Do civic returns to higher education differ across subpopulations? An analysis using propensity forests

Benjamin T. Skinner*   William R. Doyle
University of Florida   Vanderbilt University

24 June 2020

Abstract

In this paper we investigate how college participation may differentially influence civic behaviors among a recent cohort of young people. Using data from the High School Longitudinal Study of 2009, we consider two direct measures of civic behavior, voter registration and volunteering. We generate our estimates with propensity forests, a machine learning algorithm that can mitigate bias when using observational data and supports investigation of heterogeneous treatment effects. Overall, we find that college-goers are more likely register to vote and volunteer, though, conditional on volunteering at all, likely to volunteer fewer hours. We find limited evidence of heterogeneous returns across various groups, suggesting that civic returns to higher education are broadly shared by those who attend.

*Corresponding author: btskinner@coe.ufl.edu

Introduction

In the longstanding debate over the role higher education[1] plays in supporting the public good, prior research has shown connections between college participation and a range of civic behaviors (Bowman 2011; Colby et al. 2003; Evans, Marsicano, and Lennartz 2019; Hurtado 2007; Kezar, Chambers, and Burkhardt 2015; McMahon 2010). Relatively few studies provide similar estimates across subpopulations (Brand 2010; Perna 2005) or use quasi-experimental designs that mitigate bias in the estimates due to differential selection into college (Dee 2004; Doyle and Skinner 2017). In this study, we bridge this divide with a machine learning method that allows us to investigate differential civic returns across various student subpopulations.

State and federal funding for higher education is predicated on two basic tenets. First, without additional funding, too few students would make the optimal human capital investment decision, and the overall economy would suffer due to inefficient and inequitable use of human resources (Goldin and Katz 2009). Second, policymakers expect that there are positive externalities from higher education (McMahon 2010). Among the most important of the positive externalities are increased civic participation. While many of the pecuniary and non-pecuniary benefits of higher education accrue to the individual, increases in civic participation are expected to accrue to society at large (Flanagan and Levine 2010).

Without these civic benefits, federal and state policymakers expect that the country would be worse off. Accordingly, they believe the civic fabric of voting and volunteering represent important justifications for expansive investments in higher education (Marginson 2007). While the literature has noted many civic benefits of higher education (Baum, Ma, and Payea 2013; McMahon 2010), comparatively less research has been done to establish how civic participation due to college attendance might vary across groups. If civic participation varies across college-goers and the resulting benefits flow to some groups more than others—for example, if civic benefits accrue primarily to historically advantaged groups such as white, wealthy students—then higher education's role in

---

1. We use the terms *higher education*, *postsecondary education*, and *college* interchangeably throughout this paper. In all cases we broadly refer to any education that occurs after high school graduation.

supporting civic participation may be overstated due to the inequitable distribution of the benefits.

Drawing on a sample of students from the High School Longitudinal Study of 2009 (HSLS09), we examine the relationship between college participation and two key civic behaviors—registering to vote and volunteering—using propensity forests (Wager and Athey 2018). Akin to propensity score matching (PSM) models, propensity forests effectively compare college enrollees and non-enrollees of similar enrollment propensity in order to estimate the effect of college participation on the outcomes of interest. By taking into account students' relative propensity for enrollment when making comparisons, we can mitigate selection bias in our estimates that might overstate the impact of college participation on civic engagement. Though propensity forests similarly rely on observable characteristics to produce their estimates, they have benefits over traditional parametric PSM procedures that we argue make them well suited to our analysis. In particular, propensity forests more readily support investigation of heterogeneous treatment effects among subpopulations.

We find generally positive links between college participation and both voter registration and the choice to volunteer. Conversely, among those who volunteer, we find that college-goers spend less time volunteering. Though our estimates suggest heterogeneity in how students respond to college participation, with some subgroups driving the overall results, we do not generally find strong evidence for significant differences by gender, race/ethnicity, poverty status, or propensities for enrollment. We situate these findings in the broader literature on the civic benefits of higher education. Our theoretical framework identifies possible reasons why civic benefits may not be realized similarly for all groups that attend higher education. The contribution of this paper lies in its use of a novel estimation technique to recover less-biased estimate of the impact of post-secondary education on civic behavior, with an emphasis on heterogeneity of estimates across different groups.

Background

Increasing students' civic awareness has long been one of the key goals of higher education. The bulk of the available evidence shows that college-educated students appear more likely to be civically engaged (Mayhew et al. 2016). We divide our review of the available evidence into three broad areas. First, we examine studies that provide a broad overview of the association between higher education and civic behaviors. Second, we look at studies that provide observational evidence regarding the change in civic behaviors that could be due to attendance in higher education. Last, we examine the small subset of studies that attempt to recover causal evidence of the impact of higher education on civic behaviors. In general, it is difficult to recover causal estimates because of the impracticality (and general inappropriateness) of conducting randomized studies on something as important as attendance in higher education.

That higher education should increase civic participation has been widely assumed. Persson (2015) provides an overview of studies examining the role of higher education in increasing civic attainment. Persson divides studies based on three underlying theories about the role of higher education in increasing civic participation. The first type of study he terms the "absolute education model." These studies assume that any observed link between postsecondary attendance and civic participation is due to the education itself, and not any other factors. He writes, "most of these studies draw on cross-sectional data and the causal mechanisms are seldom directly tested," Persson (2015, p. 691).

Persson terms the next possible underlying mechanism "the pre-adult socialization model." In this model, the same elements that might lead to more education would also lead to more civic participation, but without any necessary connection between postsecondary education and civic participation. As Persson points out, this kind of selection bias would mean that most model estimates that show a link between education and civic outcomes are misspecified and, if properly specified, would show little or no relationship between the two.

Persson calls his last underlying mechanism "the relative education model." In this model, we observe a link between education and civic participation because more educated people are more

4

likely to obtain central positions within the society and thus be in politically influential networks. The ability to influence political outcomes directly influences the individual's intention to participate. This could account for the overall trend of decreased civic participation over time, even as educational attainment has increased (Hillygus 2005). In other words, there may be more educated people overall, but as positions of power have stayed stable, the relative proportion of educated people in such positions has decreased.

Across studies in his review, Persson finds that while most studies maintain the absolute education model as the likely explanation, the evidence does not necessarily support this position. Instead, there are contradictions in the literature wherein the combined evidence from different studies, each with a strong research design, inconsistently argues for both a causal relationship between education and civic outcomes and a "proxy" relationship between education and civic outcomes. We seek to understand this issue better by pursuing the possibility that the effects may differ by subgroup.

While many studies have described the overall relationship between civic participation and postsecondary education, fewer have focused on heterogeneity among affected individuals. An important exception to this is the work of Hillygus. Hillygus (2005) uses longitudinal data from the Baccalaureate and Beyond survey to examine which characteristics of bachelor's degree holders predict civic participation in the years following graduation from college. Hillygus finds that respondents who took more social science or humanities courses were more likely to participate in politics and to vote, while individuals who took more business and science courses were less likely to be engaged in political behavior. Hillygus also finds that college-educated African-Americans were significantly more likely to vote than college-educated white or Asian individuals, while also finding no observable difference between men and women college graduates with respect to voting. While this study does not provide causal estimates, it does provide an important indicator that not all individuals may respond to higher education in the same way, in no small part because not all individuals participate in college in the same way.

If Persson's pre-adult socialization model is correct, then postsecondary education likely has

little effect on civic behaviors. This makes finding experimental or quasi-experimental evidence quite important. In particular, research designs that reduce bias due to self-selection can help establish whether there exists a causal relationship between civic participation and education. In one quasi-experimental study, Dee (2004) uses two identification strategies, one based on the geographic availability of colleges and another based on changes in child labor laws, to explore the effect of college participation on multiple civic outcomes. Among other findings, Dee shows that additional years of education are positively associated with voting. In a follow-up using new instruments and sample, Doyle and Skinner (2017) use multiple measures of geographic variability to identify the relationship between education and civic outcomes and find that an additional year of postsecondary education leads to an increase of 7.7 percentage points in the probability of voting.

Brand (2010) examines selection bias in the context of the relationship between postsecondary education and civic outcomes in another quasi-experimental study. She identifies two types of selection bias—pre-treatment heterogeneity and treatment effect heterogeneity. To explore treatment effect heterogeneity, Brand examines variation in the impact of postsecondary education on civic outcomes by the propensity to complete college. Using data from the National Longitudinal Survey of Youth from 1979 Brand stratifies the sample on the basis of propensity to complete college and estimates the relationship between postsecondary attendance and volunteer work. Similar to Brand and Xie (2010), who studied the economic returns to college attendance based on the propensity of attendance, Brand (2010) finds that the impact of postsecondary education on civic outcomes is largest for those with the lowest propensity to complete higher education.

Our contribution with this study is two-fold. First, we update the literature on the influence that college participation has on civic engagement. Most of the cited studies use data from before 2005, limiting their relevance for current populations. Second, we demonstrate the usefulness of a new methodology that not only allows us to mitigate bias in our estimates but also enables us to focus on heterogeneity of response among different student subpopulations. We add to the contributions of Brand (2010) by examining whether treatment heterogeneity differs by race, income, identified gender, or propensity for enrollment.

Theory

Doyle and Skinner (2017) lay out a brief theory of civic participation, based on the work of Gerber, Green, and Larimer (2008). They posit that the link between postsecondary education and general civic behaviors could work by either increasing the intrinsic rewards or the extrinsic rewards for civic participation. Their study does not address heterogeneity in response in any detail.

Intrinsic rewards for civic behavior concern the satisfaction that an individual might gain from being more civically engaged. Doyle and Skinner (2017) suggest that students gaining more satisfaction from activities such as voting, volunteering, or charitable giving as they learn more about how civic participation benefits both them as individuals and society as a whole is one example of how postsecondary education might increase intrinsic rewards. Extrinsic rewards include the recognition that might come to an individual as a result of civic participation, including the approbation of peers. Extrinsic rewards for civic participation may be increased through participation in postsecondary education as an individual's peer group changes to include more and more people who value and provide approval for civic behaviors.

In our study, we examine heterogeneity in the relationship between postsecondary education and civic outcomes in three broad areas: race/ethnicity, gender, and income. We consider each in turn below.

Baker and Blissett (2018) provide a framework for examining the civic experience of racially minoritized students on college campuses. They suggest that the current policy discourse around racial or ethnic diversity on campuses focuses on descriptive analysis: what proportion of students on a given campus identify as being from a particular racial or ethnic group. More important than the simple proportion of students, they suggest, are the climates that student experience while on campus. They identify two broad areas that may affect the experience of racially minoritized students on a given campus. The first is the extent to which campus leaders are devoted to what Garces and Jayakumar (2014) define as "dynamic diversity." Dynamic diversity can be thought of as the extent to which the institution is committed to interrogating its own role in perpetuating racism and to committing to the success of a broadly representative group of students, including a commitment

to "diversity within diversity" — enrolling students who may share a similar racial background but differ in a number of other ways. The second is the extent to which racially minoritized students experience microaggressions on campus, defined by Pierce et al. (1977) as "subtle, stunning often automatic…exchanges which are 'put downs' of blacks by offenders," (p. 66). More recent work has expanded on Pierce et al. by demonstrating the degree to which microaggressions occur against other racially minoritized groups (Harwood et al. 2012).

Racially minoritized students' experience of campus climates and microaggressions may affect the degree to which postsecondary education results in higher civic participation among them. To the extent that campuses offer an experience of dynamic diversity and minimize the microaggressions racially minoritized students experience, postsecondary education may increase both the intrinsic and extrinsic rewards of civic participation for these students. On the other hand, if racially minoritized students experience an isolating campus climate that features frequent microaggressions, then both the intrinsic and extrinsic rewards of civic participation might be lowered as students may favor withdrawal from a climate that treats them with hostility. Baker and Blissett (2018) also suggest that more hostile campus environments may eventually *increase* racially minoritized students' civic participation, particularly in the realm of activism and protest, once these students reach a "tipping point" beyond which they are unwilling to abide by the campus' accommodation of microaggressive behaviors.

A similar dynamic may be at work for women on college campuses. Though women have experienced higher levels of participation in higher education over time and are now broadly represented on most college campuses, it remains the case that women are underrepresented in high-paying majors and experience wage discrimination in the labor market (Blau and Kahn 2017; Carrell, Page, and West 2010; Goldin, Katz, and Kuziemko 2006; Speer 2017). Furthermore, while women have increasingly participated in politics in the United States, they remain underrepresented in Congress and state legislatures (Cascio and Shenhav 2020; National Conference of State Legislatures 2019). As with racially minoritized students, the role higher education plays in civic participation among women is not certain. Some women in higher education may be encouraged

by trends toward greater parity and thereby experience increases in both intrinsic and extrinsic rewards to civic participation; others may be discouraged by the lack of progress in representation and favor withdrawal from civic participation; and others still may find that their higher education experience provides a negative example that spurs greater civic engagement.

For low-income students, the dynamics of college participation and its impact on civic participation also hinge on the extent to which their experience of postsecondary education affirms their ability to participate fully in the civic life of society. For low-income students in particular, the cost-benefit calculation for participating in civic activities may be particularly pertinent. Because substituting the time spent voting or volunteering for paid labor imposes a substantially higher proportional penalty on low-income students' budgets than those of their peers, the intrinsic and extrinsic rewards for civic participation may need to be higher for low-income students than their more financially secure peers (Panagopoulos 2013).

Stemming from this last point, we note another possible mechanism through which college participation might change civic behaviors: opportunity. The opportunity to engage in civic activities as a result of attending postsecondary education could work in either direction. Within a college environment, students may have more opportunities to civically engage, be it by joining organizations that volunteer as part of their mission or being proximate to student-led initiatives to engage other students (*e.g.*, voter registration drives) (Holbein and Hillygus 2020). On the flip side, students who otherwise might volunteer or be politically active may have less time to do so due the combined demands of their classes and existing commitments to work and family (Oesterle, Johnson, and Mortimer 2004). Students attending out-of-state schools may find it more difficult to register to vote (Holbein and Hillygus 2020). If opportunities for civic participation vary by institution sector, students more likely to attend residential four-year universities are likely to realize different opportunities for civic participation than those who attend community colleges or for-profit institutions. If some student groups generally have different college experiences and opportunities than their peers, then the impact of college attendance on civic participation may differ. All of these factors may play into the well-known gap between intention and turnout behavior

9

(Holbein and Hillygus 2020).

Though our estimation strategy can identify which groups show different levels of impact, it cannot disentangle which of the theorized causes account for the observed effects. We discuss the details of our methodological approach in the next section.

Method

With observational data we cannot simply compare average differences in civic behaviors between those who attended college and those who did not. People have different likelihoods of college participation and, as we discuss in the prior section, these differential likelihoods are generally co-indicated with different likelihoods of civic behaviors. Because factors that are positively associated with college enrollment—*e.g.*, income—are also positively associated with civic behaviors like registering to vote (Verba and Nie 1972; Wolfinger and Rosenstone 1980), we would expect that unadjusted estimates of changes to civic behavior based on college participation would overstate the influence of college.

In the absence of random assignment to attend college, we can reduce bias in our estimates by comparing outcomes only between young persons who are alike in their odds of going to college. One approach is to use observable characteristics associated with college participation to separate the sample into more homogeneous subgroups, within which comparisons are made. To make this process more tractable, we can use a propensity score matching procedure (Dehejia and Wahba 2002; Rosenbaum and Rubin 1983) to create a single value that reflects a young person's propensity for college participation and use it to make better matches when comparing outcomes.

Propensity score matching has received much attention in the economics and program evaluation literature (*see:* Athey and Imbens 2017; Imbens and Wooldridge 2009). Much discussed in this literature are two primary difficulties faced by applied researchers: how to specify the propensity score model and how match observations based the resulting score. More specifically, propensity score matching requires the applied researcher to decide, at the very least, how to define the

propensity score model's functional form, which observations to remove (if any) from the analysis sample, whether comparisons should only be made between observations with the most similar scores or if multiple comparisons within a range are more appropriate, and whether observations should only be used in one comparison or can be used multiple times. While a number of best practices have coalesced (Caliendo and Kopeinig 2008; Imbens 2015), applied researchers using propensity score matching face a number of decisions that may affect the validity of their results.

In this paper, we use a machine learning-based method that mitigates some of these difficulties by using the data themselves—rather than the researcher—to group and compare observations: propensity forests. Propensity forests (Wager and Athey 2018) are a special subset of the random forest algorithm (Breiman 2001; Hastie, Tibshirani, and Friedman 2009). A propensity forest is made up of a large number of decision trees in which the treatment, W (college participation), is the outcome.[2] Each decision tree is built by sorting observations into ever smaller groups, using recursive splits in covariates or predictors, X, to decide how to sort them. At each step, it does this mechanically by finding the best single criterion by which to separate the observations. The goal of these splits is to create nodes or leaves, L, in which within-leaf observations are increasingly similar to one another and, by proxy, dissimilar to those in other leafs. Because the decision tree classifies observations based on their relation to the treatment (participating in college), each decision tree effectively groups observations by their propensity for receiving the treatment.

As an illustrative example, a very simple single decision tree in our study may start with four potential predictors of college enrollment: the student's age, educational aspirations, region, and mother's education level. Based on these predictors and the observations chosen to build this tree, the algorithm may find that the first best split occurs by putting those who plan to earn a Bachelor's degree or above in one group and those who do not in another. Among those who plan to earn at least a Bachelor's degree, the best separation occurs between those who are less than 25 years old and those 25 and older. Among those who do not plan to earn a Bachelor's degree, further splits

---

2. The language of propensity forests, like parametric propensity score matching methods, uses the language of randomized control trial: *treatment, control group, average treatment effect, etc*. We use this language for consistency with existing descriptions of the method and its procedures, but note the conditions for causality later in the section.

are made by placing those from the South in one group and dividing those from other regions in the country by their mother's education level: greater than high school in one group and those below in the other. At this point, the final leafs are too small to split any further. Visually, the decision tree would look like this:
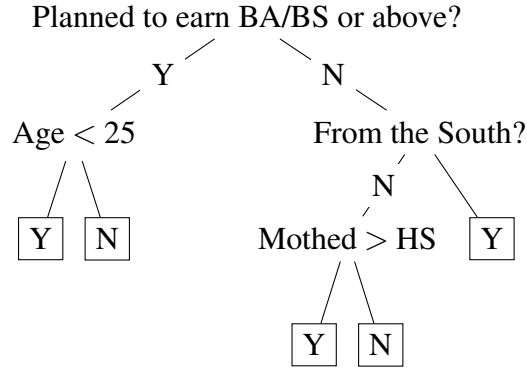
Once a tree is built, a treatment effect, $\hat{\tau}(x)$, for the primary outcome of interest (*e.g.*, registering to vote), Y, is then computed for each leaf using,

$$\hat{\tau}(x) = \frac{1}{|i : W_i = 1; X_i \in L|} \sum_{\{i:W_i=1;X_i\in L\}}^{Y_i} - \frac{1}{|i : W_i = 0; X_i \in L|} \sum_{\{i:W_i=0;X_i\in L\}}^{Y_i}, \tag{1}$$

which is simply the difference between the primary outcome of interest for the treatment and control groups within the leaf, weighted by the proportion of observations that were treated. For example, if a leaf ended with 20 observations, 8 of which attended college and 12 of which did not, and the respective number registered to vote were 5 and 6, then the treatment effect would be 1/8 * 5 - 1/12 * 6 = 0.125 or 12.5 percentage point increase due to college attendance for observations assigned to that leaf. Other leaves on the tree would receive different treatment effect estimates according to their cross tabulations of treatment exposure and primary outcome.

To introduce variation across individual trees that comprise the forest and thus prevent overfitting, a random subset of observations, I, and predictors are chosen to build each tree. To further prevent overfitting, only held out observations (those not used to build the tree) are placed in their appropriate leaf based on the rules of the decision tree and assigned the treatment effect, $\hat{\tau}(x)$, of the leaf. Wager and Athey (2018) show that as the number of trees in the forest grows, the distribu-

tion of the treatment effect assigned to each observation becomes normal and unbiased, allowing for statistical inference. We use R (R Core Team 2020) with the **grf** package (Tibshirani, Athey, and Wager 2020) to fit 20,000 separate trees for each outcome, which provides a large number of observation-level treatment estimates, $\hat{\tau}(x)_i$, that can be averaged and assigned to every student in the sample.[3]

Individualized treatment estimates can be aggregated to produce an average treatment effect (ATE) estimate. With propensity forests, ATEs can be aggregated to the full sample or among specified subgroups. In this paper, we use the doubly robust augmented inverse propensity weighting (AIPW) method to compute all ATE estimates. This method is known as "doubly robust" because in the context of a propensity score matching design, it remains consistent as long as either the propensity score model (in our case, the forest fit) or the outcome regression model (in our case, the estimation of $\hat{\tau}(x)$) is correctly specified (Glynn and Quinn 2010; Robins, Rotnitzky, and Zhao 1994). Furthermore, because some of the subgroups we explore in our heterogeneity analyses are small and contain observations with either very low or very high propensities for college enrollment, we present overlap-weighted ATEs throughout (Li, Morgan, and Zaslavsky 2018). As the name implies, overlap-weighted ATEs give more weight to "the units whose combination of characteristics could appear with substantial probability in either treatment group," (p. 394). By combining the AIPW method with overlap-weighted ATEs, we present the most conservative ATE estimates.

The benefit of propensity forests over econometric propensity score matching procedures is two fold. Unlike with econometric propensity score matching procedures, we do not have to construct a propensity score model in which we must pre-specify a functional form (Dehejia and Wahba 2002). Relatedly, we do not have pre-specify the multitude of interactions between predictors (race, gender, family income, test score, *etc.*) that quickly become too numerous to fit in a standard parametric regression, but which are nonetheless required for analyses of heteroge-

---

3. By default, the **grf** package function (`causal_forest()`) fits 2,000 trees. However, Tibshirani, Athey, and Wager (2020) note that when the accuracy of confidence intervals is important (as it is in our case), more trees may need to be fit. We choose to fit 20,000 as a compromise between an increased number of trees (10x more) that is not overly taxing to compute.

neous treatment effects. This is due to the fact that with a propensity forest, each observation is assigned its own treatment effect estimate. To compute treatment effects for a subgroup, we need only to compute estimated treatment effect estimates for observations in that subgroup. By design, the propensity forest algorithm effectively accounts for functional form differences and high dimensional interactions as it combines observations with similar treatment propensities. These two properties should mitigate bias that might otherwise accrue due a poorly specified propensity model or post hoc matching procedure.

For our estimates to have a causal interpretation, however, one must assume that the outcome is independent of treatment assignment conditional on all of the predictors we use in the model— the conditional independence assumption (Angrist 1997). This is a major assumption shared with parametric PSM models and one that we are unlikely to meet. Despite the fact that the propensity forest algorithm allows us to use a large number of predictors and gives a method for winnowing them to only those that are the most important, it cannot account for unobserved predictors of enrollment. Such that students in our sample who enrolled are systematically different from those who did not enroll along omitted dimensions, the grouping of observations and resulting estimation of $\hat{\tau}(x)$ may retain some bias. Despite this threat, we argue that propensity forests remain a useful tool for our purpose, which is to use observational data to estimate the impact of college participation on various civic outcomes.

Data

Data for our analyses come from the High School Longitudinal Study of 2009 (HSLS09), a recent nationally-representative survey of high school students that first interviewed a cohort of over 20,000 ninth graders in the fall of 2009 (Duprey et al. 2018). Subsequent data collections in 2011, 2013, and 2016 followed students as they completed high school, enrolled in college, and, in some cases, entered the workforce. The HSLS09 provides researchers with a rich set of covariates that, in addition to student demographics, include both measures of academic ability and responses

from students about their self-perception, beliefs, and expectations for the future. Survey items from the base year administration of the survey also include responses from interviews with parents/guardians, teachers, and school administrators.

We focus on two primary civic outcomes: being registered to vote and volunteering. Responses for both come from the second follow-up survey conducted in 2016. Due to the timing of the interviews, which occurred from March 2016 to January 2017, not all respondents had had the opportunity to vote. Therefore, respondents were asked, "Were you registered to vote in February 2016?", and their responses coded to one of three conditions: yes, no, and ineligible to vote.[4] We removed those who said they were ineligible to vote leaving a binary indicator which equals one for those who were registered to vote and zero those who said they were not registered.

To account for student volunteering behavior, survey administrators asked the following question:

> Now we have some questions about your community involvement. In calendar year 2015, about how many hours per month (on average) did you volunteer or perform community service that was not required by a college, trade school, an employer, or the criminal justice system?[5]

Respondents were given the option of providing a monthly average of hours they volunteered or, if they did not volunteer at all, a value of zero. Because these data are zero-inflated, that is, a large percentage of respondents did not volunteer at all in 2015, we operationalize volunteering in two ways. First, we created a binary indicator that equaled one if a student reported any hours of volunteering in 2015 and zero otherwise. Second, among those students with any positive volunteer hours in 2015 (that is, hours > 0), we created a variable that is the natural log transformation of their volunteer hours to account for the right skew of the distribution.

To construct our indicator of college enrollment, we use a question from the fourth wave in which respondents were asked about their postsecondary class-taking activities between the time

---

4. HSLS09 variable name: *S4REGVOTE*
5. HSLS09 variable name: *S4HRSVOLUNTR*

they earned their high school credential (including GED) and February 2016.[6] The question was worded broadly to include "[c]olleges and trade schools where you were just taking classes," and "[o]nline only colleges and trade schools," (Duprey et al. 2018). Respondents could answer with a yes (1) or no (2). Because students were asked about volunteering they might have done in 2015, we remove students who started college after 2015 so that the timing of first attendance is prior to the outcome. This means that the treatment of college participation is strictly whether a student enrolled college in the first year and a half after on-time high school graduation.

Because the choice of whether to attend college is not random, we rely on a host of covariates that may be predictive of college enrollment and that will help us compare students with similar enrollment propensities. All predictors come from survey items collected in the base year (2009) or first follow-up survey (2011), when respondents were in 9th and 11th grade, respectively. Limiting ourselves to items from these periods guarantees that all predictors of college enrollment are measured before first postsecondary enrollment for all students.

We chose predictors in a two-step process. First, we selected all survey items from the first two data collection periods that the college access and choice literature says should correlate with post-secondary enrollment (Hoxby 2007; Long 2004; Perna 2006; Skinner 2019) Because the random forest algorithm at the core of our estimation procedure groups students based on their propensity for taking college courses, we made no *ex ante* decision about which items might be more or less informative. To the contrary, we wanted to include a large number of items so that the algorithm—rather than we—would decide their relative predictive power. In practice this means that we included many more predictors than would have been practicable with standard regression-based propensity score models.

Broadly, the over 100 items we chose in the first step fell into the following bins: *(1) characteristics of the student*: gender, race/ethnicity, date of birth, socioeconomic and federal poverty status; *(2) characteristics of the student's household and parents*: household size and income;

---

6. HSLS09 variable name: *X4EVRATNDCLG*, which is the imputed version of *S4EVRATNDCLG*. Though some students took college classes during high school as dual-enrolled students, we define college enrollment as taking any postsecondary courses after high school graduation.

parental race/ethnicity, education levels, 2-digit occupation codes, and language used to complete questionnaire; *(3) characteristics of the student's school*: geographic locale (urbanicity), region, and control; percents of 11th graders repeating 11th grade and returning to school *(4) characteristics of the student's teachers*: race/ethnicity and teaching certification type; *(5) student's high school experience*: indicators for individualized education plan, enrollment and dropout status in 11th grade, and mathematics assessment accommodations; number of high schools attended; *(6) student test scores and identification*: mathematics score (9th and 11th grades); scales of math and science identification and student engagement; *(7) student and parental beliefs*: expectations for educational attainment and future occupation; salary expectations by degree attainment. The full list of survey items and their descriptions are listed in the appendix table A.3.

For the second step of the predictor selection process, we fit a propensity forest for each outcome using all potential predictors. We included the few survey items that took on continuous values as is. Test scores and constructed scales of student identification (*e.g.* "I/others see me as a math person") and engagement fall into this group. Most survey items, however, took on discrete values. Those that had an obvious ordering (like family income), we left as is and treated as continuous. For those without clear ordering, we converted these into separate 0/1 indicator variables for each unique category.[7] We used these vectors of indicator variables in place of the original categorical predictors in the propensity forests we fit.[8] Appendix table A.4 adds an asterisks to all discrete predictors that we recoded in this manner.

As is often the case with large-scale longitudinal surveys, not all observations have complete data across all data collection periods. For each outcome, we first limit the sample to students for whom we observe the outcome as well as college enrollment status in the 2016 update. Because we include missing values as a distinct category for all discrete predictors, we do not need to impute values or drop observations due to missingness among these predictors. For continuous predictors with missing values, the random forest algorithm will choose the best of three potential splits—

---

7. Also referred to as *dummy variables* and *one-hot encodings*.

8. To reduce the number of parental 2-digit occupational codes (23 groups), we follow the guidance of the Bureau of Labor Statistics to further aggregate the codes into 6 high-level groups (*Standard Occupational Classification and Coding Structure* 2018).

in a split into group *A* and *B* based on non-missing values: (1) all missing values are included with group *A*; (2) all missing values are included with group *B*; (3) all missing values are placed into group *A* and all non-missing values are placed into group *B*. While this procedure will work with any degree of missingness, we drop continuous predictors with greater than 50% missingness before fitting the forest model. This process leaves us with a total of 300 predictors for each outcome.

After growing each forest with all predictors, we rank the predictors by their importance in improving the fit across the trees in the forest. In addition to the categories that we explore in our heterogeneity analyses (indicators for race/ethnicity, gender, and 185% poverty level), we next fit a second set of propensity forests that each used only those predictors in the 80th percentile or above. Because random forests tend to perform better as the proportion of relevant variables increases (Hastie, Tibshirani, and Friedman 2009), we remove those predictors that do not contribute much (or any) information across fitted trees.[9] All results presented in this paper come from these second models. The subset of most-important predictors selected by our two-step procedure for each outcome are indicated in appendix table A.4.

For all models, we rely solely on the publicly available HSLS09 data elements. We do this in the interests of reproducible science and so that our code may serve as a fully accessible implementation example for others who wish to use propensity forests or other forest-based inferential methods in their own research.[10] In addition, we do not use survey weights in our models. Though HSLS09 offers a number of cross-sectional and longitudinal weights and the **grf** R packages supports their use, our estimation strategy and variable selection process made the choice of an appropriate weight unclear.[11] Accordingly, we cannot claim that our findings are externally valid to the

---

9. We fit multiple second sets of propensity forests for each outcome using predictors with positive variable importance values (*e.g.*, excluding those with zero values) as well as those at the 50th, 80th, 90th, and 95th percentiles. Fit statistics were similar across these models so we selected the 80th percentile as a balance between predictor inclusion and parsimony.

10. All replication files can be found at: https://github.com/btskinner/civic_returns_pf_rep

11. For example, base year weights do not take into account missing values for our key treatment and outcomes, which we take from later waves. Longitudinal weights that reweight observations as they appear across survey waves do not necessarily account for how we or the **grf** R package handle missing predictor values. In either case, weights likely augment the relative importance of some observations while diminishing that of others in ways that are unclear, unsupported, and untestable.

national population of students. Instead our results speak to the students in our sample: a group of young people from across the United States who entered high school in the fall of 2009.[12]

Results

We begin by describing average differences in civic participation between those who enrolled in college and those who did not. The left column facets of figure 1 focus on our three primary civic outcomes: voter registration, volunteering, and time spent volunteering among those who volunteered. Beginning with the upper left facet, 46.7% of non-enrollees reported being registered to vote in the November 2016 election compared to 63.5% of enrollees, a 16.8 percentage point (p.p.) difference. Whereas non-enrollees were slightly more likely to not be registered, enrollees were almost two times as likely to be registered. The middle left facet reveals that regardless of college participation, the majority of young persons in our sample did not report any volunteer hours in 2015. Combining the enrollment groups, only about a third of the sample (37%) volunteered. Enrollees, however, were 21.7 p.p. more likely to report any volunteer activity than non-enrollees (42.1% vs. 20.4%). The bottom left facet shows the reverse of this trend among the subset of individuals who volunteered. While there is substantial overlap between the groups, non-enrollees reported an average of 12.8 hours of volunteer time ($e^{2.55}$) compared to an average of 7.4 hours (e^2) among enrollees. This translates to enrollees who volunteer volunteering for approximately 58% less time than non-enrollees who volunteer.

Taken on the whole, we observe that college enrollees in our sample are more likely than non-enrollees to register to vote and volunteer. However, conditional on choosing to volunteer, they are likely to volunteer for fewer hours than their non-enrolled peers. For evidence as to whether these differences represent the influence of college participation or are merely co-indicated with it, we turn to average treatment effect estimates from our propensity forests.

The right column of facets in figure 1 show the distribution of observation-level "out-of-bag"

---

12. Appendix table A.1 presents descriptive differences between the full HSLS09 survey sample and the samples used for each outcome.

(OOB) $\hat{\tau}(x)$ estimates. These represent the estimated impact—at the individual level—of college participation on each respective civic outcome. From these histograms, we can derive two key findings. First, the overall average treatment effect (ATE) of college participation largely mirrors the differences observed in the data, in direction though not intensity. Among the full sample, we estimate that college enrollees are 9.9 p.p. $(p < 0.01)$[13] more likely to report being registered to vote than non-enrollees. Once adjusted by the propensity for college participation, the difference in voter registration between the two enrollment groups is about 59% of that observed in the unadjusted averages and better reflects the direct influence of college participation on registration. Enrollees are also 7.2 p.p. $(p < 0.01)$ more likely to have volunteered at all during 2015, which, though one-third the difference observed in the unadjusted averages, remains large. Finally, enrollees who volunteer are likely to volunteer about 26% less time than their non-enrolled peers, which represents a little less than half the difference observed in the data.

As a point of comparison, our ATE estimate for voter registration among the full sample is in line with the $7 - 8.4$ p.p. increase in the likelihood of voting for each additional year of college reported by Doyle and Skinner (2017) and on the lower end of the $6.8 - 21.5$ p.p. increase reported by Dee (2004). Dee did not observe that college enrollment positively affected volunteering, while Doyle and Skinner report a small effect. That our estimate for volunteering is much larger may due to differences in the margin of the effect between our study and theirs, a point we discuss later.

The second important finding shown by the histograms is the heterogeneity in students' responses to college participation. The distribution of $\hat{\tau}(x)$ for voter registration suggests a 95% ATE range of $6 - 13.6$ p.p. $(\sigma_{\hat{\tau}(x)}^{registration} = 0.019)$. For volunteering and volunteer hours, similar ranges are $2.8 - 10.6$ p.p. $(\sigma_{\hat{\tau}(x)}^{volunteer} = 0.02)$ and -36% to -16% $(\sigma_{\hat{\tau}(x)}^{vol.hours} = 0.067)$, respectively. In the remainder of this section, we explore this heterogeneity in ATE across specifically defined subgroups. In the first step, we explore differences within gender, race/ethnicity, and poverty status. In each case, we rely on the full range of categories provided by HSLS09. For race/ethnicity, categories include: American Indian/Alaska Native, Asian, Black, Hispanic, more than one race,

---

13. This and all subsequent p-values are calculated as two-tailed tests of the ATE estimate's difference from zero.

Native Hawaiian/Pacific Islander, and white.[14] Poverty status reflects whether the student's family income in the base year of the survey was above or below 185% of the federal poverty line.[15] Gender is limited to the binary male and female categories reported in HSLS09.

Figure 2 shows ATEs for each civic outcome divided by three groups: gender, race/ethnicity, and poverty status. Once again, civic outcome—voter registration, volunteering, and volunteer hours—are reported by row. Within each facet, center points represent the ATE for the subgroup and vertical lines the 95% confidence interval of the estimate. For clarity, the specific value for each estimate is reported with the *p*-value from the two-tailed test of its difference from zero. Sample sizes corresponding to each estimate are printed directly below on the *x*-axis.[16]

Recalling that the overall ATE for voter registration due to college participation was 9.9 p.p., the top row of figure 2 shows little variation across groups. Two exceptions—Asian and white students—are 16 p.p. and 14 p.p. more likely than their respective non-enrolled counterparts to register to vote. Native Hawaiian/Pacific Island students have a point estimate that suggests a nearly 28 p.p. increase in the likelihood of registering, but this estimate is not statistically significant at conventional levels. In terms of volunteering, we find that women who enroll are 11 p.p. more likely to volunteer whereas men who enroll are only 4 p.p. more likely to do so. Only Hispanic (8 p.p.) and white (9 p.p.) enrollees have statistically significant positive ATEs. Among those who volunteer, men enrolled in college are likely to volunteer 37% less time compared to women who show no difference between the enrolled and non-enrolled. Hispanic (−37%), white (−27%), and higher income students (−29%) are all likely to volunteer less time.

Using propensity forests simplifies the computing and reporting of ATE estimates by sub-

---

14. Based on the non-overlapping coding structure used in HSLS09, we combine two separate categories for Hispanic students (Hispanic students, no race specified and Hispanic students, race specified) into a single category.

15. The federal poverty threshold is updated yearly and takes into account the number of persons in a household. For a family of three, the line was set at $17,600 (somewhat higher in Alaska and Hawaii), which is around $21k in 2019 dollars. Our indicator threshold would therefore be set at $32,560 or about $38.8k in 2019 dollars for a family of three and increasing with the number of household members.

16. Appendix table A.2 shows the full range of average treatment effect estimates—with accompanying standard errors and sample sizes—that are presented in figures 2 – 5. We also offer, as a point of comparison, results from simple LPM/OLS models in appendix table A.5. Briefly, we note that results from the propensity forests and regression-based models largely agree, though with differences in point estimate size and significance. However, the underlying conceptual differences between the two approaches means that we prefer the ATE estimates presented in the paper.

groups. This includes subgroups that in a regression framework would require prespecification of interaction terms in a regression framework. In this next section, we report ATEs of postsecondary participation on voting and volunteering for interacting subgroups: race/ethnicity by gender and race/ethnicity by poverty status.

Differences in the influence of college participation on our primary civic outcomes across race/ethnicity by gender are reported in figure 3. As with figure 2, each row of panels show ATEs for a particular outcome. With figure 3, the results for men are shown in the left facet of each pair and those for women in the right. All other aspects of the figure are the same as before. Beginning with the top panel of figure 3, we find some evidence that college participation differentially influences the likelihood of voter registration. Compared to the overall estimate of 9.9 p.p., estimates for subgroups range from $-11 - 41$ p.p. We interpret these differences with caution, however, for two reasons. First, many of the estimates lack precision, with confidence intervals that cross zero. Two ethnic subgroups in particular—American Indian/Alaska Native and Native Hawaiian/Pacific Islander—have comparatively small sample sizes, in some cases approaching the lower bounds of being large enough to support inference (< 30 observations). Second, 95% confidence intervals across the estimates overlap, meaning that even in the case of statistically significant estimates, we cannot say whether they are different from one another. This pattern applies across all subgroup analyses we perform.

What we do find, however, are the subgroups that appear to drive the overall results (potentially due to their respective sample sizes). In terms of registering to vote, Hispanic and white sample members of each gender respond most strongly to college participation. Hispanic men who enroll in college are 9 p.p. more likely to register than Hispanic men who do not enroll; for Hispanic women, the difference is 10 p.p. The relative responses of white men (14 p.p.) and women (15 p.p.) are slightly larger. Asian men and women who enroll show the largest ATEs (15 p.p. and 17 p.p., respectively), though the estimate for Asian women is not statistically significant by conventional standards. For volunteering (middle row of figure 3), we find some larger differences across subgroups. White women enrollees are about 12 p.p. more likely to volunteer than their

non-enrolled counterparts, an ATE that is twice the size of that for white men (6 p.p.). Hispanic women who participate in college are 12 p.p. more likely to volunteer compared to Hispanic men, whose 3 p.p. estimate is statistically indistinguishable from zero. Among those who volunteer (last row of figure 3), almost all ATEs are statistically insignificant with the exception of Hispanic and white men enrollees, who are likely to volunteer 54% and 34% less time than non-enrolled Hispanic and white men, respectively.

In figure 4, we consider the same primary outcomes but this time divide racial/ethnic subgroups into those who lived above and below 185% the poverty line in the base year of the HSLS09 survey. For voter registration, we find more variable results. Whereas white college enrollees show differences from their non-enrolled counterparts regardless of poverty status (12 – 15 p.p., respectively), only Hispanic college students below the poverty line and Native Hawaiian/Pacific Islanders above the poverty line are statistically more likely to register to vote than their non-enrolled counterparts (9 p.p. and 51 p.p., respectively.) Black enrollees above the poverty line are estimated to be 13 p.p. less likely to register to vote than non-enrolled Black men above the poverty line.

Poverty status appears to have less influence on volunteering behavior across racial/ethnic subgroups. Hispanic and white enrollees on both sides of the line are significantly more likely volunteer than non-enrollees, though the relative increases for Hispanic students above the poverty line is higher (14 p.p.) than for Hispanic students below the poverty line or for white students in each group—all of which are closer to the overall average (7 – 9 p.p.). Though the general pattern of point estimates is negative, only white students above the poverty line who volunteer are statistically likely to volunteer less time (-34%).

Thus far, we have explored pre-specified subgroups chosen based on their importance in the literature and much higher education policy. However, we can recover the estimated enrollment propensity assigned to each student ($\hat{w}_i$ from equation (1)) and compare students within discrete bands of the estimated propensity distribution. This follows the procedure used by Brand (2010)—with the difference that propensities in our specification refer the propensity of enrollment rather

than completion—and takes full advantage of all relevant predictors when making comparisons.

Figure 5 separates students by ten bands of enrollment propensity: (0–0.50], (0.50–0.60], (0.60–0.70], (0.70–0.80], (0.80–0.85], (0.85–0.90], (0.90–0.925], (0.925–0.95], (0.95–0.975], (0.975–1]. These bands were chosen based on the distribution of estimated propensities with the intention that groups have a generally similar number of observations across bands. In her study exploring heterogeneity in college participation's effect on volunteering, Brand (2010) finds greater effects among students with lower estimated propensities for college completion. Similarly, we find that college participation may have more impact among students with lower propensities of enrollment. For students with less than an 85% propensity to enroll, those who do are $7 - 12$ p.p. more likely to register to vote. Above 85% propensity of enrollment, all ATE estimates are non-significant. For volunteering, ATE estimates for groups with propensities of enrollment below 90% are generally positive, though not always significant; among students with the highest propensity of enrollment, college enrollees may actually be less likely to volunteer (-25 p.p.). While among volunteers all point estimates are negative—suggestive of the overall finding that enrolled volunteers volunteer fewer hours—not all ATEs are significant and those that are belie any clear pattern.

Across the outcomes we study, we find some evidence that college may differentially influence civic participation and beliefs across subpopulations of young people. Practically, these differences are often small, however, and overlapping confidence intervals prevent our being able to say whether they are statistically different from one another. We find stronger evidence when we compare subgroup enrollees with their non-enrolling, same-group counterparts. In some cases, our results reveal the extent to which overall averages may mute stronger responses to college participation among particular subpopulations. We discuss potential limitations to our study design as well as implications of our results in the next section.

Discussion

Does college promote civic behaviors and beliefs among a recent cohort of young people? We find evidence that, overall, the answer is yes. In this our results are in line with the positive correlations between college and civic engagement found in prior research (Bowman 2011; Colby et al. 2003; Evans, Marsicano, and Lennartz 2019; Hurtado 2007; Kezar, Chambers, and Burkhardt 2015; McMahon 2010; Perna 2005). Due to our use of propensity forests to produce our results, we argue that our estimates better reflect the influence of college participation on civic outcomes absent bias introduced by pre-college differences among our sample (Brand 2010; Dee 2004; Doyle and Skinner 2017).

Our estimates may retain some bias, however, in that we are unable to account for unobserved differences among our sample that may both change their propensity for college participation and civic engagement. We mitigate this concern through our use of propensity forests, which allow us to both include a larger number of predictors of college enrollment and implicitly model higher-level interactions between them than would be practicable with standard parametric propensity score models. We have more confidence in the degree to which our analyses account for differential selection into college than we would had we used other selection-on-observables-based procedures. Nonetheless, propensity forests retain the core conditional independence assumption of standard propensity score models—an assumption that we cannot test. We note our reliance on observable characteristics as a limitation of our study.

Overall, our estimates suggest that college participation has a small to moderate influence on the civic behaviors we investigate. While it could be true that the civic returns to college enrollment are generally low in comparison to the cost of attendance or the private financial returns (Doyle and Skinner 2017), our operationalization of college participation may have also contributed to the muted results. Those we consider to have participated in college only necessarily did so for a short time. By the time of the outcome, some students had earned a postsecondary credential; some were still enrolled, while others were not. Thus our indicator for enrollment neither accounts for the time spent in college nor whether a student was enrolled during the periods covered by the

outcomes.

The particular civic outcomes we investigate should also be noted. Based on the timing of the second follow up, students were asked whether they were registered to vote—not if they voted. Voter registration differs by location and is more opaque than voting itself. To vote, a person must physically go to a polling location or fill out an absentee ballot. Voter registration, on the other hand, may occur in any number of ways (for example, when a person registers for a drivers license) and is not necessarily updated when a person moves between states. Individuals in our sample may have been incorrect about their registration status, in which case our estimates better reflect an intention or desire to register to vote (Holbein and Hillygus 2020).

Prior research has found small or no effect of college participation on volunteering behavior (Dee 2004; Doyle and Skinner 2017). Our results suggest that the primary impact of college participation on volunteering happens at the extensive margin rather than the intensive margin. If college participation promotes civic engagement via volunteering, it appears to do so by inducing more young people to volunteer some time—perhaps through increased opportunities to do so—rather increasing volunteer hours among those who already volunteer.

If college positively influences civic participation and beliefs, do some student populations respond differently than others? For this question, our evidence is less clear. Though we show some differences across race/ethnicity, gender, low income status, and propensity for enrollment, estimates are not generally statistically significant from one another. It may be that our sample limited our ability to find them. In particular, two ethnic subgroups—American Indian/Alaska Native and Native Hawaiian/Pacific Islander—were represented by a small number of individuals. Our statistical precision may have been increased had we combined these categories with others, but our desire to model heterogeneity combined with the unique college experiences of these groups meant that we did not.

Our models may not have recovered heterogeneity in the sample. Checking various specifications of our models—specifically for differences in model fit as a function of the number of relevant predictors we chose to include—one calibration check offered suggestive evidence that

while our mean forest prediction was correct across specifications, only our model for volunteering found and accurately accounted for heterogeneity (see appendix table A.6 for the full range of results across various model specifications). We note this as a potential limitation of our results.

Finally, it also may be that the subpopulations we investigate do not have a strong differential response to college enrollment (which is also a possible cause of the calibration test results discussed above). Our conceptual framework suggests that a differential response to higher education in terms of civic engagement might be expected, given how differently groups can be served by our postsecondary system. In general, we find that responses are mostly similar. Were all subgroup ATE point estimates we present to stay the same and, through increased precision, we were able to show that all were statistically different from one another, it would still be the case that with only a few exceptions, our evidence would suggest that changes to civic participation due to college participation are broadly similar across groups.

Important societal differences in civic participation appear to occur between those who experience different intrinsic and extrinsic reward levels, with those who experience greater rewards being more inclined to civic participation. Though we conceptually separate them above, the rewards framework can include opportunity to participate. If the opportunity costs for civic engagement are low, then the rewards necessary to motivate engagement may not need be high. Conversely, greater opportunity costs may require commensurably greater rewards. Based on our findings, we posit that college participation effectively changes the rewards, intrinsic or extrinsic, for civic participation to a similar degree across groups. This claim suggests two possibilities: college participation changes the rewards of civic engagement to the same degree regardless of an individual's background (*i.e.*, truly homogeneous effect) or college participation differentially affects intrinsic and extrinsic rewards across groups, but in such a way that the combined effect as revealed by civic behavior is the same (*i.e.*, homogeneous effect in the aggregate). The second of these two possibilities leaves open the prospect that for some groups, college participation may have an inverse relationship to the rewards of civic engagement due to poor campus climate (see Baker and Blissett 2018). Because our analyses cannot differentiate between intrinsic and extrinsic rewards nor the

specific opportunities or experiences of individuals in our sample, we cannot decide between these two possibilities and instead leave this area for future research.

Conclusion

In this paper we offer evidence that college participation continues to positively influence the likelihood of voter registration and the choice to volunteer, while negatively influencing the time spent volunteering for a new generation of young people. Using propensity forests, we provide evidence that overall average treatment effect estimates may be driven by subpopulations of students. Therefore, though we are generally unable to say whether differences across subpopulations are statistically different from one another, we believe that further work that explores heterogeneous effects of college participation on civic outcomes is warranted. Propensity forests and causal forests more broadly (Wager and Athey 2018) represent useful approaches that should be added to the quantitative toolkit of education.

Previous work has established the increase in civic participation due to postsecondary participation. This paper provides updated estimates of this impact. In addition, our exploration of possible differential responses, enabled by a new methodological approach, suggests that changes in civic participation are broadly shared among all who enroll in postsecondary education. This provides further evidence that the public goods provided by higher education can be expected to be shared across a wide swathe of our society—a heartening result for those who support increased participation in higher education.

References

Angrist, Joshua D. 1997. "Conditional independence in sample selection models." *Economics Letters* 54 (2): 103–112.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31 (2): 3–32.

Baker, Dominique J, and Richard SL Blissett. 2018. "Beyond the incident: Institutional predictors of student collective action." *The Journal of Higher Education* 89 (2): 184–207.

Baum, Sandy, Jennifer Ma, and Kathleen Payea. 2013. *Education Pays 2013: The Benefits of Higher Education for Individuals and Society.* Trends in Higher Education Series. New York: College Board.

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature* 55 (3): 789–865.

Bowman, Nicholas A. 2011. "Promoting Participation in a Diverse Democracy A Meta-Analysis of College Diversity Experiences and Civic Engagement." *Review of Educational Research* 81, no. 1 (March): 29–68.

Brand, Jennie E, and Yu Xie. 2010. "Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education." *American sociological review* 75 (2): 273–302.

Brand, Jennie E. 2010. "Civic Returns to Higher Education: A Note on Heterogeneous Effects." *Social Forces* 89 (2): 417–433.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32.

Caliendo, Marco, and Sabine Kopeinig. 2008. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22 (1): 31–72.

Carrell, Scott E, Marianne E Page, and James E West. 2010. "Sex and science: How professor gender perpetuates the gender gap." *The Quarterly Journal of Economics* 125 (3): 1101–1144.

Cascio, Elizabeth U, and Na'ama Shenhav. 2020. "A Century of the American Woman Voter: Sex Gaps in Political Participation, Preferences, and Partisanship Since Women's Enfranchisement." *Journal of Economic Perspectives* 34 (2): 24–48.

Colby, Anne, Thomas Ehrlich, Elizabeth Beaumont, and Jason Stephens. 2003. *Educating Citizens: Preparing America's Undergraduates for Lives of Moral and Civic Responsibility.* John Wiley & Sons, June.

Dee, Thomas S. 2004. "Are there civic returns to education?" *Journal of Public Economics* 88 (9–10): 1697–1720.

Dehejia, Rajeev H, and Sadek Wahba. 2002. "Propensity score-matching methods for nonexperimental causal studies." *Review of Economics and statistics* 84 (1): 151–161.

Doyle, William R., and Benjamin T. Skinner. 2017. "Does Postsecondary Education Result in Civic Benefits?" *The Journal of Higher Education* 88 (6): 863–893.

Duprey, Michael A., Daniel J. Pratt, Donna M. Jewell, Melissa B. Cominole, Laura Burns Fritch, Ethan A. Ritchie, James E. Rogers, Jamie D. Wescott, and David H. Wilson. 2018. *High School Longitudinal Study of 2009 (HSLS:09) Base-Year to Second Follow-Up Data File Documentation.* NCES 2018-140. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Evans, Brent J, Christopher R Marsicano, and Courtney J Lennartz. 2019. "Cracks in the Bedrock of American Democracy: Differences in Civic Engagement Across Institutions of Higher Education." *Educational Researcher* 48 (1): 31–44.

Flanagan, Constance, and Peter Levine. 2010. "Civic engagement and the transition to adulthood." *The future of children* 20 (1): 159–179.

Garces, Liliana M, and Uma M Jayakumar. 2014. "Dynamic diversity: Toward a contextual understanding of critical mass." *Educational Researcher* 43 (3): 115–124.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *The American Political Science Review* 102 (1): 33–48.

Glynn, Adam N, and Kevin M Quinn. 2010. "An introduction to the augmented inverse propensity weighted estimator." *Political analysis,* 36–56.

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko. 2006. "The homecoming of American college women: The reversal of the college gender gap." *Journal of Economic perspectives* 20 (4): 133–156.

Goldin, Claudia Dale, and Lawrence F. Katz. 2009. *The Race between Education and Technology.* Harvard University Press.

Harwood, Stacy A, Margaret Browne Huntt, Ruby Mendenhall, and Jioni A Lewis. 2012. "Racial microaggressions in the residence halls: Experiences of students of color at a predominantly White university." *Journal of Diversity in Higher Education* 5 (3): 159.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction.* Springer Series in Statistics. New York: Springer.

Hillygus, D. Sunshine. 2005. "The MISSING LINK: Exploring the Relationship Between Higher Education and Political Engagement." *Political Behavior* 27 (1): 25–47.

Holbein, John B., and D. Sunshine Hillygus. 2020. *Making Young Voters: Converting Civic Attitudes into Civic Action.* Cambridge University Press.

Hoxby, Caroline M. 2007. *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It.* University of Chicago Press.

Hurtado, Sylvia. 2007. "Linking Diversity with the Educational and Civic Missions of Higher Education." *The Review of Higher Education* 30 (2): 185–196.

Imbens, Guido W. 2015. "Matching methods in practice: Three examples." *Journal of Human Resources* 50 (2): 373–419.

Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.

Kezar, Anthony C. Chambers, and John C. Burkhardt. 2015. *Higher Education for the Public Good: Emerging Voices from a National Movement.* John Wiley & Sons.

Li, Fan, Kari Lock Morgan, and Alan M. Zaslavsky. 2018. "Balancing Covariates via Propensity Score Weighting." *Journal of the American Statistical Association* 113 (521): 390–400.

Long, Bridget Terry. 2004. "How have college decisions changed over time? An application of the conditional logistic choice model." *Journal of Econometrics* 121:271–296.

Marginson, Simon. 2007. "The public/private divide in higher education: A global revision." *Higher Education* 53 (3): 307–333.

Mayhew, Matthew J, Ernest T Pascarella, Nicholas A Bowman, Alyssa N Rockenbach, Tricia AD Seifert, Patrick T Terenzini, and Gregory C Wolniak. 2016. *How college affects students: 21st century evidence that higher education works.* Vol. 3. John Wiley & Sons.

McMahon, Walter W. 2010. *Higher Learning, Greater Good: The Private and Social Benefits of Higher Education.* JHU Press, April.

National Conference of State Legislatures. 2019. *Women in State Legislatures for 2019.* Technical report. Https://www.ncsl.org/legislators-staff/legislators/womens-legislative-network/women-in-state-legislatures-for-2019.aspx. Denver, CO: National Conference of State Legislatures.

Oesterle, Sabrina, Monica Kirkpatrick Johnson, and Jeylan T Mortimer. 2004. "Volunteerism during the transition to adulthood: A life course perspective." *Social forces* 82 (3): 1123–1149.

Panagopoulos, Costas. 2013. "Extrinsic rewards, intrinsic motivation and voting." *The Journal of Politics* 75 (1): 266–280.

Perna, Laura W. 2005. "The Benefits of Higher Education: Sex, Racial/Ethnic, and Socioeconomic Group Differences." *The Review of Higher Education* 29, no. 1 (September): 23–52.

———. 2006. "Studying college access and choice: A proposed conceptual model." In *Higher Education: Handbook of Theory and Research,* edited by John C. Smart, 99–157. Springer Netherlands.

Persson, Mikael. 2015. "Education and Political Participation." *British Journal of Political Science* 45, no. 3 (July): 689–703.

Pierce, Chester M, Jean V Carew, Diane Pierce-Gonzalez, and Deborah Wills. 1977. "An experiment in racism: TV commercials." *Education and Urban Society* 10 (1): 61–87.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Robins, James M., Andrea Rotnitzky, and Lue Ping Zhao. 1994. "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed." *Journal of the American Statistical Association* 89 (427): 846–866.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70, no. 1 (April): 41–55.

Skinner, Benjamin T. 2019. "Choosing College in the 2000s: An Updated Analysis Using the Conditional Logistic Choice Model." *Research in Higher Education* 60 (2): 153–183.

Speer, Jamin D. 2017. "The gender gap in college major: Revisiting the role of pre-college factors." *Labour Economics* 44:69–88.

Standard Occupational Classification and Coding Structure. 2018. U.S. Bureau of Labor Statistics.

Tibshirani, Julie, Susan Athey, and Stefan Wager. 2020. *grf: Generalized Random Forests.* R package version 1.2.0.

Verba, Sidney, and Norman H. Nie. 1972. *Participation in America: Political Democracy and Social Equality.* Harper / Row.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–1242.

Wolfinger, Raymond E, and Steven J Rosenstone. 1980. *Who Votes?* Vol. 22. Yale University Press.

Figure 1: Variation in civic participation as seen in the data (left) and as a function of college participation (right).
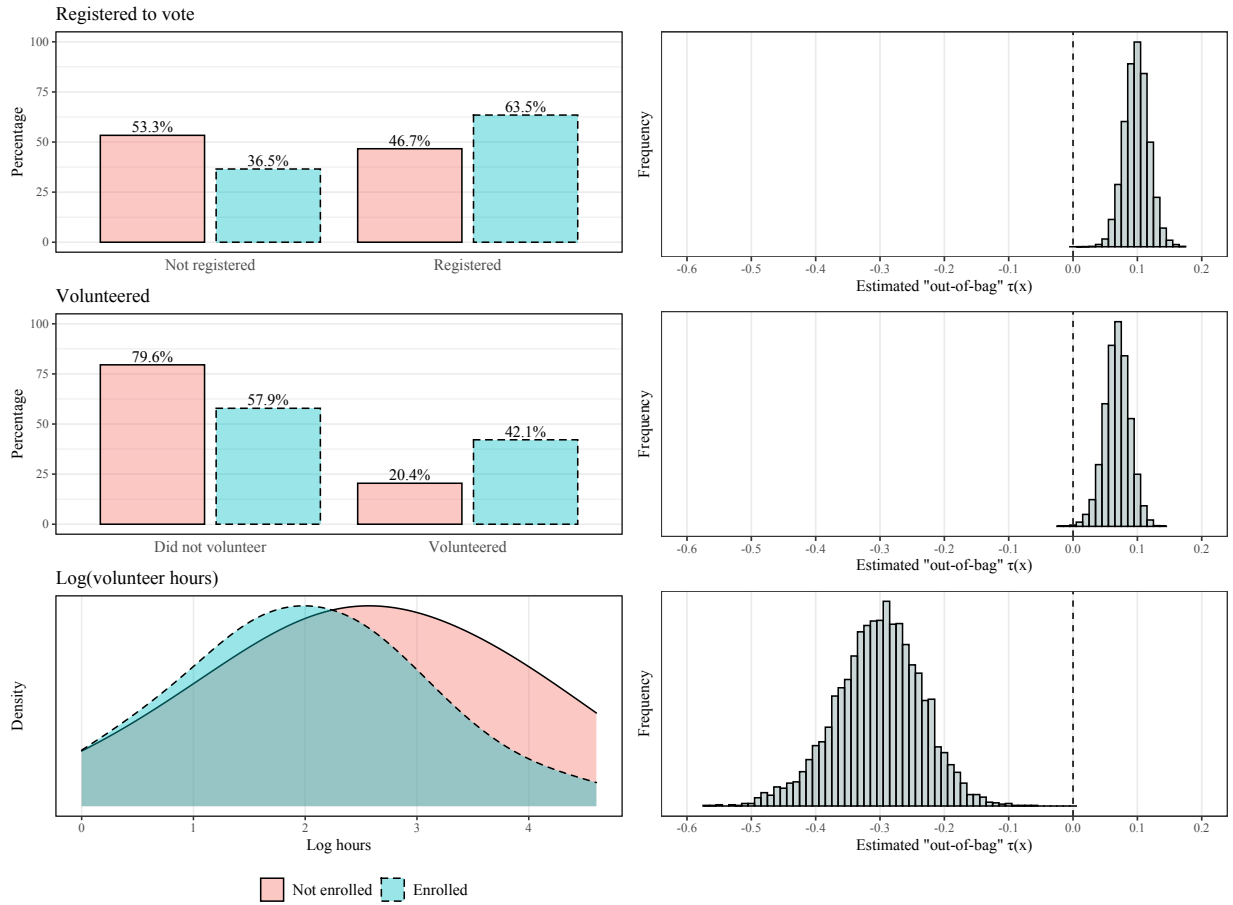
Note. All values represent unweighted averages by subgroup. Percentages show in left-hand side facets are relative within enrollment condition. Histograms on the right-hand side show observation-level variation in the $\tau(x)$ estimates produced by the propensity forest models.

Figure 2: Estimated returns of college enrollment on voting and volunteering behavior by gender, race / ethnicity, and poverty status

Note. Center points represent the overlap-weighted conditional average treatment effect estimate for each subgroup, with vertical lines plotting the 95% confidence interval. Each point is labeled with its estimate as well as the *p*-value from the test of its difference from zero. Both values are rounded to two significant digits. Sample sizes for each subgroup are printed on the *x*-axis.

Figure 3: Estimated returns of college enrollment on voting and volunteering behavior: gender by race/ethnicity
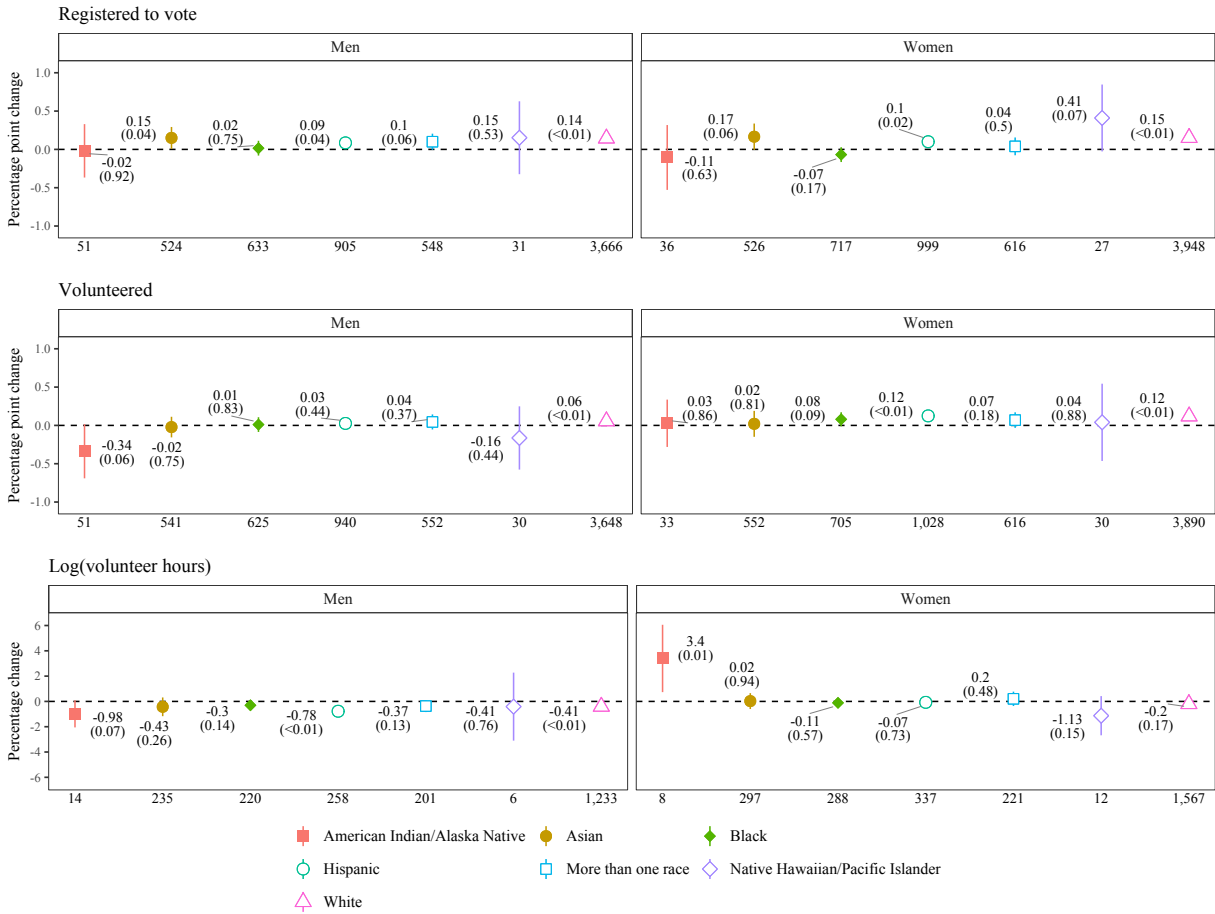
Note. Center points represent the overlap-weighted conditional average treatment effect estimate for each subgroup, with vertical lines plotting the 95% confidence interval. Each point is labeled with its estimate as well as the *p*-value from the test of its difference from zero. Both values are rounded to two significant digits. Sample sizes for each subgroup are printed on the *x*-axis.

Figure 4: Estimated returns of college enrollment on voting and volunteering behavior: poverty status by race/ethnicity
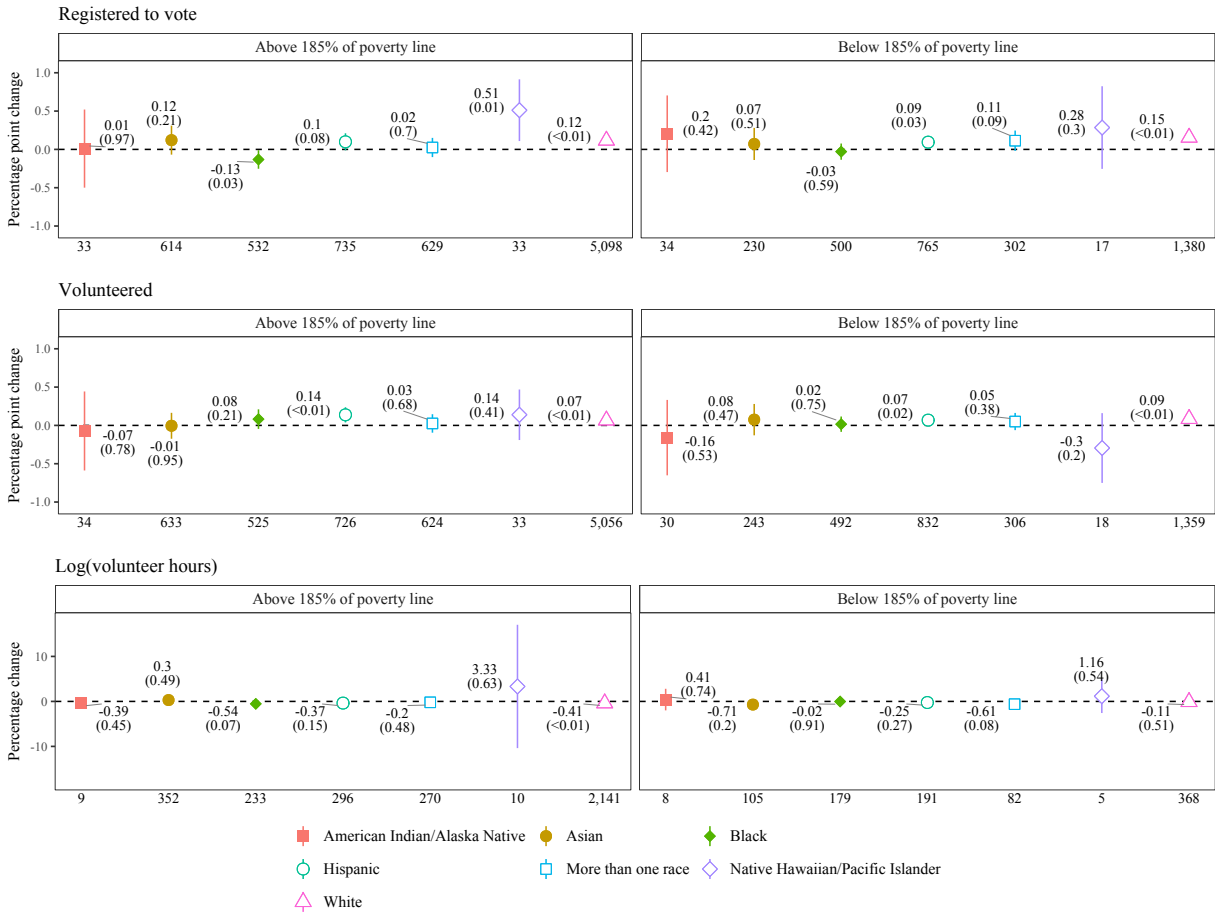
Note. Center points represent the overlap-weighted conditional average treatment effect estimate for each subgroup, with vertical lines plotting the 95% confidence interval. Each point is labeled with its estimate as well as the *p*-value from the test of its difference from zero. Both values are rounded to two significant digits. Sample sizes for each subgroup are printed on the *x*-axis.

Figure 5: Estimated returns of college enrollment on voting and volunteering behavior by propensity of enrollment
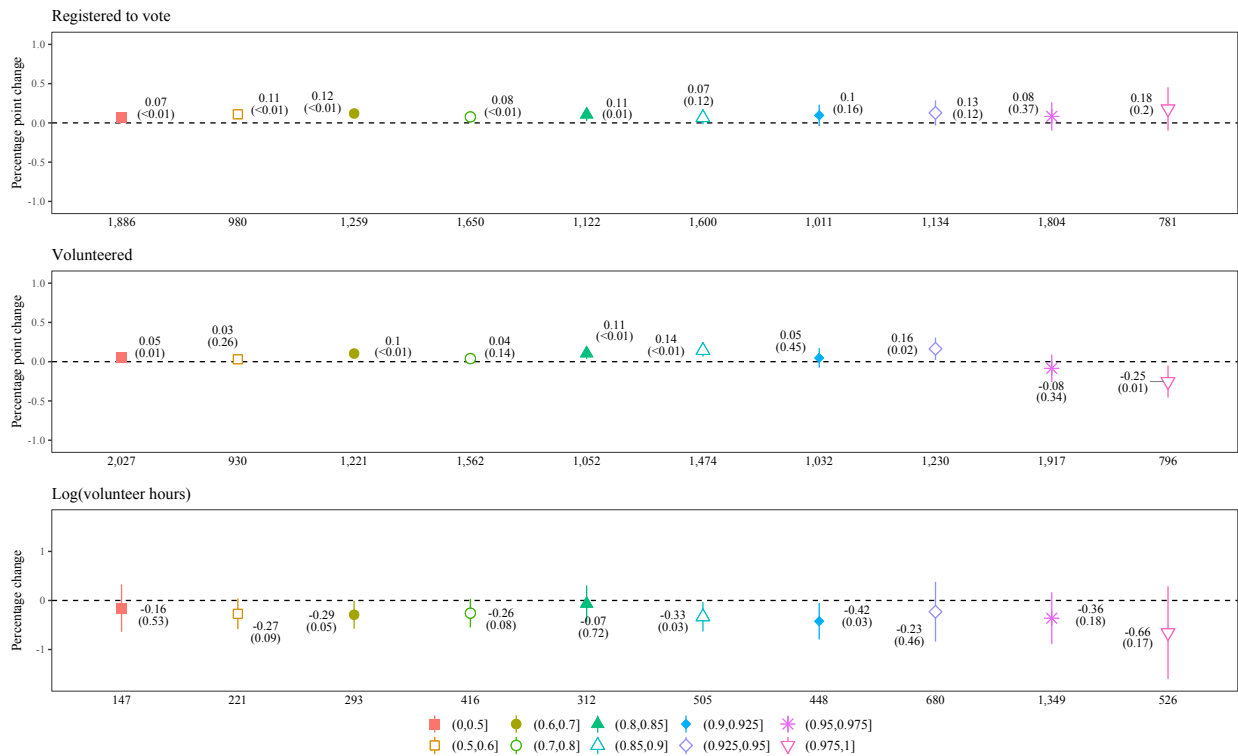
Note. Center points represent the overlap-weighted conditional average treatment effect estimate for each subgroup, with vertical lines plotting the 95% confidence interval. Each point is labeled with its estimate as well as the *p*-value from the test of its difference from zero. Both values are rounded to two significant digits. Sample sizes for each subgroup are printed on the *x*-axis.

Table A.1: Comparison of model samples with full HSLS sample

| | Full HSLS | | Outcomes | | |
	(1)	(2)	Registered to vote	Volunteer	Log(volunteer hours)
Enrollment					
Non-enroll	18.21	24.7	23.63	23.78	13.15
Enroll	55.54	75.3	76.37	76.22	86.85
Missing	26.24				
Gender					
Male	50.94	50.96	48.07	48.24	44.25
Female	49.03	49.04	51.93	51.76	55.75
Missing	0.03				
Race/Ethnicity					
American Indian/Alaska Native	0.7	0.73	0.66	0.63	0.45
Asian	8.31	8.68	7.94	8.25	10.86
Black	10.42	10.89	10.21	10.04	10.37
Hispanic	16.16	16.88	14.39	14.86	12.15
More than one race	8.26	8.63	8.8	8.82	8.62
Native Hawaiian/Pacific Islander	0.47	0.49	0.44	0.45	0.37
White	51.41	53.7	57.56	56.93	57.18
Missing	4.28				
Poverty status					
Above 185% of poverty line	47.55	66.78	58.02	57.63	67.61
Below 185% of poverty line	23.65	33.22	24.4	24.77	19.15
Missing	28.8		17.58	17.6	13.23
N	23503		13227	13241	4897

Note. All values are percentages. Column (1) includes missing values in the full HSLS sample as their own category. Column (2) recomputes the percentages after dropping missing categorical values. Percentages in the outcome columns represent the analytic samples used to fit each model

Table A.2: Average treatment effect estimates across subgroups

	Registered to vote		Volunteer		Log(volunteer hours)	
	ATE	*N*	ATE	*N*	ATE	*N*
Overall	**0.099**	**13227**	**0.072**	**13241**	**-0.307**	**4897**
	(0.0124)		**(0.0108)**		**(0.0598)**	
Single group						
Gender						
Men	0.106	6358	0.036	6387	-0.459	2167
	(0.0169)		(0.0152)		(0.0785)	
Women	0.099	6869	0.106	6854	-0.103	2730
	(0.0181)		(0.0153)		(0.0921)	
Race/ethnicity						
American Indian/Alaska Native	-0.06	87	-0.197	84	-0.35	22
	(0.1316)		(0.1287)		(0.5383)	
Asian	0.156	1050	-0.002	1093	-0.228	532
	(0.0555)		(0.0533)		(0.2486)	
Black	-0.027	1350	0.049	1330	-0.201	508
	(0.0347)		(0.0338)		(0.1422)	
Hispanic	0.09	1904	0.076	1968	-0.468	595
	(0.0294)		(0.0236)		(0.1427)	
More than one race	0.067	1164	0.053	1168	-0.17	422
	(0.0395)		(0.0359)		(0.1867)	
Native Hawaiian/Pacific Islander	0.283	58	-0.064	60	-0.85	18
	(0.159)		(0.1627)		(0.6324)	
White	0.142	7614	0.088	7538	-0.32	2800
	(0.0168)		(0.0145)		(0.0883)	
Poverty line						
Below 185% of poverty line	0.098	3228	0.066	3280	-0.196	938
	(0.0212)		(0.0179)		(0.1028)	
Above 185% of poverty line	0.086	7674	0.069	7631	-0.341	3311
	(0.0192)		(0.0175)		(0.0915)	
Gender by race/ethnicity						
Men						
American Indian/Alaska Native	-0.019	51	-0.337	51	-0.98	14
	(0.1778)		(0.1805)		(0.548)	
Asian	0.15	524	-0.022	541	-0.427	235
	(0.0727)		(0.0686)		(0.3772)	
Black	0.016	633	0.011	625	-0.305	220
	(0.0492)		(0.049)		(0.2042)	
Hispanic	0.085	905	0.026	940	-0.777	258
	(0.0409)		(0.0335)		(0.1931)	
More than one race	0.1	548	0.045	552	-0.366	201
	(0.0536)		(0.0501)		(0.2429)	
Native Hawaiian/Pacific Islander	0.152	31	-0.164	30	-0.413	6
	(0.2426)		(0.2105)		(1.3713)	
White	0.144	3666	0.055	3648	-0.408	1233
	(0.0228)		(0.0206)		(0.1115)	
Women						
American Indian/Alaska Native	-0.105	36	0.028	33	3.396	8
	(0.2159)		(0.1576)		(1.3578)	

	Registered to vote		Volunteer		Log(volunteer hours)	
	ATE	*N*	ATE	*N*	ATE	*N*
Asian	0.166 (0.0875)	526	0.02 (0.0863)	552	0.025 (0.3201)	297
Black	-0.068 (0.0493)	717	0.08 (0.0469)	705	-0.115 (0.2013)	288
Hispanic	0.1 (0.0424)	999	0.124 (0.033)	1028	-0.074 (0.211)	337
More than one race	0.04 (0.0592)	616	0.069 (0.0518)	616	0.204 (0.29)	221
Native Hawaiian/Pacific Islander	0.41 (0.2227)	27	0.04 (0.2572)	30	-1.129 (0.7904)	12
White	0.151 (0.0248)	3948	0.119 (0.0203)	3890	-0.2 (0.146)	1567
Poverty status by race/ethnicity						
Below 185% of poverty line						
American Indian/Alaska Native	0.204 (0.2551)	34	-0.16 (0.251)	30	0.407 (1.2234)	8
Asian	0.071 (0.1068)	230	0.075 (0.1044)	243	-0.708 (0.5516)	105
Black	-0.029 (0.054)	500	0.016 (0.051)	492	-0.023 (0.2084)	179
Hispanic	0.094 (0.0426)	765	0.069 (0.0303)	832	-0.248 (0.2261)	191
More than one race	0.113 (0.0677)	302	0.051 (0.0573)	306	-0.612 (0.3517)	82
Native Hawaiian/Pacific Islander	0.285 (0.275)	17	-0.295 (0.2323)	18	1.159 (1.9073)	5
White	0.155 (0.0323)	1380	0.086 (0.0277)	1359	-0.108 (0.1626)	368
Above 185% of poverty line						
American Indian/Alaska Native	0.011 (0.2601)	33	-0.073 (0.263)	34	-0.39 (0.5212)	9
Asian	0.121 (0.0964)	614	-0.005 (0.0863)	633	0.299 (0.4322)	352
Black	-0.132 (0.0615)	532	0.082 (0.065)	525	-0.54 (0.2928)	233
Hispanic	0.099 (0.0568)	735	0.138 (0.0504)	726	-0.373 (0.2588)	296
More than one race	0.025 (0.0639)	629	0.026 (0.0615)	624	-0.201 (0.2841)	270
Native Hawaiian/Pacific Islander	0.512 (0.2054)	33	0.139 (0.1682)	33	3.327 (6.9925)	10
White	0.116 (0.0237)	5098	0.066 (0.0213)	5056	-0.406 (0.1178)	2141
Propensity of enrollment						
(0,0.5]	0.071 (0.0265)	1886	0.054 (0.0214)	2027	-0.157 (0.2466)	147
(0.5,0.6]	0.11 (0.0317)	980	0.031 (0.0273)	930	-0.272 (0.1589)	221
(0.6,0.7]	0.12	1259	0.103	1221	-0.291	293

	Registered to vote		Volunteer		Log(volunteer hours)	
	ATE	*N*	ATE	*N*	ATE	*N*
	(0.0294)		(0.0246)		(0.1463)	
(0.7,0.8]	0.077	1650	0.039	1562	-0.259	416
	(0.0293)		(0.0268)		(0.1478)	
(0.8,0.85]	0.106	1122	0.106	1052	-0.067	312
	(0.0416)		(0.0369)		(0.1903)	
(0.925,0.95]	0.127	1134	0.164	1230	-0.231	680
	(0.0818)		(0.0727)		(0.31)	
(0.95,0.975]	0.083	1804	-0.085	1917	-0.362	1349
	(0.0923)		(0.0888)		(0.27)	
(0.975,1]	0.179	781	-0.253	796	-0.656	526
	(0.1409)		(0.104)		(0.4817)	

Note. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. *ATE*: Average treatment effect; *N* is sample size. These estimates are the same as shown in the figures 2-5, with the exception that standard errors (rather than *p*-values) are shown in parentheses.

Table A.3: Predictor names and descriptions

Predictor name	Predictor description
X1SEX	Student's sex
X1RACE	Student's race/ethnicity-composite
X1DUALLANG	Student dual-first language indicator
X1STDOB	Student's date of birth (YYYYMM)
X1TXMTH	Mathematics theta score
X1MACC	Mathematics assessment accommodations
X1PARRESP	Whether parent questionnaire respondent is Parent 1
X1P1RELATION	Parent 1: relationship to 9th grader
X1PAR1EDU	Parent 1: highest level of education
X1PAR1EMP	Parent 1: employment status
X1PAR1OCC2	Parent 1: current/most recent occupation: 2-digit ONET code
X1PAR1RACE	Parent 1: race/ethnicity
X1P2RELATION	Parent 2: spouse's relationship to 9th grader
X1PAR2EDU	Parent 2: highest level of education
X1PAR2EMP	Parent 2: employment status
X1PAR2OCC2	Parent 2: current/most recent occupation: 2-digit ONET code
X1PAR2RACE	Parent 2: race/ethnicity
X1PAREDU	Parents'/guardians' highest level of education
X1HHNUMBER	Number of 2009 household members
X1FAMINCOME	Total family income from all sources 2008
X1POVERTY	Poverty indicator (relative to 100% of Census poverty threshold)
X1POVERTY130	Poverty indicator (relative to 130% of Census poverty threshold)
X1POVERTY185	Poverty indicator (relative to 185% of Census poverty threshold)
X1SES	Socio-economic status composite
X1MTHID	Scale of student's mathematics identity
X1MTHUTI	Scale of student's mathematics utility
X1MTHEFF	Scale of student's mathematics self-efficacy
X1MTHINT	Scale of student's interest in fall 2009 math course
X1SCIID	Scale of student's science identity
X1SCIUTI	Scale of student's science utility
X1SCIEFF	Scale of student's science self-efficacy
X1SCIINT	Scale of student's interest in fall 2009 science course
X1SCHOOLBEL	Scale of student's sense of school belonging
X1SCHOOLENG	Scale of student's school engagement
X1STU30OCC2	Student occupation at age 30: 2-digit ONET code
X1STUEDEXPCT	How far in school 9th grader thinks he/she will get
X1PAREDEXPCT	How far in school parent thinks 9th grader will go
X1IEPFLAG	Individualized Education Plan
X1PQLANG	Parent questionnaire language (English v. Spanish)
X1TMRACE	Math teacher's race/ethnicity-composite
X1TMCERT	Math teacher's math teaching certification
X1TMCOMM	Scale of math teacher's perceptions of math professional learning community
X1TMEFF	Scale of math teacher's self-efficacy
X1TMEXP	Scale of math teacher's perceptions of math teachers' expectations
X1TMPRINC	Scale of math teacher's perceptions of principal support
X1TMRESP	Scale of math teacher's perceptions of collective responsibility
X1TSRACE	Science teacher race/ethnicity-composite

Predictor name	Predictor description
X1TSCERT	Science teacher's science teaching certification
X1TSCOMM	Scale of science teacher's perceptions of science professional learning community
X1TSEFF	Scale of science teacher's self-efficacy
X1TSEXP	Scale of science teacher's perceptions of science teachers expectations
X1TSPRINC	Scale of science teacher's perceptions of principal support
X1TSRESP	Scale of science teacher's perceptions of collective responsibility
X1CONTROL	School control
X1LOCALE	School locale (urbanicity)
X1REGION	School geographic region
X1SCHOOLCLI	Scale of administrator's assessment of school climate
X1COUPERTEA	Scale of counselor's perceptions of teacher expectations
X1COUPERCOU	Scale of counselor's perceptions of counselor expectations
X1COUPERPRI	Scale of counselor's perceptions of principal's expectations
X2ENROLSTAT	Student enrollment status
X2EVERDROP	Ever dropout
X2DROPSTAT	F1 dropout status
X2SAMEPAR1	Same parent 1 as in the base year
X2SAMEPAR2	Same parent 2 as in the base year
X2NUMHS	Number of high schools attended
X2TXMTH	Mathematics theta score
X2MACC	Mathematics assessment accommodations
X2P1RELATION	Parent 1: relationship to sample member
X2PAR1EDU	Parent 1: highest level of education
X2PAR1EMP	Parent 1: employment status
X2PAR1OCC2	Parent 1: current/most recent occupation: 2-digit ONET code
X2PAR1RACE	Parent 1: race/ethnicity
X2P2RELATION	Parent 2: spouse's relationship to sample member
X2PAR2EDU	Parent 2: highest level of education
X2PAR2EMP	Parent 2: employment status
X2PAR2OCC2	Parent 2: current/most recent occupation: 2-digit ONET code
X2PAR2RACE	Parent 2: race/ethnicity
X2PAREDU	Parents'/guardians' highest level of education
X2HHNUMBER	Number of 2012 household members
X2POVERTY	Poverty indicator (relative to 100% of Census poverty threshold)
X2POVERTY130	Poverty indicator (relative to 130% of Census poverty threshold)
X2POVERTY185	Poverty indicator (relative to 185% of Census poverty threshold)
X2SES	Socio-economic status composite
X2REPEATG11	Percent of 11th graders repeating 11th grade-categorical
X2RETURNG11	Percent of 11th graders returning to school-categorical
X2BEHAVEIN	Scale of school motivation
X2MEFFORT	Scale of math class effort
X2SEFFORT	Scale of science class effort
X2PROBLEM	Scale of problems at high school
X2MTHID	Scale of student's mathematics identity
X2MTHUTI	Scale of student's mathematics utility
X2MTHEFF	Scale of student's mathematics self-efficacy
X2MTHINT	Scale of student's interest in fall 2009 math course
X2SCIID	Scale of student's science identity
X2SCIUTI	Scale of student's science utility

...table A.3 continued

Predictor name	Predictor description
X2SCIEFF	Scale of student's science self-efficacy
X2SCIINT	Scale of student's interest in fall 2009 science course
X2STU30OCC2	Student occupation at age 30: 2-digit ONET code
X2STUEDEXPCT	How far in school sample member thinks he/she will get
X2PAREDEXPCT	How far in school parent thinks sample member will go
X2S2SSPR12	Teenager taking science/computer science/tech class(es) in spring 2012
X2REQLEVEL	Highest level of education student indicates will meet minimum requirements
X2S2EARNNOHS	Earnings without HS diploma standardized by year
X2S2EARNHS	Earnings with HS diploma standardized by year
X2S2EARNOCC	Earnings with occupational training diploma standardized by year
X2S2EARN2YPUB	Earnings with two year college degree standardized by year
X2S2EARN4Y	Earnings with four year college degree standardized by year
X2PQLANG	Parent questionnaire language (English v. Spanish)
X2CONTROL	School control
X2LOCALE	School locale (urbanicity)
X2REGION	School geographic region
X2SCHOOLCLI	Scale of administrator's assessment of school climate

Note. Predictor names and labels come directly from the HSLS 2009 variable list file found at the National Center for Education Statistics website: https://nces.ed.gov/surveys/hsls09/hsls09_data.asp.

Table A.4: Predictors used across models

	Registered to vote	Volunteered	Log(volunteer hours)
X1SEX*	O	O	O
X1RACE*	O	O	O
X1DUALLANG*	.	.	.
X1STDOB	X	.	X
X1TXMTH	X	X	X
X1MACC*	.	.	.
X1PARRESP*	.	.	.
X1P1RELATION*	.	.	.
X1PAR1EDU	.	.	.
X1PAR1EMP	.	.	.
X1PAR1OCC2	.	.	.
X1PAR1RACE*	X	.	.
X1P2RELATION*	.	.	.
X1PAR2EDU	.	.	.
X1PAR2EMP	.	.	.
X1PAR2OCC2	.	.	.
X1PAR2RACE*	.	.	.
X1PAREDU	.	.	.
X1HHNUMBER	.	.	.
X1FAMINCOME	.	.	.
X1POVERTY*	.	.	.
X1POVERTY130*	.	.	.
X1POVERTY185*	O	O	O
X1SES	X	X	X
X1MTHID	.	.	.
X1MTHUTI	.	X	.
X1MTHEFF	.	X	X
X1MTHINT	X	X	X
X1SCIID	.	.	.
X1SCIUTI	.	X	X
X1SCIEFF	.	X	.
X1SCIINT	X	.	X
X1SCHOOLBEL	X	X	X
X1SCHOOLENG	X	X	X
X1STU30OCC2	.	.	.
X1STUEDEXPCT	.	X	X
X1PAREDEXPCT	.	.	X
X1IEPFLAG*	.	.	.
X1PQLANG*	.	.	.
X1TMRACE*	.	.	.
X1TMCERT*	.	.	.
X1TMCOMM	X	X	.
X1TMEFF	X	X	X
X1TMEXP	X	X	.
X1TMPRINC	X	X	X
X1TMRESP	.	.	.
X1TSRACE*	.	.	.
X1TSCERT*	.	.	.

	Registered to vote	Volunteered	Log(volunteer hours)
X1TSCOMM	X	X	·
X1TSEFF	·	X	X
X1TSEXP	X	·	X
X1TSPRINC	X	·	·
X1TSRESP	X	·	X
X1CONTROL*	·	·	·
X1LOCALE*	·	·	·
X1REGION*	·	·	·
X1SCHOOLCLI	X	X	X
X1COUPERTEA	X	·	X
X1COUPERCOU	X	·	·
X1COUPERPRI	X	X	·
X2ENROLSTAT*	·	·	·
X2EVERDROP*	·	·	·
X2DROPSTAT*	·	X	·
X2SAMEPAR1*	·	·	·
X2SAMEPAR2*	·	·	·
X2NUMHS	·	X	X
X2TXMTH	X	X	X
X2MACC*	·	·	·
X2P1RELATION*	·	·	·
X2PAR1EDU	·	·	·
X2PAR1EMP	·	·	·
X2PAR1OCC2	·	·	·
X2PAR1RACE*	·	·	·
X2P2RELATION*	·	·	·
X2PAR2EDU	·	·	·
X2PAR2EMP	·	·	·
X2PAR2OCC2	·	·	·
X2PAR2RACE*	·	·	·
X2PAREDU	·	·	·
X2HHNUMBER	·	·	·
X2POVERTY*	·	·	·
X2POVERTY130*	·	·	·
X2POVERTY185*	·	·	·
X2SES	X	X	X
X2REPEATG11	·	·	·
X2RETURNG11	·	·	·
X2BEHAVEIN	X	X	X
X2MEFFORT	X	X	X
X2SEFFORT	·	·	·
X2PROBLEM	X	X	X
X2MTHID	·	·	·
X2MTHUTI	X	·	·
X2MTHEFF	X	·	X
X2MTHINT	X	·	·
X2SCIID	·	X	·
X2SCIUTI	·	·	·
X2SCIEFF	·	X	X
X2SCIINT	·	·	·

	Registered to vote	Volunteered	Log(volunteer hours)
X2STU30OCC2	·	·	·
X2STUEDEXPCT	·	X	X
X2PAREDEXPCT	·	X	X
X2S2SSPR12*	·	·	·
X2REQLEVEL	·	·	X
X2S2EARNNOHS	·	·	·
X2S2EARNHS	·	·	·
X2S2EARNOCC	·	·	·
X2S2EARN2YPUB	·	·	·
X2S2EARN4Y	·	·	·
X2PQLANG*	·	·	·
X2CONTROL*	·	·	·
X2LOCALE*	·	·	·
X2REGION*	·	·	·
X2SCHOOLCLI	·	X	X

Note. Initial propensity forest models for each outcome included all predictors listed in the table. Factor predictors, which were converted to sets of binary indicators, are marked with an asterisk. *O*s represent subgroup predictors; *X*s are the most important predictors (exclusive of subgroup predictors) from each initial estimation. Results presented in the paper come from propensity forest estimations using only these two sets of predictors for each outcome.

Table A.5: Results from outcomes regressed on indicator for college participation

	Registered to vote	Volunteer	Log(volunteer hours)
Model 1			
Enrolled	0.14***	0.12***	-0.43***
	(0.011)	(0.01)	(0.049)
Model 2			
Enrolled *X* Men	0.16***	0.08***	-0.47***
	(0.012)	(0.011)	(0.053)
Enrolled *X* Women	0.11***	0.17***	-0.38***
	(0.011)	(0.011)	(0.051)
Model 3			
Enrolled *X* American Indian/Alaska Native	0.14*	-0.03	-0.18
	(0.067)	(0.066)	(0.317)
Enrolled *X* Asian	0.01	0.19***	-0.3***
	(0.018)	(0.018)	(0.068)
Enrolled *X* Black	0.15***	0.17***	-0.26***
	(0.018)	(0.018)	(0.07)
Enrolled *X* Hispanic	0.09***	0.12***	-0.42***
	(0.016)	(0.015)	(0.066)
Enrolled *X* More than one race	0.13***	0.12***	-0.42***
	(0.019)	(0.018)	(0.074)
Enrolled *X* Native Hawaiian/Pacific Islander	0.07	0.05	-0.39
	(0.079)	(0.074)	(0.305)
Enrolled *X* White	0.16***	0.12***	-0.51***
	(0.012)	(0.011)	(0.052)
Model 4			
Enrolled *X* Above 185% of poverty line	0.16***	0.15***	-0.52***
	(0.012)	(0.012)	(0.053)
Enrolled *X* Below 185% of poverty line	0.15***	0.13***	-0.41***
	(0.014)	(0.013)	(0.06)
Model 5			
Enrolled *X* American Indian/Alaska Native *X* Men	0.18	-0.06	-0.42
	(0.091)	(0.088)	(0.446)
Enrolled *X* Asian *X* Men	0.01	0.14***	-0.32***
	(0.024)	(0.023)	(0.087)
Enrolled *X* Black *X* Men	0.16***	0.12***	-0.31**
	(0.025)	(0.024)	(0.096)
Enrolled *X* Hispanic *X* Men	0.1***	0.08***	-0.52***
	(0.022)	(0.021)	(0.091)
Enrolled *X* More than one race *X* Men	0.15***	0.12***	-0.4***
	(0.026)	(0.025)	(0.098)
Enrolled *X* Native Hawaiian/Pacific Islander *X* Men	0.07	-0.08	-0.47
	(0.105)	(0.104)	(0.545)
Enrolled *X* White *X* Men	0.18***	0.07***	-0.55***
	(0.014)	(0.013)	(0.058)
Enrolled *X* American Indian/Alaska Native *X* Women	0.09	0.02	0.08
	(0.099)	(0.097)	(0.446)
Enrolled *X* Asian *X* Women	0	0.24***	-0.27***
	(0.024)	(0.023)	(0.08)
Enrolled *X* Black *X* Women	0.14***	0.21***	-0.2*
	(0.023)	(0.022)	(0.083)
Enrolled *X* Hispanic *X* Women	0.07***	0.16***	-0.33***
	(0.02)	(0.019)	(0.077)
Enrolled *X* More than one race *X* Women	0.1***	0.13***	-0.42***
	(0.024)	(0.023)	(0.09)
Enrolled *X* Native Hawaiian/Pacific Islander *X* Women	0.07	0.18	-0.33
	(0.117)	(0.104)	(0.365)
Enrolled *X* White *X* Women	0.13***	0.16***	-0.47***
	(0.013)	(0.013)	(0.055)
Model 6			
Enrolled *X* American Indian/Alaska Native *X* Above 185% of poverty line	0.09	-0.03	-0.13
	(0.097)	(0.092)	(0.414)
Enrolled *X* Asian *X* Above 185% of poverty line	0.02	0.22***	-0.34***
	(0.023)	(0.022)	(0.079)

Continued on next page...

...table A.5 continued

	Registered to vote	Volunteer	Log(volunteer hours)
Enrolled *X* Black *X* Above 185% of poverty line	0.16***	0.18***	-0.35***
	(0.025)	(0.024)	(0.089)
Enrolled *X* Hispanic *X* Above 185% of poverty line	0.11***	0.17***	-0.51***
	(0.022)	(0.021)	(0.081)
Enrolled *X* More than one race *X* Above 185% of poverty line	0.13***	0.15***	-0.55***
	(0.023)	(0.023)	(0.085)
Enrolled *X* Native Hawaiian/Pacific Islander *X* Above 185% of poverty line	0.05	0.06	-0.7*
	(0.09)	(0.087)	(0.348)
Enrolled *X* White *X* Above 185% of poverty line	0.18***	0.14***	-0.6***
	(0.013)	(0.012)	(0.055)
Enrolled *X* American Indian/Alaska Native *X* Below 185% of poverty line	0.23*	0.1	-0.37
	(0.111)	(0.113)	(0.489)
Enrolled *X* Asian *X* Below 185% of poverty line	0.03	0.25***	-0.38**
	(0.035)	(0.033)	(0.118)
Enrolled *X* Black *X* Below 185% of poverty line	0.15***	0.21***	-0.25*
	(0.028)	(0.028)	(0.107)
Enrolled *X* Hispanic *X* Below 185% of poverty line	0.09***	0.1***	-0.43***
	(0.023)	(0.022)	(0.102)
Enrolled *X* More than one race *X* Below 185% of poverty line	0.19***	0.1**	-0.32*
	(0.036)	(0.034)	(0.149)
Enrolled *X* Native Hawaiian/Pacific Islander *X* Below 185% of poverty line	0.23	-0.05	1.89
	(0.196)	(0.176)	(1.089)
Enrolled *X* White *X* Below 185% of poverty line	0.16***	0.13***	-0.46***
	(0.019)	(0.019)	(0.079)

Note. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Primary point estimates from linear probability models (LPM) and ordinary least squares (OLS) regressions are shown, with standard errors in parentheses. All models include indicators for gender, race/ethnicity, and poverty status (under 185% federal poverty line) as well as controls for base year socioeconomic status and region.

Table A.6: Test calibration statistics for each propensity forest fit

Model	Mean forest prediction	Differential forest prediction
Registered to vote	1.016***	0.544
Variable importance subset	(0.1309)	(0.82)
Positive	1.018***	-0.117
	(0.1318)	(0.7064)
50th quantile	1.003***	0.201
	(0.1299)	(0.682)
80th quantile	1.006***	-0.489
	(0.1253)	(0.6871)
90th quantile	1.011***	-0.517
	(0.0976)	(0.5731)
95th quantile	1.003***	0.352
	(0.0965)	(0.4375)
Volunteered	0.993***	1.95**
Variable importance subset	(0.1588)	(0.7093)
Positive	1.011***	1.374*
	(0.1601)	(0.6189)
50th quantile	0.981***	1.995***
	(0.1593)	(0.6028)
80th quantile	0.988***	1.833***
	(0.1591)	(0.59)
90th quantile	0.949***	2.044***
	(0.146)	(0.4725)
95th quantile	0.95***	1.648***
	(0.1432)	(0.3531)
Log(volunteer hours)	0.975***	-3.295
Variable importance subset	(0.1889)	(1.2418)
Positive	0.972***	-2.2
	(0.1902)	(1.0611)
50th quantile	0.98***	-1.044
	(0.2002)	(1.0712)
80th quantile	0.991***	-1.587
	(0.1987)	(0.9502)
90th quantile	0.971***	-1.116
	(0.2134)	(0.8696)
95th quantile	0.889***	-0.78
	(0.2159)	(0.7994)

Note. Bold rows represent propensity forests fit using all variables. Rows under each model represent models run with only most important variables that fall within the cut point (any positive value or at/above quantile level of importance). A significant mean forest prediction estimate of 1 offers evidence that the mean forest prediction is correct; a differential forest prediction estimate of 1 or greater suggests the predictions also capture any underlying heterogeneity. The p-value of the differential forest prediction can be understood as test of underlying heterogeneity against a null hypothesis of no heterogeneity. See `grf::test_calibration()` help file: https://grf-labs.github.io/grf/reference/test_calibration.html