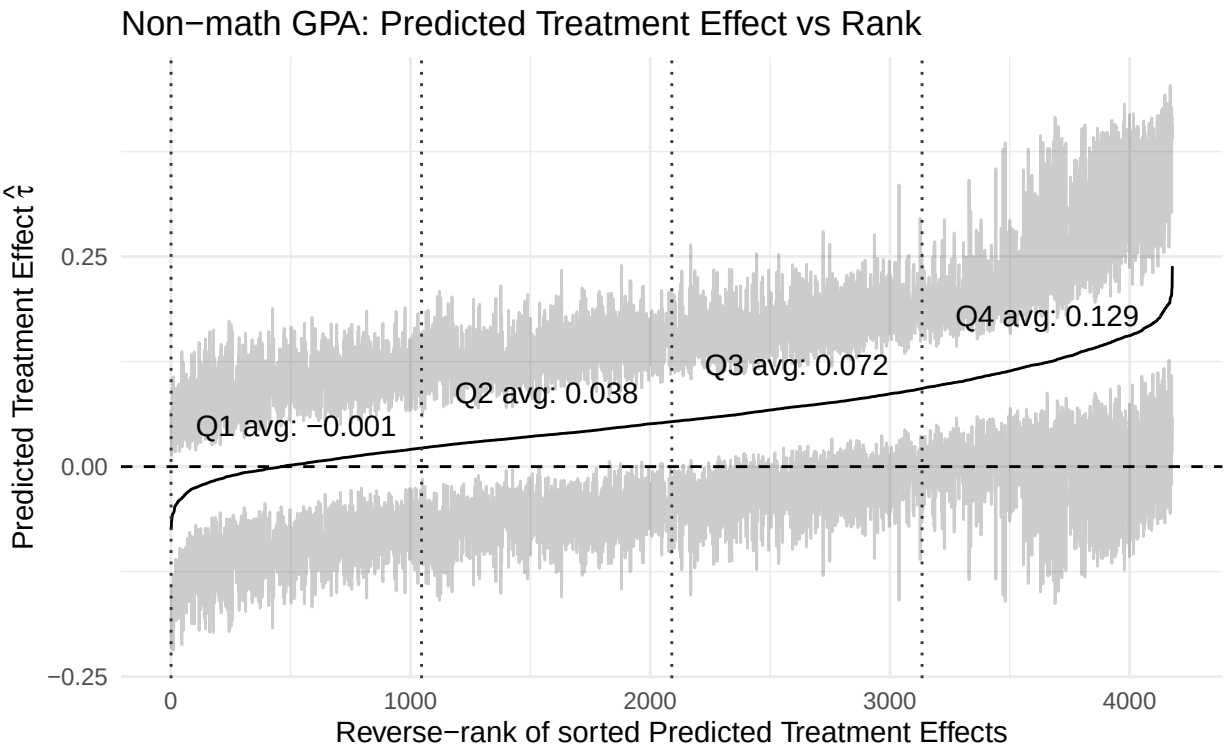
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.139, 0.083]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.165, -0.009]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.146, 0.087].

[3]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 3: Summary table by Quartile of Predicted Treatment Effects on Math Course Failures

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.031	-0.051	-0.089	-0.194
N	1037.000	1037.000	1037.000	1038.000
Age	14.293	14.518	14.722	15.056
Female	0.151	0.200	0.174	0.099
Has IEP	0.123	0.214	0.202	0.151
Has Free/Reduced Lunch	0.912	0.883	0.848	0.910
Black	0.484	0.501	0.555	0.535
Hispanic	0.462	0.452	0.395	0.436
Other Race	0.054	0.046	0.049	0.029
In 9th Grade	0.885	0.714	0.649	0.553
In 10th Grade	0.114	0.285	0.348	0.431
Baseline GPA	2.859	2.761	2.136	1.111
Num. A's	11.998	8.924	3.997	1.139
Num. B's	10.718	9.139	5.243	1.781
Num. C's	9.865	9.355	7.377	4.296
Num. D's	2.767	2.290	3.236	5.048
Num. F's	0.936	1.133	2.246	6.747
Missing Baseline GPA/Grades	0.001	0.010	0.184	0.124
Days Absent	9.988	9.467	17.999	32.654
Missing Attendance Data	0.001	0.010	0.143	0.025
Math Test Score (Z)	0.629	-0.104	-0.392	-0.388
Reading Test Score (Z)	0.422	-0.022	-0.267	-0.342
Missing Math Test	0.067	0.031	0.224	0.194
Missing Reading Test	0.072	0.032	0.226	0.198
Out-of-School Suspensions	0.407	0.406	1.231	3.356
Disciplinary Incidents	0.279	0.307	0.989	2.516
Any Arrests at Baseline	0.076	0.069	0.170	0.303
Arrests: Violent Crime	0.031	0.032	0.134	0.239
Arrests: Property Crime	0.027	0.019	0.069	0.143
Arrests: Drug Crime	0.014	0.014	0.040	0.140
Math Score - Decile in Previous School	7.667	5.428	4.656	5.005
In Study 2	0.551	0.508	0.489	0.332
Participated in Year 1 of Study	0.238	0.276	0.279	0.255

Outcome: Non-math GPA



Our causal forest estimates an overall average treatment effect of 0.059 (0.023).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.005 and a 'differential forest prediction' (DFP) coefficient of 0.149, with a corresponding DFP p-value of 0.383.[4]

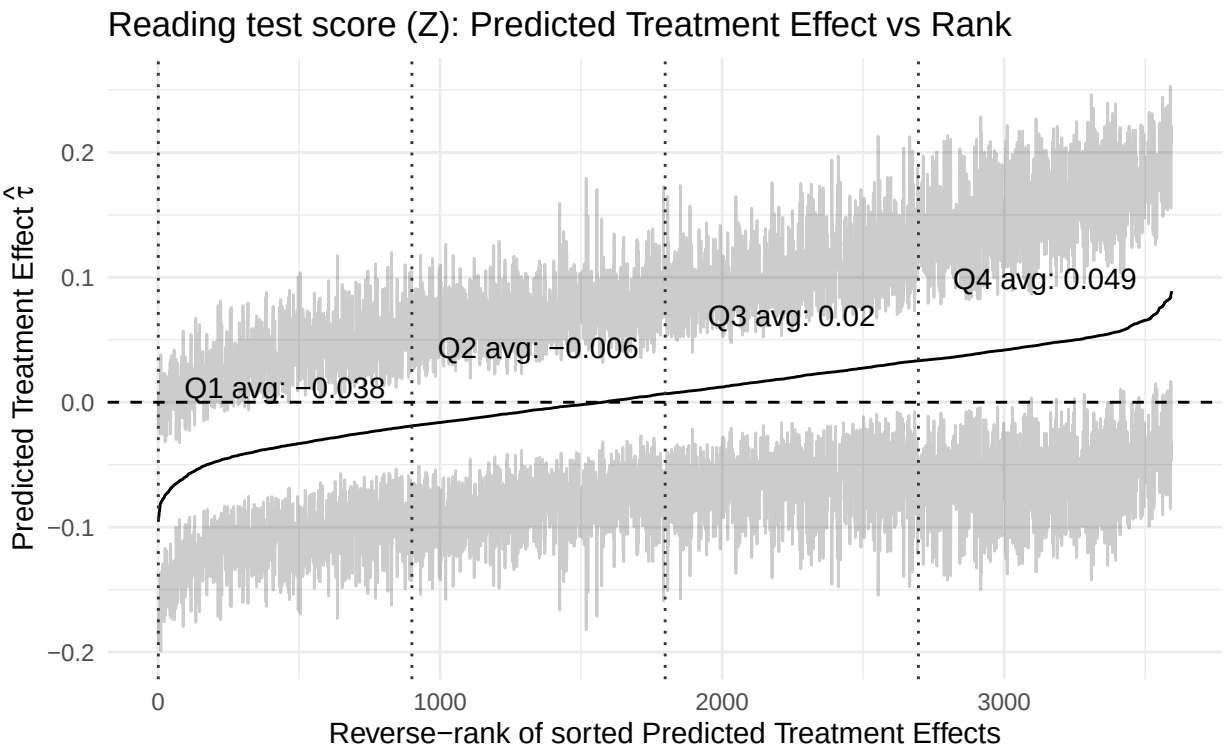
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.021, 0.194]. The 95% confidence interval for the difference between the above-median and below-median group is [0.005, 0.187]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.025, 0.239].

[4]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 4: Summary table by Quartile of Predicted Treatment Effects on Non-math GPA

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.001	0.038	0.072	0.129
N	1045.000	1044.000	1044.000	1045.000
Age	14.481	14.597	14.570	14.964
Female	0.111	0.201	0.176	0.135
Has IEP	0.160	0.203	0.206	0.121
Has Free/Reduced Lunch	0.922	0.894	0.893	0.848
Black	0.460	0.513	0.522	0.584
Hispanic	0.500	0.445	0.428	0.369
Other Race	0.039	0.041	0.050	0.047
In 9th Grade	0.741	0.710	0.752	0.594
In 10th Grade	0.259	0.287	0.245	0.389
Baseline GPA	2.472	2.375	2.362	1.642
Num. A's	6.622	7.023	8.509	3.778
Num. B's	9.206	7.514	7.166	2.861
Num. C's	11.158	9.525	7.483	2.621
Num. D's	3.744	3.729	3.158	2.733
Num. F's	1.277	2.209	3.050	4.674
Missing Baseline GPA/Grades	0.000	0.004	0.065	0.249
Days Absent	10.840	12.570	16.165	34.148
Missing Attendance Data	0.000	0.000	0.001	0.177
Math Test Score (Z)	0.173	-0.161	-0.170	0.039
Reading Test Score (Z)	0.226	-0.142	-0.181	-0.012
Missing Math Test	0.104	0.049	0.052	0.319
Missing Reading Test	0.109	0.049	0.057	0.322
Out-of-School Suspensions	0.337	0.540	0.993	4.134
Disciplinary Incidents	0.330	0.468	0.704	2.984
Any Arrests at Baseline	0.058	0.108	0.151	0.312
Arrests: Violent Crime	0.022	0.054	0.124	0.251
Arrests: Property Crime	0.022	0.030	0.073	0.138
Arrests: Drug Crime	0.010	0.019	0.035	0.146
Math Score - Decile in Previous School	6.384	5.334	5.328	6.191
In Study 2	0.466	0.526	0.510	0.377
Participated in Year 1 of Study	0.279	0.292	0.266	0.205

Outcome: Reading test score (Z)



Our causal forest estimates an overall average treatment effect of 0.006 (0.027).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.156 and a 'differential forest prediction' (DFP) coefficient of -1.561, with a corresponding DFP p-value of 0.964.[5]

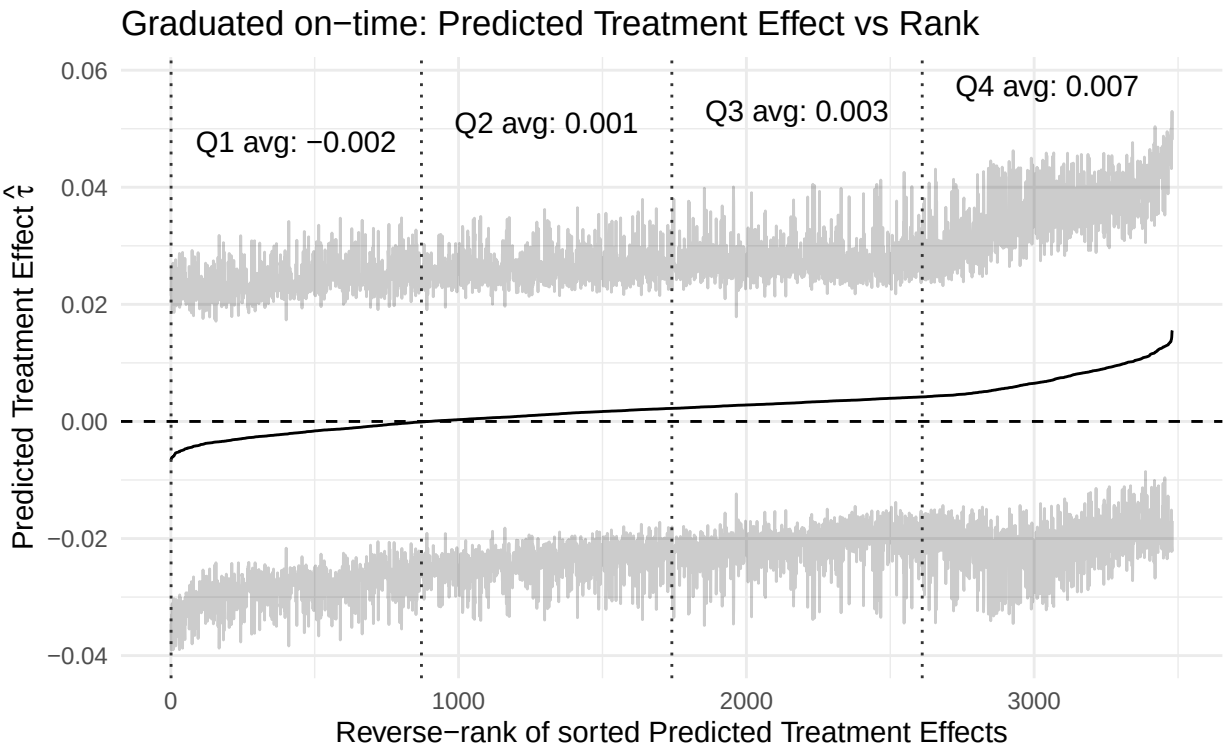
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.315, -0.035]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.102, 0.111]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.284, 0.027].

[5]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 5: Summary table by Quartile of Predicted Treatment Effects on Reading test score (Z)

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.038	-0.006	0.020	0.049
N	899.000	898.000	898.000	899.000
Age	14.691	14.624	14.511	14.363
Female	0.119	0.171	0.163	0.185
Has IEP	0.335	0.200	0.098	0.044
Has Free/Reduced Lunch	0.934	0.923	0.878	0.834
Black	0.495	0.539	0.552	0.455
Hispanic	0.483	0.429	0.400	0.472
Other Race	0.022	0.032	0.048	0.073
In 9th Grade	0.654	0.716	0.762	0.782
In 10th Grade	0.341	0.283	0.238	0.217
Baseline GPA	1.799	2.247	2.620	2.967
Num. A's	3.577	6.228	8.656	10.958
Num. B's	4.610	7.089	9.675	8.491
Num. C's	8.058	9.442	9.197	6.263
Num. D's	5.416	3.759	2.442	1.454
Num. F's	4.358	2.590	1.566	0.892
Missing Baseline GPA/Grades	0.034	0.036	0.047	0.159
Days Absent	20.082	15.712	11.387	6.608
Missing Attendance Data	0.000	0.000	0.008	0.127
Math Test Score (Z)	-0.837	-0.308	0.344	1.037
Reading Test Score (Z)	-0.740	-0.207	0.338	0.865
Missing Math Test	0.001	0.069	0.086	0.174
Missing Reading Test	0.002	0.075	0.087	0.177
Out-of-School Suspensions	1.446	0.997	0.842	0.182
Disciplinary Incidents	1.270	0.791	0.524	0.195
Any Arrests at Baseline	0.154	0.122	0.088	0.049
Arrests: Violent Crime	0.090	0.073	0.051	0.022
Arrests: Property Crime	0.046	0.028	0.031	0.014
Arrests: Drug Crime	0.037	0.018	0.019	0.006
Math Score - Decile in Previous School	3.487	4.890	6.882	8.705
In Study 2	0.402	0.458	0.533	0.564
Participated in Year 1 of Study	0.307	0.273	0.275	0.212

Outcome: Graduated on-time



Our causal forest estimates an overall average treatment effect of 0.002 (0.012).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 0.871 and a 'differential forest prediction' (DFP) coefficient of -11.162, with a corresponding DFP p-value of 0.999.[6]

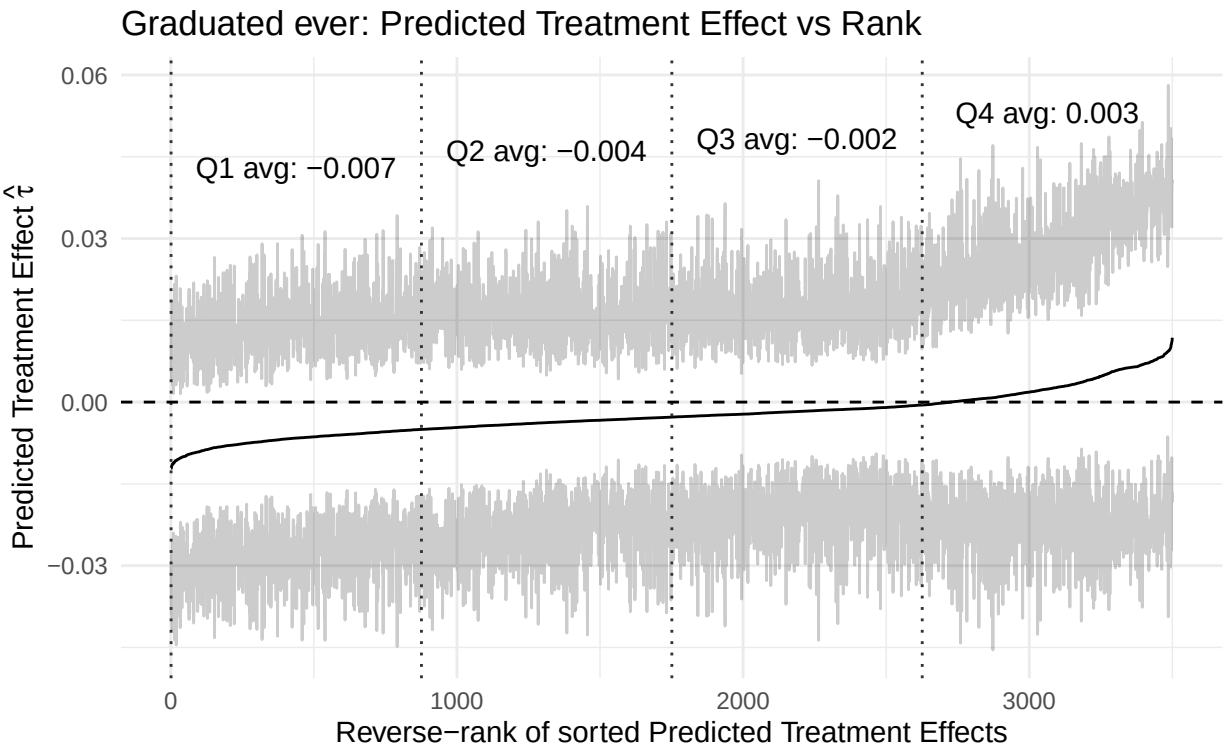
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.121, 0.001]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.102, -0.004]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.145, 0.001].

[6]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 6: Summary table by Quartile of Predicted Treatment Effects on Graduated on-time

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.002	0.001	0.003	0.007
N	870.000	870.000	870.000	870.000
Age	15.060	14.499	14.152	14.634
Female	0.057	0.175	0.200	0.194
Has IEP	0.174	0.185	0.166	0.174
Has Free/Reduced Lunch	0.869	0.910	0.876	0.897
Black	0.501	0.546	0.451	0.663
Hispanic	0.452	0.424	0.493	0.310
Other Race	0.047	0.030	0.056	0.026
In 9th Grade	0.299	0.822	0.934	0.766
In 10th Grade	0.698	0.175	0.063	0.228
Baseline GPA	2.285	2.477	2.720	1.942
Num. A's	2.733	7.409	12.300	6.282
Num. B's	5.015	8.111	10.447	5.472
Num. C's	6.516	9.230	8.845	6.968
Num. D's	3.247	3.120	2.538	3.768
Num. F's	1.152	2.269	1.841	4.166
Missing Baseline GPA/Grades	0.043	0.070	0.053	0.107
Days Absent	12.897	12.359	10.583	26.182
Missing Attendance Data	0.000	0.003	0.024	0.097
Math Test Score (Z)	0.054	-0.037	0.175	-0.120
Reading Test Score (Z)	0.074	-0.012	0.139	-0.178
Missing Math Test	0.071	0.066	0.069	0.207
Missing Reading Test	0.074	0.068	0.074	0.215
Out-of-School Suspensions	0.008	0.054	0.115	4.756
Disciplinary Incidents	0.083	0.167	0.274	3.233
Any Arrests at Baseline	0.076	0.077	0.051	0.334
Arrests: Violent Crime	0.037	0.051	0.026	0.286
Arrests: Property Crime	0.029	0.052	0.025	0.162
Arrests: Drug Crime	0.011	0.015	0.016	0.123
Math Score - Decile in Previous School	6.099	5.628	6.269	5.629
In Study 2	0.285	0.533	0.591	0.514
Participated in Year 1 of Study	0.267	0.237	0.253	0.237

Outcome: Graduated ever



Our causal forest estimates an overall average treatment effect of -0.002 (0.011).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.053 and a 'differential forest prediction' (DFP) coefficient of -10.921, with a corresponding DFP p-value of 0.999.[7]

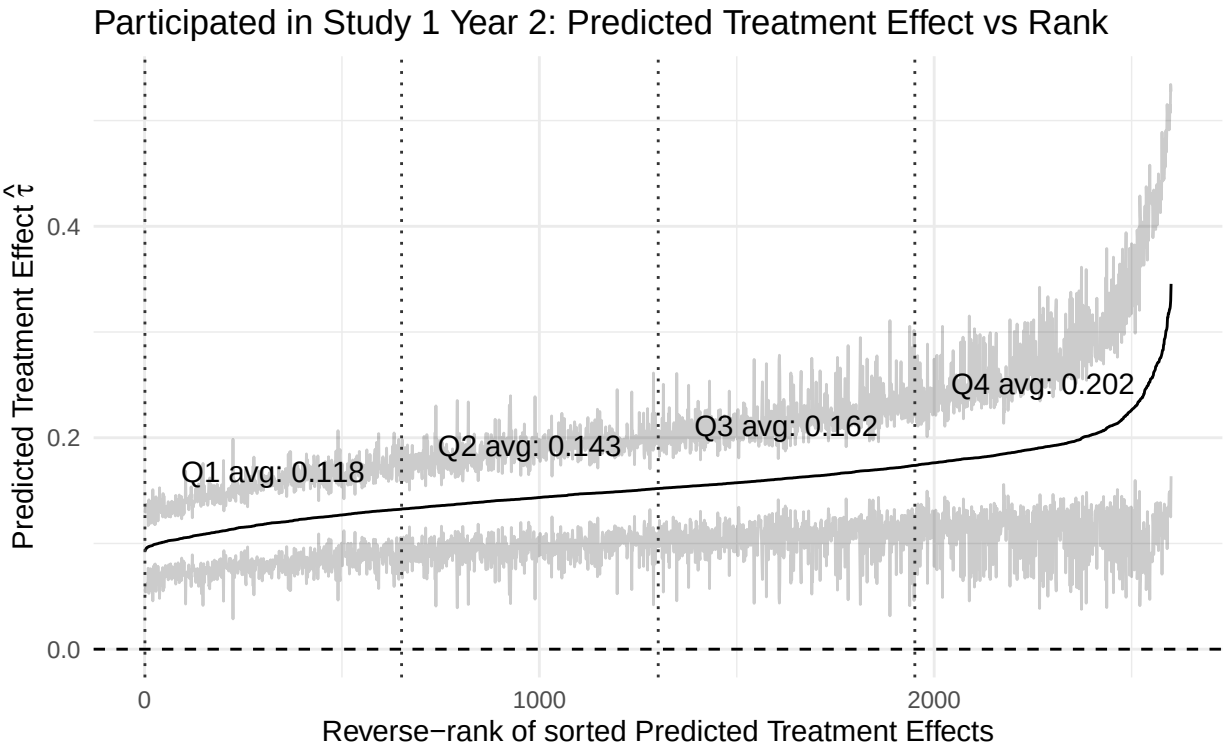
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.112, 0.005]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.089, -0.001]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.15, -0.012].

[7]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 7: Summary table by Quartile of Predicted Treatment Effects on Graduated ever

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	-0.007	-0.004	-0.002	0.003
N	875.000	875.000	875.000	875.000
Age	14.627	14.534	14.350	14.825
Female	0.101	0.184	0.179	0.161
Has IEP	0.129	0.139	0.174	0.255
Has Free/Reduced Lunch	0.905	0.880	0.891	0.873
Black	0.583	0.547	0.434	0.597
Hispanic	0.385	0.422	0.514	0.357
Other Race	0.032	0.031	0.051	0.047
In 9th Grade	0.646	0.694	0.819	0.647
In 10th Grade	0.350	0.302	0.173	0.333
Baseline GPA	2.112	2.522	2.861	1.873
Num. A's	3.934	7.109	12.907	4.765
Num. B's	6.394	7.871	10.195	4.528
Num. C's	10.333	8.462	7.544	5.222
Num. D's	4.712	2.823	2.150	3.027
Num. F's	2.721	1.790	1.195	3.807
Missing Baseline GPA/Grades	0.007	0.065	0.023	0.178
Days Absent	12.709	11.750	10.042	28.282
Missing Attendance Data	0.000	0.000	0.000	0.123
Math Test Score (Z)	0.252	0.234	0.222	-1.037
Reading Test Score (Z)	0.102	0.198	0.218	-0.782
Missing Math Test	0.000	0.003	0.019	0.390
Missing Reading Test	0.001	0.006	0.025	0.398
Out-of-School Suspensions	0.391	0.534	0.613	3.372
Disciplinary Incidents	0.254	0.487	0.545	2.460
Any Arrests at Baseline	0.101	0.088	0.074	0.279
Arrests: Violent Crime	0.056	0.050	0.041	0.262
Arrests: Property Crime	0.029	0.042	0.027	0.169
Arrests: Drug Crime	0.013	0.019	0.023	0.120
Math Score - Decile in Previous School	6.781	6.529	6.205	2.839
In Study 2	0.449	0.520	0.525	0.434
Participated in Year 1 of Study	0.259	0.232	0.249	0.247

Outcome: Participated in Study 1 Year 2



Our causal forest estimates an overall average treatment effect of 0.156 (0.01).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 0.998 and a 'differential forest prediction' (DFP) coefficient of 1.445, with a corresponding DFP p-value of 0.[8]

The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [0.031, 0.132]. The 95% confidence interval for the difference between the above-median and below-median group is [0.039, 0.119]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [0.047, 0.16].

[8]A coefficient of 1 for MFP suggests the mean forest prediction is correct, and a DFP coefficient of 1 'additionally suggests that the forest has captured heterogeneity in the underlying signal.' The p-value from the DFP estimate 'acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity'.

Table 8: Summary table by Quartile of Predicted Treatment Effects on Participated in Study 1 Year 2

Baseline	$\hat{\tau}$ Quartile 1	$\hat{\tau}$ Quartile 2	$\hat{\tau}$ Quartile 3	$\hat{\tau}$ Quartile 4
Mean $\hat{\tau}$	0.118	0.143	0.162	0.202
N	650.000	650.000	650.000	650.000
Age	15.014	14.820	14.732	14.675
Female	0.002	0.000	0.000	0.003
Has IEP	0.066	0.075	0.177	0.366
Has Free/Reduced Lunch	0.852	0.908	0.875	0.826
Black	0.398	0.478	0.520	0.465
Hispanic	0.602	0.511	0.457	0.362
Other Race	0.000	0.011	0.023	0.174
In 9th Grade	0.503	0.518	0.568	0.655
In 10th Grade	0.472	0.466	0.429	0.340
Baseline GPA	2.040	1.989	2.163	2.212
Num. A's	7.075	4.171	4.668	4.983
Num. B's	5.232	5.343	5.774	6.482
Num. C's	1.985	5.992	8.203	9.797
Num. D's	1.580	3.586	3.963	4.109
Num. F's	3.888	3.222	2.588	1.858
Missing Baseline GPA/Grades	0.083	0.055	0.062	0.091
Days Absent	29.787	19.882	16.852	16.276
Missing Attendance Data	0.000	0.002	0.035	0.088
Math Test Score (Z)	0.713	0.282	-0.163	-0.887
Reading Test Score (Z)	0.627	0.137	-0.306	-0.412
Missing Math Test	0.195	0.135	0.122	0.169
Missing Reading Test	0.198	0.137	0.125	0.168
Out-of-School Suspensions	2.335	1.461	1.694	1.069
Disciplinary Incidents	1.763	1.296	1.161	0.904
Any Arrests at Baseline	0.277	0.191	0.155	0.111
Arrests: Violent Crime	0.225	0.145	0.094	0.078
Arrests: Property Crime	0.151	0.083	0.065	0.038
Arrests: Drug Crime	0.155	0.057	0.034	0.028
Math Score - Decile in Previous School	8.314	7.310	5.711	3.220
In Study 2	0.000	0.000	0.000	0.000
Participated in Year 1 of Study	0.129	0.195	0.257	0.231