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:

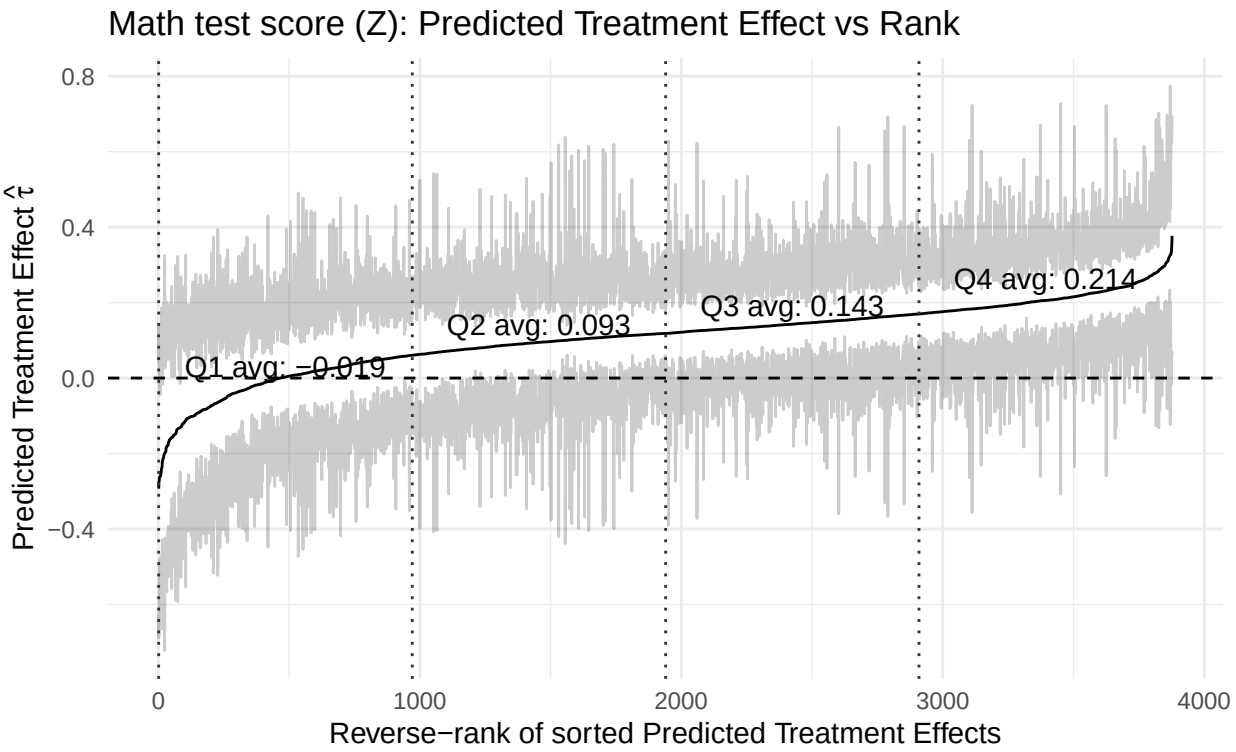
- Math Test Scores
- Math Class Failures
- Math GPA
- Reading Test Scores
- Non-math GPA
- Graduated on-time
- Ever graduated
- Participation in Saga tutoring in year 2

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 0.107 (0.025).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.003 and a 'differential forest prediction' (DFP) coefficient of 0.406, with a corresponding DFP p-value of 0.075.[1]

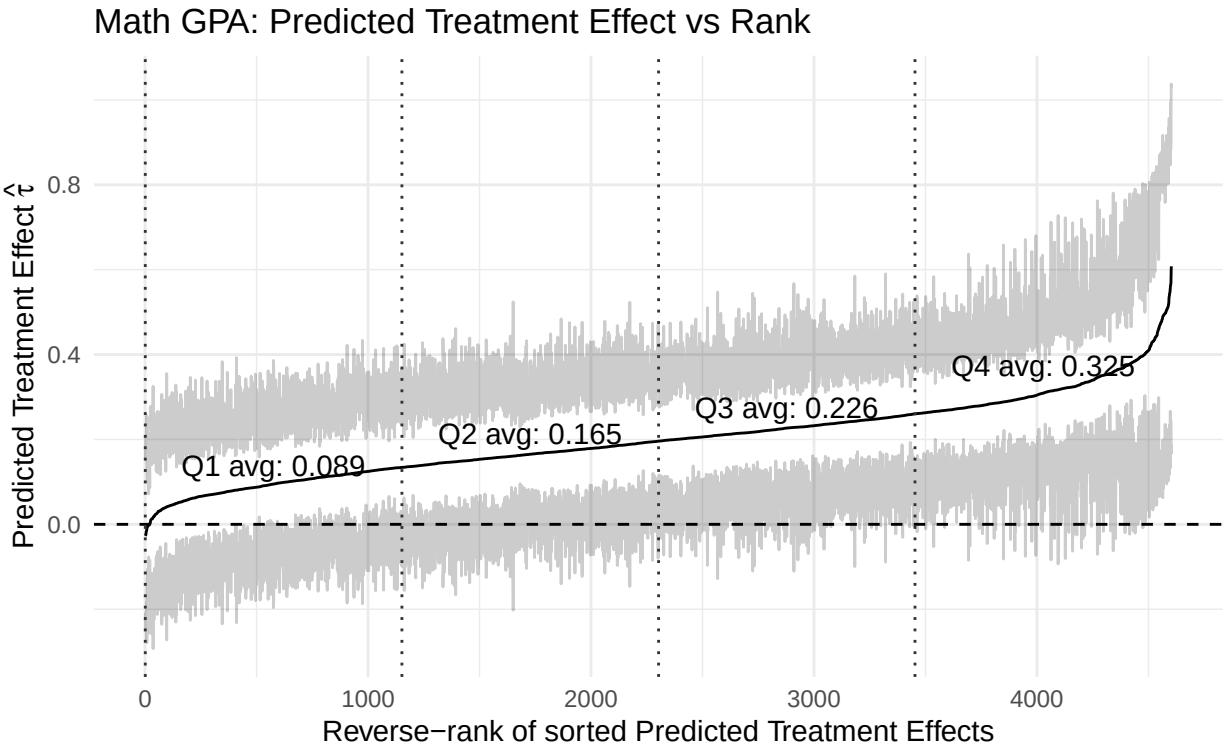
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.071, 0.141]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.056, 0.137]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.064, 0.214].

[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 $\hat{\tau}$	-0.022	0.093	0.143	0.211
N	1254.000	1208.000	1190.000	1193.000
Age	14.769	14.531	14.435	14.446
Female	0.133	0.163	0.165	0.164
Has IEP	0.342	0.181	0.115	0.055
Has Free/Reduced Lunch	0.896	0.902	0.929	0.882
Black	0.625	0.613	0.582	0.429
Hispanic	0.353	0.344	0.382	0.486
Other Race	0.022	0.043	0.035	0.085
In 9th Grade	0.781	0.802	0.799	0.733
In 10th Grade	0.215	0.198	0.201	0.267
Baseline GPA	1.813	2.372	2.484	2.624
Num. A's	4.291	6.426	8.176	10.085
Num. B's	5.072	8.201	8.955	8.789
Num. C's	8.224	9.734	9.483	7.936
Num. D's	4.325	3.454	3.233	2.701
Num. F's	4.705	2.325	1.824	1.522
Missing Baseline GPA/Grades	0.097	0.058	0.029	0.045
Days Absent	21.057	12.679	12.134	11.500
Missing Attendance Data	0.076	0.013	0.005	0.003
Math Test Score (Z)	-0.689	0.060	0.295	0.282
Reading Test Score (Z)	-0.364	0.112	0.299	0.139
Missing Math Test	0.257	0.084	0.014	0.013
Missing Reading Test	0.231	0.085	0.019	0.028
Out-of-School Suspensions	2.218	0.761	0.595	0.201
Disciplinary Incidents	1.437	0.612	0.479	0.183
Any Arrests at Baseline	0.180	0.127	0.126	0.080
Arrests: Violent Crime	0.099	0.079	0.049	0.050
Arrests: Property Crime	0.050	0.017	0.045	0.027
Arrests: Drug Crime	0.041	0.017	0.006	0.019
Math Score - Decile in Previous School	3.899	5.857	6.638	6.460
In Study 2	0.544	0.562	0.594	0.533
Participated in Year 1 of Study	0.306	0.278	0.271	0.268

Outcome: Math GPA



Our causal forest estimates an overall average treatment effect of 0.204 (0.027).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.02 and a 'differential forest prediction' (DFP) coefficient of 0.842, with a corresponding DFP p-value of 0.005.[2]

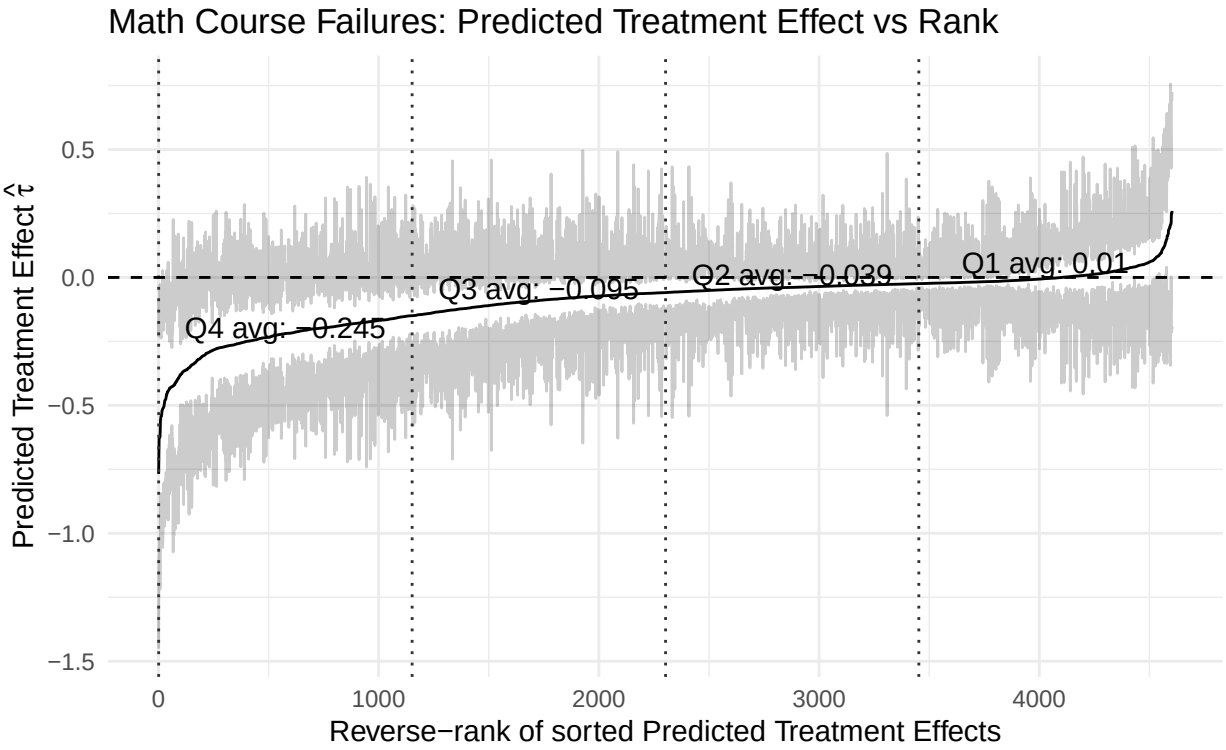
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [0.067, 0.324]. The 95% confidence interval for the difference between the above-median and below-median group is [0.105, 0.319]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [0.186, 0.495].

[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 $\hat{\tau}$	0.091	0.164	0.225	0.322
N	1797.000	1576.000	1819.000	1473.000
Age	14.389	14.653	14.837	14.902
Female	0.198	0.147	0.091	0.076
Has IEP	0.116	0.197	0.167	0.145
Has Free/Reduced Lunch	0.972	0.926	0.896	0.847
Black	0.642	0.591	0.595	0.463
Hispanic	0.334	0.366	0.350	0.436
Other Race	0.024	0.043	0.055	0.101
In 9th Grade	0.965	0.845	0.717	0.518
In 10th Grade	0.035	0.151	0.278	0.477
Baseline GPA	2.110	1.993	2.045	2.076
Num. A's	8.306	6.340	5.671	2.997
Num. B's	7.421	7.183	6.623	4.795
Num. C's	10.401	8.224	7.542	4.946
Num. D's	5.331	3.525	3.259	2.322
Num. F's	4.415	4.471	3.501	1.535
Missing Baseline GPA/Grades	0.000	0.001	0.005	0.239
Days Absent	23.216	25.612	21.151	16.386
Missing Attendance Data	0.000	0.000	0.000	0.126
Math Test Score (Z)	-0.098	-0.059	-0.077	0.124
Reading Test Score (Z)	0.093	-0.054	-0.097	0.101
Missing Math Test	0.117	0.126	0.158	0.238
Missing Reading Test	0.125	0.121	0.161	0.239
Out-of-School Suspensions	1.568	2.648	1.987	0.783
Disciplinary Incidents	0.932	1.684	1.385	0.700
Any Arrests at Baseline	0.206	0.264	0.231	0.213
Arrests: Violent Crime	0.120	0.273	0.132	0.129
Arrests: Property Crime	0.075	0.151	0.067	0.054
Arrests: Drug Crime	0.026	0.119	0.074	0.026
Math Score - Decile in Previous School	5.272	5.656	5.403	6.279
In Study 2	0.677	0.571	0.518	0.401
Participated in Year 1 of Study	0.304	0.303	0.235	0.253

Outcome: Math Course Failures



Our causal forest estimates an overall average treatment effect of -0.093 (0.019).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 1.032 and a 'differential forest prediction' (DFP) coefficient of 1.348, with a corresponding DFP p-value of 0.[3]

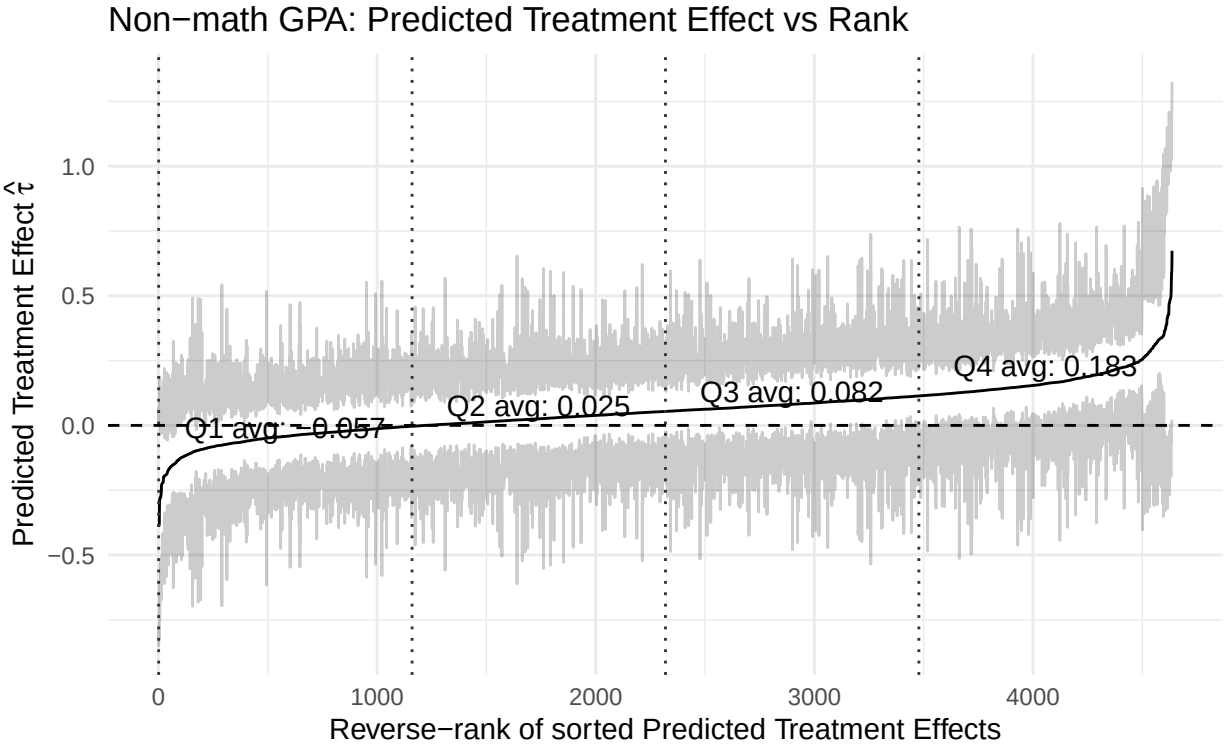
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.315, -0.108]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.256, -0.11]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.368, -0.137].

[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 $\hat{\tau}$	0.026	-0.039	-0.094	-0.271
N	1720.000	1414.000	1473.000	2058.000
Age	14.471	14.615	14.693	14.913
Female	0.157	0.161	0.168	0.057
Has IEP	0.121	0.156	0.213	0.142
Has Free/Reduced Lunch	0.914	0.871	0.908	0.944
Black	0.634	0.530	0.623	0.531
Hispanic	0.345	0.397	0.336	0.392
Other Race	0.022	0.073	0.041	0.078
In 9th Grade	0.869	0.719	0.775	0.720
In 10th Grade	0.131	0.279	0.222	0.274
Baseline GPA	2.363	2.569	2.068	1.384
Num. A's	8.125	8.770	5.554	2.473
Num. B's	7.738	9.021	6.741	3.776
Num. C's	9.140	8.151	8.623	6.174
Num. D's	3.201	2.563	3.609	4.878
Num. F's	2.474	1.966	3.134	5.810
Missing Baseline GPA/Grades	0.029	0.016	0.057	0.101
Days Absent	20.415	15.852	19.878	28.689
Missing Attendance Data	0.025	0.013	0.040	0.031
Math Test Score (Z)	0.300	0.111	-0.373	-0.211
Reading Test Score (Z)	0.197	0.163	-0.216	-0.121
Missing Math Test	0.138	0.074	0.123	0.254
Missing Reading Test	0.148	0.075	0.126	0.250
Out-of-School Suspensions	1.929	0.850	1.361	2.641
Disciplinary Incidents	1.036	0.677	1.012	1.823
Any Arrests at Baseline	0.208	0.146	0.179	0.338
Arrests: Violent Crime	0.112	0.140	0.177	0.207
Arrests: Property Crime	0.074	0.064	0.094	0.104
Arrests: Drug Crime	0.017	0.066	0.068	0.089
Math Score - Decile in Previous School	6.523	6.004	4.624	5.196
In Study 2	0.622	0.533	0.578	0.474
Participated in Year 1 of Study	0.252	0.286	0.327	0.245

Outcome: Non-math GPA



Our causal forest estimates an overall average treatment effect of 0.057 (0.022).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 0.97 and a 'differential forest prediction' (DFP) coefficient of 1.132, with a corresponding DFP p-value of 0.[4]

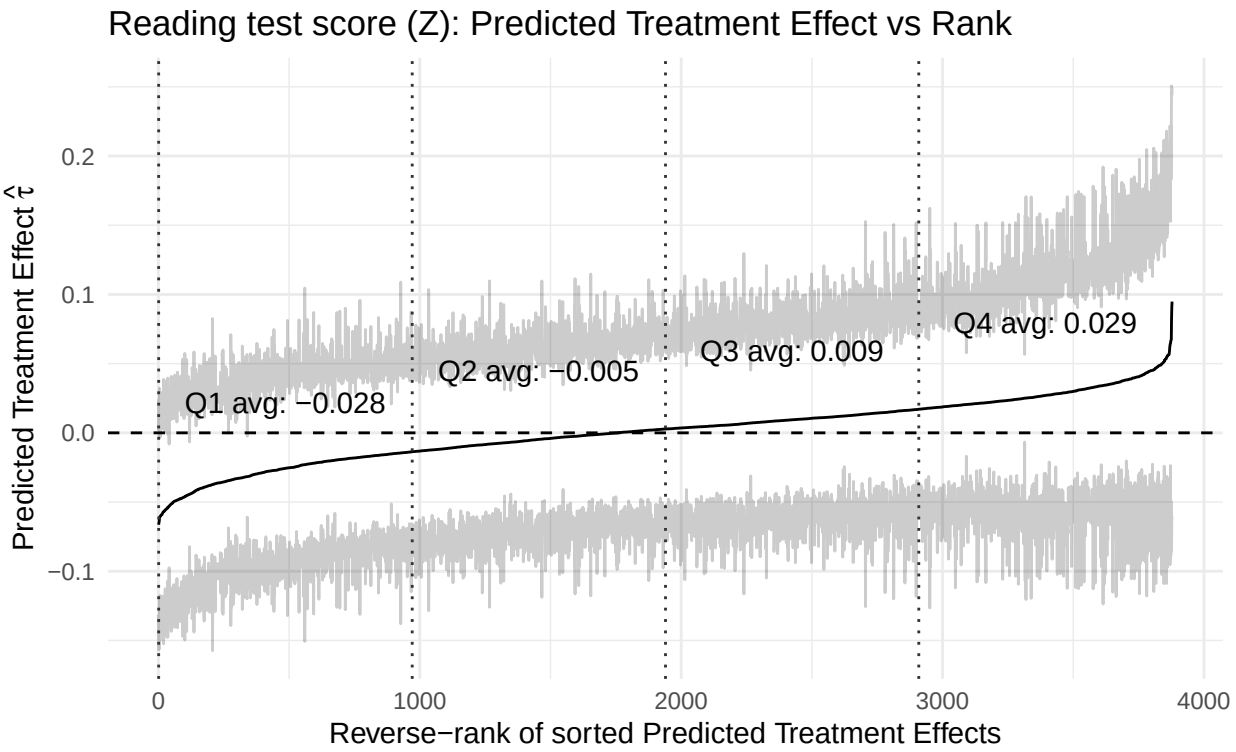
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [0.083, 0.284]. The 95% confidence interval for the difference between the above-median and below-median group is [0.007, 0.178]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [0.074, 0.319].

[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 $\hat{\tau}$	-0.080	0.025	0.083	0.199
N	1798.000	1488.000	1487.000	1969.000
Age	14.513	14.708	14.631	14.882
Female	0.122	0.159	0.163	0.086
Has IEP	0.138	0.143	0.196	0.146
Has Free/Reduced Lunch	0.930	0.907	0.902	0.912
Black	0.561	0.542	0.587	0.624
Hispanic	0.377	0.405	0.373	0.320
Other Race	0.062	0.054	0.040	0.056
In 9th Grade	0.814	0.726	0.763	0.771
In 10th Grade	0.186	0.270	0.233	0.222
Baseline GPA	2.230	2.062	2.106	1.794
Num. A's	6.661	6.005	6.584	4.622
Num. B's	8.160	6.033	6.410	5.455
Num. C's	9.547	7.565	7.779	6.524
Num. D's	3.831	3.695	3.695	3.414
Num. F's	2.568	3.647	3.778	4.389
Missing Baseline GPA/Grades	0.015	0.034	0.067	0.094
Days Absent	19.459	23.836	18.872	27.527
Missing Attendance Data	0.007	0.015	0.028	0.055
Math Test Score (Z)	0.114	-0.117	-0.182	-0.004
Reading Test Score (Z)	0.303	-0.118	-0.128	-0.062
Missing Math Test	0.121	0.150	0.126	0.235
Missing Reading Test	0.132	0.153	0.117	0.237
Out-of-School Suspensions	1.153	1.478	1.591	2.986
Disciplinary Incidents	0.796	0.994	1.078	1.897
Any Arrests at Baseline	0.186	0.175	0.196	0.347
Arrests: Violent Crime	0.065	0.147	0.139	0.281
Arrests: Property Crime	0.045	0.074	0.058	0.153
Arrests: Drug Crime	0.046	0.060	0.046	0.104
Math Score - Decile in Previous School	5.921	5.474	5.262	5.710
In Study 2	0.576	0.541	0.586	0.499
Participated in Year 1 of Study	0.300	0.307	0.298	0.196

Outcome: Reading test score (Z)



Our causal forest estimates an overall average treatment effect of 0 (0.026).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 0.993 and a 'differential forest prediction' (DFP) coefficient of -2.994, with a corresponding DFP p-value of 0.987.[5]

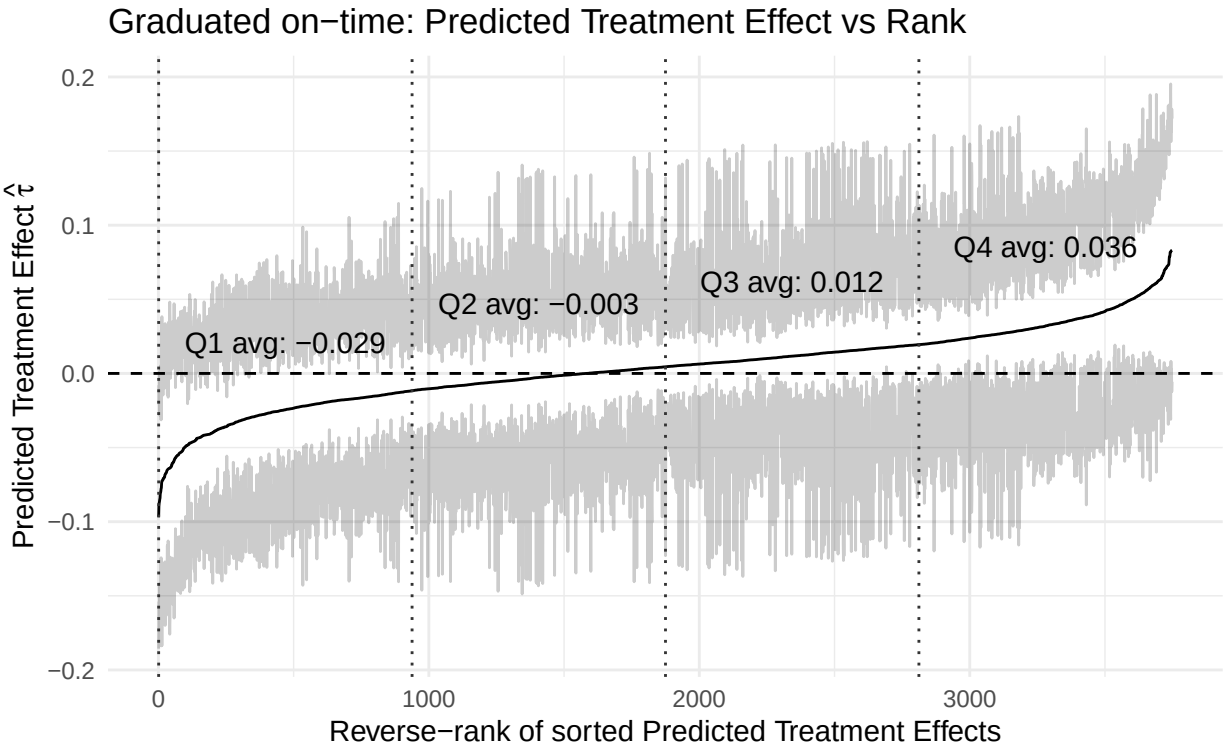
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.322, -0.057]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.183, 0.022]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.308, -0.006].

[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 $\hat{\tau}$	-0.028	-0.005	0.009	0.029
N	1234.000	1259.000	1253.000	1112.000
Age	14.589	14.504	14.658	14.434
Female	0.086	0.176	0.194	0.167
Has IEP	0.390	0.212	0.068	0.014
Has Free/Reduced Lunch	0.944	0.905	0.922	0.831
Black	0.506	0.585	0.635	0.531
Hispanic	0.483	0.383	0.298	0.394
Other Race	0.011	0.033	0.066	0.076
In 9th Grade	0.784	0.795	0.762	0.775
In 10th Grade	0.212	0.204	0.237	0.225
Baseline GPA	1.944	2.313	2.327	2.803
Num. A's	5.004	6.718	7.131	10.231
Num. B's	6.113	8.065	8.192	8.537
Num. C's	10.169	10.258	8.281	6.371
Num. D's	5.660	3.377	3.057	1.459
Num. F's	3.939	2.674	2.571	1.291
Missing Baseline GPA/Grades	0.016	0.027	0.053	0.143
Days Absent	18.841	14.639	14.388	8.864
Missing Attendance Data	0.000	0.000	0.003	0.105
Math Test Score (Z)	-0.627	-0.189	0.192	0.952
Reading Test Score (Z)	-0.463	-0.126	0.246	0.791
Missing Math Test	0.050	0.040	0.119	0.190
Missing Reading Test	0.053	0.040	0.107	0.195
Out-of-School Suspensions	1.366	0.921	0.957	0.377
Disciplinary Incidents	1.048	0.535	0.659	0.389
Any Arrests at Baseline	0.147	0.141	0.165	0.068
Arrests: Violent Crime	0.071	0.083	0.084	0.039
Arrests: Property Crime	0.060	0.037	0.032	0.024
Arrests: Drug Crime	0.019	0.017	0.040	0.006
Math Score - Decile in Previous School	3.990	5.110	6.463	8.369
In Study 2	0.535	0.577	0.566	0.560
Participated in Year 1 of Study	0.307	0.300	0.294	0.224

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 0.969 and a 'differential forest prediction' (DFP) coefficient of 0.334, with a corresponding DFP p-value of 0.265.[6]

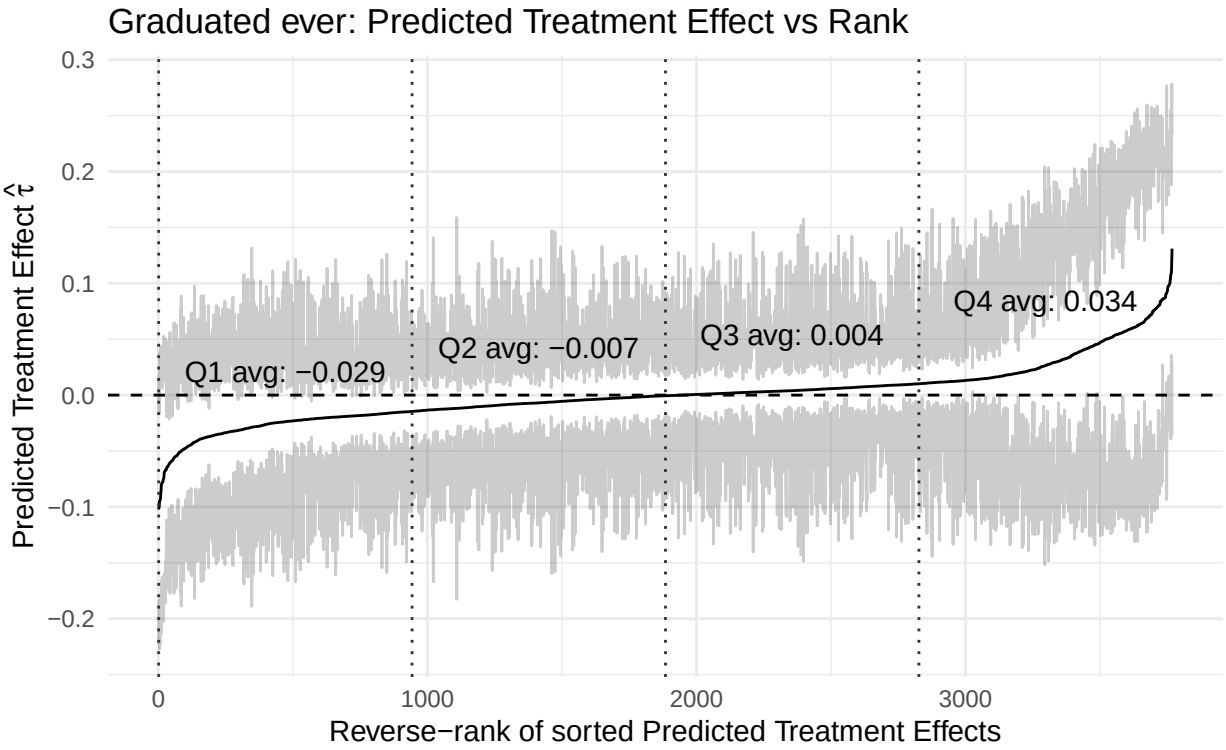
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.049, 0.06]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.059, 0.035]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.062, 0.077].

[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 $\hat{\tau}$	-0.033	-0.003	0.012	0.038
N	1193.000	1197.000	1181.000	1333.000
Age	14.784	14.511	14.474	14.508
Female	0.118	0.185	0.181	0.103
Has IEP	0.176	0.184	0.211	0.090
Has Free/Reduced Lunch	0.903	0.892	0.896	0.905
Black	0.573	0.567	0.604	0.550
Hispanic	0.392	0.365	0.332	0.399
Other Race	0.035	0.068	0.064	0.051
In 9th Grade	0.562	0.809	0.830	0.819
In 10th Grade	0.434	0.189	0.169	0.172
Baseline GPA	2.104	2.447	2.428	2.066
Num. A's	3.769	7.403	8.527	8.344
Num. B's	5.852	8.172	7.912	7.461
Num. C's	8.269	9.492	9.169	6.974
Num. D's	4.341	2.967	3.250	3.697
Num. F's	1.847	1.962	2.477	4.486
Missing Baseline GPA/Grades	0.028	0.072	0.070	0.026
Days Absent	18.075	13.610	11.822	22.133
Missing Attendance Data	0.003	0.024	0.048	0.014
Math Test Score (Z)	-0.137	-0.089	0.005	0.330
Reading Test Score (Z)	0.070	-0.084	0.047	0.401
Missing Math Test	0.142	0.091	0.088	0.114
Missing Reading Test	0.146	0.094	0.083	0.125
Out-of-School Suspensions	0.255	0.233	0.810	3.765
Disciplinary Incidents	0.295	0.339	0.590	2.561
Any Arrests at Baseline	0.122	0.112	0.126	0.293
Arrests: Violent Crime	0.059	0.054	0.061	0.214
Arrests: Property Crime	0.028	0.054	0.062	0.142
Arrests: Drug Crime	0.013	0.046	0.041	0.079
Math Score - Decile in Previous School	5.435	5.270	5.765	6.849
In Study 2	0.431	0.617	0.628	0.557
Participated in Year 1 of Study	0.257	0.287	0.287	0.194

Outcome: Graduated ever



Our causal forest estimates an overall average treatment effect of 0.001 (0.011).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 2.023 and a 'differential forest prediction' (DFP) coefficient of -0.106, with a corresponding DFP p-value of 0.577.[7]

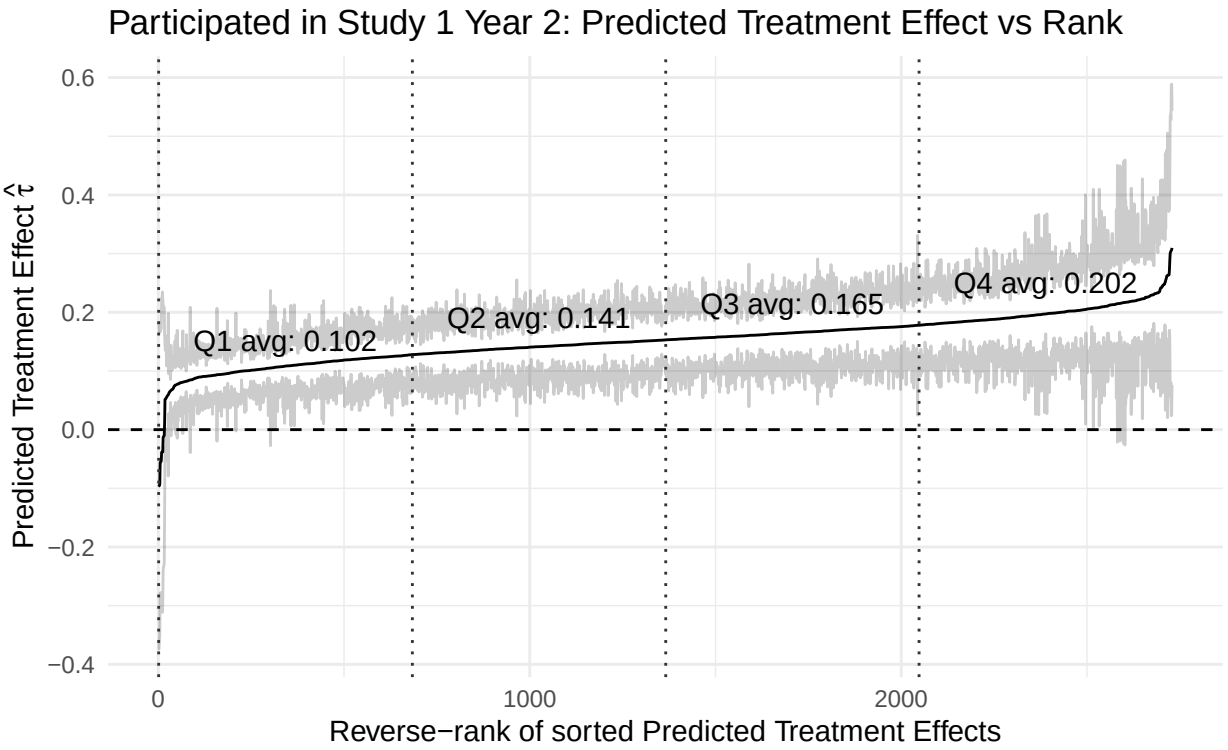
The 95% confidence interval for the difference in predicted treatment effect between the highest quartile group and the bottom 3 quartiles is [-0.053, 0.056]. The 95% confidence interval for the difference between the above-median and below-median group is [-0.049, 0.036]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [-0.065, 0.072].

[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 $\hat{\tau}$	-0.032	-0.007	0.005	0.039
N	1274.000	1176.000	1134.000	1340.000
Age	14.682	14.612	14.361	14.590
Female	0.135	0.167	0.160	0.124
Has IEP	0.194	0.195	0.116	0.145
Has Free/Reduced Lunch	0.911	0.886	0.915	0.885
Black	0.596	0.603	0.575	0.522
Hispanic	0.297	0.377	0.391	0.426
Other Race	0.107	0.020	0.034	0.052
In 9th Grade	0.676	0.700	0.839	0.803
In 10th Grade	0.317	0.295	0.155	0.183
Baseline GPA	2.000	2.448	2.691	1.927
Num. A's	3.432	6.282	11.652	7.247
Num. B's	5.677	7.912	9.849	6.301
Num. C's	9.864	9.218	8.036	6.710
Num. D's	5.073	2.816	2.430	3.779
Num. F's	2.785	2.070	1.964	4.008
Missing Baseline GPA/Grades	0.020	0.042	0.036	0.093
Days Absent	16.835	11.316	11.130	26.607
Missing Attendance Data	0.000	0.000	0.000	0.081
Math Test Score (Z)	-0.265	0.003	0.417	0.001
Reading Test Score (Z)	-0.217	-0.046	0.505	0.253
Missing Math Test	0.087	0.033	0.047	0.249
Missing Reading Test	0.091	0.036	0.049	0.252
Out-of-School Suspensions	0.513	0.423	0.332	4.016
Disciplinary Incidents	0.447	0.296	0.236	2.931
Any Arrests at Baseline	0.148	0.088	0.089	0.323
Arrests: Violent Crime	0.071	0.037	0.052	0.230
Arrests: Property Crime	0.030	0.060	0.040	0.155
Arrests: Drug Crime	0.031	0.030	0.014	0.106
Math Score - Decile in Previous School	4.995	5.844	6.871	5.731
In Study 2	0.520	0.581	0.608	0.535
Participated in Year 1 of Study	0.265	0.266	0.237	0.246

Outcome: Participated in Study 1 Year 2



Our causal forest estimates an overall average treatment effect of 0.151 (0.01).We test the calibration of the forest, and estimate a 'mean forest prediction' (MFP) coefficient of 0.99 and a 'differential forest prediction' (DFP) coefficient of 2.002, 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 [0.039, 0.136]. The 95% confidence interval for the difference between the above-median and below-median group is [0.066, 0.145]. The 95% confidence interval for the difference between the top quartile and bottom quartile is [0.078, 0.19].

[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 $\hat{\tau}$	0.095	0.141	0.165	0.203
N	821.000	744.000	742.000	802.000
Age	14.772	14.835	14.740	14.677
Female	0.001	0.000	0.001	0.001
Has IEP	0.097	0.109	0.162	0.284
Has Free/Reduced Lunch	0.864	0.878	0.908	0.901
Black	0.396	0.462	0.519	0.545
Hispanic	0.521	0.516	0.450	0.320
Other Race	0.083	0.022	0.031	0.135
In 9th Grade	0.622	0.519	0.598	0.686
In 10th Grade	0.348	0.461	0.399	0.310
Baseline GPA	2.125	1.875	2.117	2.114
Num. A's	8.161	3.870	4.566	4.635
Num. B's	6.630	4.914	6.388	6.288
Num. C's	3.408	5.595	8.606	10.446
Num. D's	1.481	3.579	4.069	4.718
Num. F's	4.108	4.099	3.173	2.190
Missing Baseline GPA/Grades	0.056	0.069	0.030	0.087
Days Absent	30.824	22.422	18.230	19.358
Missing Attendance Data	0.001	0.009	0.011	0.081
Math Test Score (Z)	0.825	0.132	-0.149	-0.852
Reading Test Score (Z)	0.858	0.069	-0.263	-0.452
Missing Math Test	0.210	0.149	0.097	0.181
Missing Reading Test	0.212	0.153	0.097	0.180
Out-of-School Suspensions	1.991	2.038	1.869	1.635
Disciplinary Incidents	1.289	1.512	1.317	1.142
Any Arrests at Baseline	0.291	0.281	0.158	0.161
Arrests: Violent Crime	0.207	0.218	0.096	0.167
Arrests: Property Crime	0.130	0.106	0.058	0.059
Arrests: Drug Crime	0.133	0.081	0.051	0.044
Math Score - Decile in Previous School	8.313	6.879	5.665	3.206
In Study 2	0.000	0.000	0.000	0.000
Participated in Year 1 of Study	0.134	0.198	0.241	0.234