Who Matches? Propensity Scores and Bias in the Causal Effects of Education on Participation*

John Henderson† & Sara Chatfield‡

Travers Department of Political Science
University of California, Berkeley

April 24, 2009

Abstract

Matching is increasingly being used in political science to reduce selection bias in casual estimation. In a recent study, Cindy Kam and Carl Palmer employ propensity score matching to assess the decades-long consensus that college attendance increases political participation. The authors find no effect, upending a major pillar in social science. While we agree that selection is a serious concern, we argue it is so substantial in the choice to attend college that good matches may be very difficult to obtain. In this situation, propensity score matching may actually increase bias. We match on nearly 800,000 propensity scores and use genetic matching to recover better matches with lower covariate 'imbalances'. We find even modest improvements reverses the null finding, however, no matching approach yields unbiased estimates. Ultimately, we show that 'balance' in the covariates and robustness to sensitivity diagnostics should guide the matching enterprise.

Key words: causal inference, matching, propensity scores, college education, political participation.

*We thank Jas Sekhon, Rocio Titiunik, Eric Schickler, Henry Brady, Claudine Gay, and Erin Hartman for reading over countless drafts and providing invaluable advice over the course of this project. We also thank Cindy Kam for generously making code and data available.

†Ph.D. Candidate, Travers Department of Political Science, University of California, Berkeley. Email: jahenderson@berkeley.edu. Web: http://www.jahenderson.com

‡Ph.D. Candidate, Travers Department of Political Science, University of California, Berkeley. Email: schatfield@calmail.berkeley.edu

1 Introduction

Mass participation is the mainspring of politics in modern democracies. Its ebbs and flows determine the outcomes of elections, structure the outlines of legislative conflict, and influence the direction of national policy. At its highest peaks, citizen participation may bolster the strength of democratic government, but in its lowest valleys, it may expose democratic institutions to crises of legitimacy and stability. Understanding why some people participate when others opt out, therefore, is fundamental to both the study and practice of politics. Yet, after more than fifty years of study, we still possess more well-reasoned conjectures than we do well-supported conclusions about what in fact *causes* political participation.

One of the most robust findings in political science is that going to college is strongly associated with greater participation later in life. According to a typical view, college attendance confers long-lasting participatory returns, including cognitive, economic, and social resources that reduce the costs and increase the benefits to political activity (Jennings and Stoker 2008; Schlozman 2002; Miller and Shanks 1996; Brady, Verba, and Schlozman 1995a; 1995b; Jackson 1995; Rosenstone and Hansen 1993; Jennings and Niemi 1981; Wolfinger and Rosenstone 1980; Campbell et al. 1960).[1] Indeed, the illumination of an 'education effect' may constitute one of the major contributions of political science to the general body of knowledge (Schlozman 2002).

Kam and Palmer (2008) launch a serious and timely challenge to this consensus. The authors reject the claim that college attendance *causally* impacts participation, and

[1] A comprehensive list of scholarship that finds an effect of education on participation would fill volumes. To get a sense of the scope of this work, an abridged set of important citations would include: Jennings and Stoker 2008; Hillygus 2005; Schlozman 2002; Galston 2001; Brady, Schlozman, and Verba 1999; Niemi and Junn 1998; Jackson 1996; Miller and Shanks 1996; Nie, Junn, and Stehlik-Barry 1996; Brady, Verba, and Schlozman 1995a; 1995b; Jackson 1995; Leighley 1995; Rosenstone and Hansen 1993; Verba, Schlozman Brady, and Nie, 1993; Cox and Munger 1989; Dalton 1988; Abramson and Aldrich 1982; Erikson 1981; Jennings and Niemi 1981; Wolfinger and Rosenstone 1980; Brody 1978; Ashenfelter and Kelley 1975; Converse 1972; Verba and Nie 1972; Ranney and Epstein 1966; Almond and Verba 1963; Campbell et al. 1960.

argue that college is really a *proxy* for the many preceding life experiences that drive both educational attainment and political engagement. At issue is the fact that selection into college is nonrandom, a problem that previous studies have be unable to resolve through the use of traditional statistical approaches.

Kam and Palmer take advantage of some recent developments in matching theory and technology, which may give promise to the causal enterprise in non-experimental settings (Rubin 2006; Rosenbaum 2002; Neyman 1990). They employ a propensity score matching design to assess the effect of education on participation in the now classic Youth-Parent Socialization Panel Study (Y-PSPS) (Jennings and Niemi 1981). Notably, their approach potentially provides a way to account for selection confounders using an innovative method and some of the best observational data available for this question. In a major reversal from prior results, the authors find no education effect at all after matching on their estimated propensity score.

In spite of their results, we believe that the question is far from settled. We agree with Kam and Palmer that selection is a central obstacle in causal inference, especially when using non-experimental data. However, selection may be so strong and other unobserved confounders so numerous that it may be practically impossible to recover unbiased causal estimates through even the most sophisticated matching methods as yet available. In particular, we argue that the propensity score matching approach is not well-suited to using this data to study this question.

To build our case, we explore two prominent matching approaches, *propensity score matching* and *genetic matching*, to investigate how well each does in addressing the selection problems inherent in the political returns to education. Ultimately, we show that Kam and Palmer's estimate is highly sensitive to choices in propensity score model specification and subject to serious shortcomings in 'balance' diagnostics. Throughout, we emphasize that the basis for judging any matching approach is grounded in its ability

to obtain matches so that treatment and control groups are essentially similar on all relevant covariates (i.e. the experimental benchmark). In the absence of such balance, it is unlikely that the matching estimate will be free from bias. Also, if an estimate fails well-grounded sensitivity tests, it is generally unwise to interpret that estimate as causal. Every matching estimator we examine in this undertaking fails on both accounts.

In the sections below, we first discuss the key mechanisms through which college education is thought to influence political participation and the basic challenge posed by Kam and Palmer (2008) to this literature. Next, we explore the propensity score approach taken by the authors in the context of the Neyman-Rubin framework in which matching is grounded. We test balance and sensitivity diagnostics on Kam and Palmer's propensity score estimate. Then we match on nearly 800,000 propensity scores, sampled from permutations in the variables of Kam and Palmer's model, to elicit a comparable distribution of matching estimates and balance statistics in the data. We find that the propensity score matching approach is highly sensitive to minor alterations in modeling choices and assumptions. Finally, we utilize a genetic matching algorithm to find matches that recover the best possible balance. Though there is improvement, the genetic matching estimates again fail balance and sensitivity tests. All of this leads us to conclude that a true causal estimate of the effect of college attendance on political participation remains elusive.

2 The Cost-Benefit Model of Participation and Mobilization

2.1 The Participatory Returns to College Education

Most models of political behavior start with the finding that participating is costly, and its benefits largely immaterial (Wolfinger and Rosenstone 1980; Olson 1965; Downs

1957). Where the costs seem to expand exponentially, the benefits to participation seem to shrink in kind. The probability of being the decisive voter in any election is so slim as to be effectively zero, and other forms of participation are similarly unlikely to have a tangible impact on the eventual outcome. Yet, people still participate in spite of the act's instrumental irrationality. Some culprit is moving around the costs and benefits that motivate participation, and education is a usual suspect.

With this behavioral insight in mind, scholars argue that participation occurs (a) when an individual possesses enough of the right resources to be able to 'afford' the costs of doing so, and (b) when the benefits (whether instrumental, expressive, solidarity, or purposive), outweigh those costs (Brady, Verba, and Schlozman 1995a; Wilson 1995; Wolfinger and Rosenstone 1980). In the costly participation model, college education is a wellspring of such material and immaterial resources.

Accordingly, college provides people with the ability to digest complex and abstract concepts, improving cognitive, research, and reasoning skills. These abilities facilitate the collection and consumption of political information relevant to the participatory process (Brady, Verba, and Schlozman 1995a; Rosenstone and Hansen 1993; Wolfinger and Rosenstone 1980). Greater education appears to make voting and other kinds of participation less costly relatively to the benefits, and therefore seems to encourage participatory behavior.

Formal education is also thought to transform the types of benefits that people receive through participating (Brady, Verba, and Schlozman 1995a; Rosenstone and Hansen 1993). College immerses students in an environment rich in political conversation, campus activism, and intellectual inquiry. These experiences may transmit civic values, facilitate student organization and leadership, and heighten awareness of political matters, thereby altering the prism through which people judge their participatory acts. In this way, college also encourages greater political involvement by attaching a substantial, consumptive and

non-instrumental reward to the act of participating itself (Schlozman 2002; Galston 2001; Brady, Verba, and Schlozman 1995a; Wilson 1995).

Finally, individuals participate in politics not only because they derive value from it or possess the resources to afford it, but also because political elites recruit them into the process (Schlozman 2002; Jackson 1996; Leighley 1995; Rosenstone and Hansen 1993). The well-educated are more likely to be central members of dense social networks, active in voluntary associations, and leaders in their business or community (Rosenstone and Hansen 1993). Educated people also tend to be wealthier, able to give more time and money to campaigns, and able to leverage their respective social networks to translate their political involvement into participatory windfalls. Thus, college education is both a signal for elites to identify whom to recruit and also a pool of resources that when tapped reaps wider participatory dividends for parties, candidates, and activists.[2]

2.2 Selection as Cause: Socialization and Stratification in Education and Participation Outcomes

The central challenge to the costly participation model lies not in disputing the importance of costs, benefits, or resources to the participatory act, but in contesting the underlying origins of those individual qualities and experiences that help people navigate the political arena. A person's decision to attend college and to participate in the political process rests upon a deep foundation of prior personal, familial, and social pressures that originate early in life and continue throughout pre-adulthood. By the time individuals reach college age, their past weighs heavily on their future educational and

[2]Upon our reading, the mobilization account does not necessarily require a causal link between education and turnout for the basic argument to go through, though the major authors in this vein typically assert one. As some economists have argued, most notably Arrow (1973) and Cameron and Heckman (1993), the acquisition of a college degree may simply provide an informative signal about an individual's future level of productivity. In the political arena, college may signal a certain receptivity to mobilization efforts, the possession of useful resources, or a greater degree of productivity or efficiency in mobilizing friends and colleagues. None of these effects require or imply education as cause.

participatory opportunities and outcomes (Kam and Palmer 2008, Sears and Levy 2003, Jennings and Niemi 1968). The difficulty in assessing the complex causal path that might exist between college and participation stems from the fact that the same social forces that drive educational choices may also motivate participatory actions. Indeed, college is not a cause if it merely reflects or 'proxies' rather than contributes to those things that increase participation (Kam and Palmer 2008).

The primary social forces at work in this regard are socialization and socioeconomic stratification, which have the result of differentiating people in their basic resources and attributes over time (Sears and Levy 2003; Grusky 2001; Baker and Velez 1996; Luster and McAdo 1996; Saunders 1990). Modern societies are subject to considerable inequality in social and economic status, as well as in the distribution of natural talents. Those who start life with fewer advantages face steeper challenges in forming and pursuing certain goals and beliefs, compared to those born under more favorable conditions (Grusky 2001; Saunders 1990). Moreover, in childhood, people are most inundated with and susceptible to social learning. Thus, individuals' particular socialization experiences also strongly influence their development of important skills, attitudes, and resources (Sears and Levy 2003; Baker and Velez 1996; Luster and McAdo 1996).

According to Kam and Palmer's 'education as proxy' model, the attributes at stake in these processes include an individual's (a) cognitive ability, (b) personality, (c) values and attitudes, and (d) financial or other endowments (Kam and Palmer 2008; Duncan et al. 2005; Entwisle et al. 2005; Baker and Velez 1996; Jencks et al. 1972).[3] Each of these attributes are rooted in early-life development, and motivate both education- and participation-seeking behavior. Therefore, these qualities contain many of the sources of selection pressure that frustrate causal estimation and necessitate statistical or other adjustment.[4] With that in mind, we note some key variables associated with each attribute

[3]See Kam and Palmer (2008) for a particularly superb review of these large and disparate literatures.

[4]Systematic differences between college and non-college students in these key variables provide a picture of how much selection pressure may be contributing to bias in estimation. Thus, controlling for

that may help account for the selection that arises along each dimension.

(a) *Cognitive ability*: A person's intellectual gifts determine not only their interest in pursuing more education, but also their keenness in following domestic and foreign affairs and engaging in public matters (Cacioppo et al. 1996). People who enjoy the interplay of competing ideas probably find political and academic conflict interesting, if not entertaining. **Key variables:** School GPA, Public Affairs Courses, Political Conversations, Newspaper Readership.

(b) *Personality*: Certain personality traits, especially a sense of efficacy, trust, and confidence, may lead individuals to take risks and sustain their efforts in the face of setbacks (Kam and Palmer 2008; Entwisle et al. 2005; Tomlinson-Keasey and Little 1990). Yet, if one views the world as imminently frustrating and one's efforts rather ineffectual or repeatedly insufficient, the expectation may be that participating and educating will also result in failure and thus should be avoided. **Key variables:** Trust in Government, Social Trust, External Efficacy.

(c) *Values and Attitudes*: The values and attitudes a person has towards civic engagement and higher education are strong predictors of their future investment in those activities. Parents who attach high value to educational attainment or who themselves graduated from college likely provide examples to or apply pressure on their children to do the same (Jencks et al. 1972; Jennings and Niemi 1968). Parents who transmit education-related values may also transmit their own positive (or negative) attitudes about politics and participation (Kam and Palmer 2008). **Key variables:** Participation in Student Government, Parental Education, Parental Participation, Strength of Partisanship.

(d) *Endowments*: Parents also invest time, money, and energy into helping their children develop the intellectual and practical skills, talents, and abilities that will guide them through their adult life. It is no accident that adults who care strongly about attaining intellectual, personal, or social success in life also encourage their children to do

at least these confounders is necessary to make persuasive causal inferences.

the same, or seek to create the sorts of environments that facilitate this development. **Key variables:** Parental Income, Parental Employment, Parental Homeownership, Parental Knowledge.

Kam and Palmer argue that college education contributes no additional resources relevant to the calculus of costly participation. Going to college and participating are indeed causally linked, yet as the end products of a prior set of forces that begin early in life and continue throughout adulthood. This is a sharp departure from the traditional view in political science. Notably, it implies that whatever resources accrue to an individual during her college years are unique to that person, and not to the experience itself.

However, few interventions in life are as extensive and intensive as going to college. We agree that college attendance is subject to considerable selection pressures that simultaneously affect educational and political outcomes. Yet, college may also cause greater rates of participation on top of these prior influences, for instance, by actualizing resources that might be difficult to obtain otherwise. The task at hand for scholars is (and has been) to address whether or not college has some *independent* effect on participation in addition to its role as a proxy for the confounding characteristics and experiences discussed above. We now turn to this task.

3 Neyman-Rubin Causal Framework

The Neyman-Rubin framework, when tethered to a powerful research design, enables researchers to draw persuasive causal inferences from observational data. The basic model is grounded in the experimental tradition, but recent work has shown that it unifies the underlying logic behind causal inference in observational social science research (See Sekhon 2009; 2008).[5] Noteworthy for our purposes, it develops the tools necessary to go beyond simple correlations, which is required for us to estimate the causal effect of

[5]Throughout this paper, we use the notation found in Sekhon (2008).

college education on political participation.

In the Neyman-Rubin model, causal inference is a missing data problem. Let Y_{i1} be the potential outcome for the *ith* individual if she receives treatment (e.g a drug regimen, exposure to negative campaign ads, or college attendance), and Y_{i0} if she does not. The causal effect of treatment then is: $\gamma_i = Y_{i1} - Y_{i0}$. However, Y_{i1} and Y_{i0} cannot both be observed, since we cannot both *give* and *not give* individual i the treatment. Instead, we must give the treatment to some people and not others and observe differences across the groups. How treatment is assigned, then, will determine whether or not we can conclude that any observed differences between groups are due to the treatment or to other factors.

In an experimental design, treatment is assigned randomly. In doing so, we know that individuals assigned to the treatment group ($T_i = 1$) will be drawn from the same population as those assigned to the control group ($T_i = 0$). This means that the distribution of both observed and unobserved variables in the two groups should be 'balanced' (Sekhon 2008; Rosenbaum 2002). Stated differently, randomization (and a sufficiently large sample size) ensures that people in a treatment group could be replaced with those in the control, and vice versa, without altering the distribution of each group's underlying characteristics. Since treatment assignment is the only thing that differs across the groups, any differences we observe on the outcome of interest can be attributed to treatment assignment and not to any of the other covariates.

Formally i's observed outcome is:

$$Y_i = T_i Y_{i1} + (1 - T_i) Y_{i0}.$$

A good candidate for a causal estimate is the Average Treatment Effect (ATE):

$$\gamma = E(Y_i \mid T_i = 1) - E(Y_i \mid T_i = 0).$$

Notice that this is just the difference in means of the observed outcomes between individuals in the treated group and those in the control. This is true since treatment assignment is independent of the outcomes, $\{T \perp\!\!\!\perp Y_1, Y_0\}$, and thus the distribution of the potential outcomes is the same for those receiving treatment as those receiving control.

3.1 Causal Inference in Observational Data

The choice to go to college is clearly not random and is usually based on the outcomes a person expects to receive from doing so. This means that there will almost certainly be large differences between the populations for many of the baseline covariates. Randomizing college attendance for the purpose of analysis, moreover, is practically difficult and morally undesirable. Thus, we must turn to observational methods to study its effects.

Drawing causal inferences from non-random treatments generally requires both strong assumptions about the nature of selection in the data and statistical adjustments to control for confounders. A necessary assumption is that selection into treatment depends only on X observable covariates (such as parental income and education), and on no other observable or unobservable characteristics. We can then assume, conditional on X, that the potential outcomes of receiving treatment or control are not confounded with the particular treatment assignment, or $\{T \perp\!\!\!\perp Y_1, Y_0\} \mid X$. In other words, by conditioning on the set of observable factors that drive the choice to go to college, we can assess the independent effect of college on participatory outcomes.

This is typically called the Selection on Observables Assumption (SOA), which is assumed by all causal approaches (including matching) in observational social science. SOA requires that there are no unobserved or excluded characteristics that drive selection into treatment after conditioning on X. For instance, a student's work ethic likely predicts

his scholastic success and may be related to later political activity. But work ethic is extremely difficult to survey and measure. Moreover, there is no guarantee that conditioning on X will eliminate its confounding influence. Therefore, SOA must be assumed to hold after conditioning. Though there is no direct way to measure unobservable characteristics and see if they confound causal inference, sensitivity tests can help determine whether violations of SOA are problematic to a research design. [6]

Assuming SOA holds, there are two reasonable conditional causal estimators, the Average Treatment Effect on the Treated (ATT):

$$\gamma \mid (T_i = 1) = E\{E(Y_i \mid T_i = 1, X_i) - E(Y_i \mid T_i = 0, X_i) \mid T_i = 1\}$$

And the Average Treatment Effect on the Controlled (ATC):

$$\gamma \mid (T_i = 0) = E\{E(Y_i \mid T_i = 1, X_i) - E(Y_i \mid T_i = 0, X_i) \mid T_i = 0\}$$

Each estimator is a difference in means of the outcome across treated and control groups that are sampled through conditioning on X. Generally, these estimators will not be equal since conditioning on X selection confounders for treatment may produce different samples of Y_i than when doing so for control.

Matched sampling has emerged as a popular way to compute these conditional causal estimates. This is so primarily because matching makes no functional form or distributional assumptions about the relationship between treatment and outcome, but also because interpreting its results is relatively straightforward. A simple example of exact matching illustrates this point. Suppose that individuals select into college based only

[6] A second assumption that must be made for *all* causal inference (including randomized trials, as well as observational analysis) is the Stable Unit Treatment Value Assumption (SUTVA). This assumption states that one individual's treatment assignment should have no effect on the outcome that results for another individual being assigned into treatment or control. See Sekhon (2009; 2008) and Rosenbaum (2002) for a more extensive discussion.

upon their high school GPA and their gender. One could then match a college-bound woman who earned a B-average in high school with a B-average woman who did not attend college, and observe the difference in their participation later in life. We could then do the same for the remaining female B-students, and so on, until each male and female individual is exactly matched accordingly. This process allows us to recover a causal estimate of college attendance on participation in the matched sample without ever having to construct a model (e.g. $Y_i = \alpha_i + T_i\gamma + X_i\beta + \epsilon_i$).

Reality, of course, is much more complex. A variety of factors, including categorical and continuous phenomena, influence selection into college, making exact matching of this sort impossible. In this multivariate and continuous world, two matching techniques are frequently utilized: matching on a *propensity score* and matching on *multivariate distances* (e.g. Malahanobis Distance, Genetic Matching).

3.2 Propensity Score and Genetic Matching

In propensity score matching, researchers assign a conditional probability (p-score) of receiving treatment to each individual based on the set of relevant continuous and categorical observed covariates in X. Rosenbaum and Rubin (1983) show that matching on such a propensity score will eliminate all overt bias in the sample, that is, bias that stems from observed covariates. However, this finding *requires* that we know the correct propensity score or the right model to estimate it, and thus the exact probability that each observation received treatment. In practice, we do not know this propensity to receive treatment and must instead estimate it using traditional parametric or nonparametric models, with logit regression being the most popular.[7] But, how do we know if any particular estimated propensity score model is appropriate?

[7]Most frequently, researchers match on logit predicted probabilities: $Pr(Y_i \mid T_i = 1) = \frac{exp(X_i\beta)}{1+exp(X_i\beta)}$. However, Sekhon (2009) argues that matching on the logit linear predictor ($\hat{\mu}_i = X_i\hat{\beta}$) may be better since it avoids compressing propensity scores around zero and one.

A typical answer is that a propensity score should be estimated by including all of the important covariates that may influence selection to treatment into the logit model (Kam and Palmer 2008; Ho et al. 2007; Galiani et al. 2005). However, there are no strong reasons, *a priori*, to prefer one propensity score over any other based solely on the variables included or the model used to estimate it (Sekhon 2009; Smith and Todd 2001; Deheja and Wahba 1999; LaLonde 1986). The goal of matching, in general, is to sample from the treatment and control groups to construct a matched population that resembles as closely as possible a population that could be obtained under random treatment assignment. Choosing variables that influence selection or that are theoretically relevant may be a good place to start, but the ultimate test of a good set of matches must be *balance* in the observed characteristics of treated and control groups and *robustness* to sensitivity checks.

This claim is counterintuitive and goes against the logic of parametric statistical modeling, which typically fails in the case of omitted variables. In estimating a propensity score, including too many variables, even when these correlate with treatment, can induce considerable bias in the matched samples. The reason for this is simple. The predicted probabilities in logit models tend increasingly to cluster near 0 and 1 with the inclusion of greater numbers of variables in the model, due the intercorrelations amongst the covariates that predict treatment. This typically will result in bad matches for those observations at the tail ends of the propensity score, since finding controls with high predicted propensities or treateds with low ones becomes increasingly rare. If matching is done with replacement, the consequence is that a tiny fraction of controls could be matched to a very large proportion of treateds. This may ensure balance in the propensity score, but in little else.

An alternative to propensity score matching is genetic matching, a technique developed by Sekhon (forthcoming) that uses a genetic optimizer (Sekhon and Mebane 1998)

to match individuals based on their weighted distances in multivariate space. Let i and j be individuals in a population and their distance in k-dimensional space be $d(X_i, X_j)$. Their weighted Malahanobis distance is defined as:

$$d(X_i, X_j) = \{(X_i - X_j)'(S^{-1/2})'WS^{-1/2}(X_i - X_j)\}^{1/2},$$

where W is a $k \times k$ positive definite weight matrix, with w_p (for $p = 1, 2, ..., k$) along the diagonal elements and zeros in the off-diagonal cells, and $S^{-1/2}$ is the Cholesky decomposition of S, which is the variance-covariance matrix of X.[8] If all w_p equal 1, then $d(X_i, X_j)$ is the k-dimensional distance between two observations standardized by the variance-covariance in X. If different weights are supplied to each w_p, then $d(X_i, X_j)$ is re-weighted to emphasize the distance between individuals along some covariates more than others.

A form of multivariate matching is to match treateds to controls with the smallest weighted Malahanobis distances between them, given the best choice for each covariate weight, w_p. Picking the best weights and matches, however, is an extremely difficult optimization problem. The fundamental advance in genetic matching is that it uses a genetic algorithm to find the weights that improve covariate balance when matching on re-weighted Malahanobis distances. It works by finding improvements in the most imbalanced variables, gradually making its way upwards over a sequence of successive generations.

From a practical standpoint, genetic matching tends to improve overall balance drastically as compared to propensity score matching. And unlike p-score matching, in which the researcher must manually check the balance of each model to determine whether acceptable levels have been achieved, genetic matching automatically optimizes the balance of included covariates. Although this technique does not necessarily guarantee that

[8] See Diamond and Sekhon (2005) for more details.

acceptable levels of balance will be achieved, nor does it guarantee that SOA will be satisfied, genetic matching does provide a more efficient way to search the interminable space to find matches that get us closer to the experimental benchmark.

4 Assessing the Propensity Score Matching Approach

Our first task is to examine the particular propensity score approach Kam and Palmer develop to control for selection in the Y-PSPS data. Certain features of Y-PSPS give leverage to researchers using the matching approach to make causal inferences. Y-PSPS contains data from both parent and student interviews in 1965 that establish baseline covariates, and then re-interviews with the same respondents in 1973, 1982, and 1997 that assess college attendance, political participation, and a battery of demographic, social, and attitudinal questions (Jennings et al. 2004).

Given the panel structure, we can be assured that all of the baseline covariates are 'pre-treatment', as they were all measured prior to treatment occurring (college attendance between 1965 and 1973). Moreover, baseline covariates are particularly rich in this dataset, and tap at least some aspects of those attributes (e.g. cognitive ability, personality, values and attitudes, and endowments) that likely drive selection into college. This provides a considerable amount of information that may improve the quality of our matches and the strength of our inferences. Most usefully, the richness of baseline covariates in the survey facilitates a relatively expansive assessment of how well propensity score and genetic matching approaches perform in terms of their ability to achieve balance on literally hundreds of important, underlying characteristics. Indeed, we were able to match individuals and check balance on over 80 survey instruments in our analysis. (See Appendix A for the full list.)

4.1 Replicating Kam and Palmer

We start by assessing Kam and Palmer's propensity score analysis from the 1965, 1973, and 1982 waves of Y-PSPS, making as few deviations from their research design as possible.[9] Treatment is a dichotomous measure of college attendance by the 1973 survey, and the outcome is an additive index of political participation variables taken in 1973 and 1982. Specifically, the outcome index consists of voting in the 1972 (or 1980) election, attending a campaign rally or meeting, wearing a campaign button, working on a campaign, donating money, contacting a representative, going to a demonstration, and participating in local politics.

In their analysis, Kam and Palmer perform 1 to 3 matching to estimate ATT, which takes each college-attending observation and matches it to its three nearest neighbors in the non-attending control group based on the predicted probabilities estimated in the propensity score model. The model includes the 81 covariates the authors consider most relevant to selection into college, each transformed into factors. After matching, Kam and Palmer find no effect of college attendance on participation in either 1973 or 1982 (See TABLE 2 below).

After replicating Kam and Palmer's original 81-covariate propensity score model, we recover their estimates for the outcomes in 1973 and 1982 and then test for balance in the matched population on 86 covariates that are transformed into 217 dichotomous variables.[10] Surprisingly, we find that overall balance actually gets much worse. Before

[9] We follow Kam and Palmer in eliminating students with no parental data and those who dropped out of the survey by the second wave. For those students who had interview data from both parents, we randomly selected one parent to include in our dataset.

[10] This includes the 81 covariates used in estimating the propensity score, plus dummies for HOUSEHOLD COLLEGE, WIFE'S COLLEGE, and MEDIAN INCOME, and two PARENT PARTICIPATION indices. Generally, it is wise to check balance on as many relevant covariates as are available in the data. In assessing overall balance, we restrict our diagnostics to include *all* the covariates in the survey that are theorized to influence the choice to go to college. We also report balance on the 16 *key* covariates we believe are the most important in predicting selection into college and that were the most persistently imbalanced after multiple matching iterations.

matching, 42% of the 217 dichotomized covariates are balanced at the $p \geq 0.1$ level.[11] After matching, only 22% and 28% of the covariates are similarly balanced for the 1973 and 1982 samples, respectively. This raises the concern that overt bias has actually *increased* after matching on the propensity score. Yet, balance on some covariates is more relevant than on others due to their respective importance in predicting whether someone goes to college. For instance, we might believe that parents' income may strongly influence the educational and participatory outcomes of their children, but worry less about a student's participation in high school sports. Thus, the critical question is: How well does this propensity model achieve balance on the more important baseline covariates?

[TABLE 1 about here]

While balance does improve on some important covariates after matching on Kam and Palmer's propensity score, most of the crucial ones still remain imbalanced. As TABLE 1 shows, after p-score matching, imbalances remain in both key parental (e.g. PARTICIPATION, EMPLOYMENT, EDUCATION, INCOME, and PARTY ID) and student (e.g. HIGH SCHOOL GPA, GENDER, and PARTY ID) covariates in 1973. TABLE 1 also shows the same basic pattern for 1982, with some key variables making gains, but others taking losses in balance. This covariate imbalance, especially in the most important predictors, gives us strong reasons to suspect that Kam and Palmer's resulting ATT estimates are seriously biased and that major selection confounders remain uncontrolled (and may have even gotten worse) after matching on the propensity score. In the absence of some extremely strong assumptions, we simply cannot interpret these particular null findings as the correct causal inferences of education on participation.

[11] These are t-test p-values, since the variables are all dichotomous. The $p \geq 0.1$ level is chosen out of convention and not due to any deeper justification. We should always aim for better balance, in light of the fact that our fundamental goal is to obtain matches that could be recovered if we randomized treatment. Also note that dichotomizing the covariates does not affect the relative proportion of balance achieved by the various matching methods.

Additional robustness tests may reveal how badly selection confounders are biasing Kam and Palmer's findings. These tests assess the sensitivity of the ATT estimates after dropping college and non-college observations that may be difficult to match due to their distance from possible matches in the interior (*clustering bias*) or at the extremes (*extremity bias*) of the propensity score. In logit estimation, both predicted probabilities and linear predictors tend to cluster at the tails (or potentially around other values) of the resulting distribution, especially as greater numbers of intercorrelated variables are added to the logit model. Kam and Palmer do test for *clustering bias* throughout the propensity score by applying a caliper of 0.25 standard deviations ($\sigma = 0.3512$) during matching, which would delete all college-attenders that have predicted probabilities that are are farther than ± 0.0878 from the nearest non-attenders. They also test for *extremity bias* by deleting observations with the 5% highest and 5% lowest propensity scores. Yet, these tests are insufficient as each college-attenders has *at least* one non-attender within the respective bounds.

[FIGURE 1 about here]

The problem, however, is that there may be *only* one control within this bound for a large number of the treateds to be matched to, especially for treateds near the extremes of the propensity score. FIGURE 1 displays the densities of Kam and Palmer's propensity scores for attenders and non-attenders. In examining these, we observe that clustering around 1 is so pronounced that the p-scores for the top 5% of treated observations range in value from .9998174 to .9999998. In fact, over half of all treateds have propensity scores greater than .9 and over a quarter have scores greater than .99. In contrast, only 14 controls have propensities greater than .9 (3%) and only 5 (1%) have propensities greater than .95, with the highest propensity for any control being .9889.

As a consequence of this particular pattern of clustering at the tails, an extremely small number of control observations must bear the weight of being matched to a large proportion of treated individuals. In Kam and Palmer's matched sample for 1973, the five controls with the highest predicted probabilities (.989, .982, .974, .964, and .953) are matched to about 44% of the total treated observations (respectively, 287 or 11.7%, 316 or 12.9%, 335 or 13.7%, 70 or 2.8%, and 70 or 2.8%).[12] Initially, this result should be striking. For instance, one might seriously question how likely it is that a single individual could plausibly represent a useful counterfactual for over 280 different people, much less that there exist three such people in this survey sample.

[TABLE 2 about here]

Further, three of these five controls are big outliers in terms of their participatory behavior. The mean value on the 1973 (1982) participatory outcome variable is 2.79 (3.04) for all college attendees and 1.43 (1.98) for all non-college respondents. As one might expect, for individuals who attended college and who also have predicted propensities greater than .95, the average participation value is 3.26 (3.32). Yet, for the five controls with predicted probabilities above .95, the average participation score is 4.00 (4.40). Since these five outlier controls get matched to almost half of the treateds in the data, it is no surprise that a null effect is recovered. Indeed, after dropping these five controls, a positive and significant ATT is recovered for both 1973 and 1982, as TABLE 2 illustrates.[13] This reversal in the significance of the estimate, after dropping less than 1% of the controls,

[12]This is after matching one treated to three control observations, which is roughly equal to tripling the size of the matched population. These proportions are comparable to those recovered for the 1982 matched samples.

[13]Note that the ATT estimate on the 1982 participatory outcome is slightly different than the one Kam and Palmer (2008) report. Apparently, we drop one additional treated observation from our 1982 sample, which reduces the conditional difference in means after matching from 0.15 to .071. Since the ATT we report for 1982 is more conservative, we do not believe that dropping this observation fundamentally alters the basic findings.

is a clear indication of bias due to clustering in the propensity score.

4.2 Large Sample Robustness of Propensity Score Modeling

The particular propensity score Kam and Palmer use to estimate the conditional treatment effect fails both balance diagnostics and robustness checks. Yet, does this mean we should completely cast aside the propensity score approach as a means to control selection confounders into college in the Y-PSPS data? Is it possible to find a p-score model that produces much better balance *and* is robust to perturbations in model specification?

We address these questions by performing a second set of robustness checks. We match on nearly 800,000 propensity score models, sampling from all linear combinations of the original 81-variable model, to recover a more general distribution of propensity score estimators and balance statistics. Such a distribution will give us a useful way to compare Kam and Palmer's findings to the broader set of possible propensity score results, will provide an impressionistic assessment of the utility of propensity score matching for this question in general, and will illuminate how problematic selection into treatment really is in studying the causal effects of college education.

In running these models, we again deviate as little as possible from Kam and Palmer's basic research design, changing only which combination of the 81 covariates gets included in the logit model. Thus, we still factorize each variable, and conduct 1 to 3 matching on the predicted probability to estimate ATT of college attendance on the participation outcome index. For each of the models, we record the resulting ATT estimate and pvalue, balance statistics (the proportion of the 217 dichotomized covariates balanced at the $p \geq 0.1$ level and balance on the 16 'key' variables included in TABLE 1), and propensity score distribution (the proportion of the score in the middle 50% of the distribution). Thus, these models are directly comparable to Kam and Palmer's propensity score and the other models we examine below.

Our starting point is to randomly sample from all possible combinations of the 81 variables to give us an idea of how bounded the propensity score findings will be. The challenge in doing this is that the number of all possible combinations is extremely large: $2^{81} = 2.42 \times 10^{24}$. Moreover, it is suspected that the actual number of variables included in the logit model may actually impact the range of propensity score findings. Even if it were possible, uniformly sampling from all possible combinations of the variables either would require an enormous sample size or would exclude all variable combinations at most of the discrete levels.[14]

To get around this problem, we employ a 'stratified' uniform sampling method. In this approach, we sample 10,000 variable combinations at each level taken uniformly within that level, to construct our population of propensity models. Therefore, for the outcome in 1973, we sample 10,000 combinations (or all of the combinations if less than 10,000) each taken uniformly at the $\binom{81}{1}$ level, the $\binom{81}{2}$ level, the $\binom{81}{3}$ level and so on, all the way up to the $\binom{81}{80}$ level.[15] This produces a total sample population of 766,642 propensity score models of the effect of college attendance on the 1973 political participation index.[16]

[FIGURE 2 about here]

The distributions that are produced from running these propensity models illustrate an interesting but troubling picture. In terms of overall balance, no propensity score model reaches higher than 62% covariate balance. Moreover, until balance generally rises above 50%, it seems that any possible result (positive, negative, zero) may be ob-

[14] With a sample size of 1 million and a uniform probability, the expected number of samples from the $\binom{81}{n}$ combinations is 2 or less, for all $n = (1, 2, ..., 20)$ and $n = (61, 62, ..., 81)$).

[15] The sample size of 10,000 was chosen based on computational time. Larger sample sizes would exponentially increase computer time from approximately one month of parallel processing to orders of magnitude longer.

[16] We conducted similar propensity analysis for the 1982 participation outcome. For these, we uniformly sampled from the 766,642 propensity scores to produce a population of 50,000 models. The findings here are the same as in the 1973 population, so we leave them out of our discussion.

tained, though the range of coefficients gets larger as overall balance declines. FIGURE 2 shows the relationship between levels of covariate balance achieved by the propensity score models and the corresponding ATT estimates. One observable trend is that as balance increases, the estimates tend to be increasingly bounded above zero and convergent towards one. When balance is greater than 0.55, all of the ATT estimates of college attendance on participation are positive and significant.

The results in FIGURE 2 allow us to put Kam and Palmer's propensity model in larger context. Their matching estimator (indicated by $\triangle$) appears to be an outlier. The vast majority (76.4%,) of all the propensity models produce an estimate that is positive and significant (though this is difficult to visualize). And over 98.8% of the propensity matches produce better *overall* levels of balance compared to Kam and Palmer's model. Further, thirty-two of the propensity score models (indicated by $\bullet$) actually recover monotonic improvements on *all* of the 16 key covariates, though only three of these report significant estimates. This leads us to conclude that, in terms of balance, the original model does pretty well on a few difficult, but important variables (e.g. STUDENT POLITICAL KNOWLEDGE, STUDENT COLLEGE PLANS, and PARENTAL POLITICAL PERSUASION), yet performs very poorly on the rest as compared to the other propensity models. In terms of the estimates, Kam and Palmer recover a quite rare zero effect.

However, none of the 766,642 propensity models comes close to recovering the type of balance we would expect from randomization. Thus, the critical shortcoming to this sampling approach is that it is impossible to know what matching estimates would emerge from models that could achieve balance greater than 0.62. In this sense, the distribution is censored from above. This censoring could be due to three sources: (a) by chance, not sampling the right models from the enormous set of all combinations, (b) by construction, being unable to use propensity scores to recover optimal balance, and (c) by selection, being unable to recover optimal balance regardless of the matching approach taken.

We cannot rule out that by chance we simply happened to sample all the wrong propensity models in our sampling iterations. After all, this is an astronomically small fraction of all of the possible combinations of variables we could have used. Further, we could have explored other modeling specifications (e.g. not factorizing the variables, using interactive terms). Yet, we argue that the limitations in propensity score matching and the deeply rooted problem of selection in the data are likely the biggest culprits impeding the ability to get good matches here.

FIGURE 2 also displays the association between the estimated propensity scores and the ATT estimates and levels of covariate balance that result in the matching samples. The x-axis in the bottom two graphs is the proportion of propensity scores that lie in the range between .25 and .75, which measures the level of extremity in the predicted probabilities. An interesting, but not surprising finding, is that covariate balance is optimized when about half of the predicted probabilities lie in the middle 50% of the range of propensity scores. Balance declines as more of the data is compressed either at the tails or in the middle of the distribution, but for opposite reasons. When most of the data is located in the middle range, balance is lower because matching on the propensity score is basically not controlling for *any* selection confounders in the data. Conversely, when most of the data is located in the tails, the matching estimator is highly *sensitive* to whichever controls happen to have received the highest predicted probabilities depending on the inclusion of some variables in the logit rather than others.

[FIGURE 3 about here]

Further, these results are a function of how many variables are included in the propensity score model and also of how badly selection confounds participation and education outcomes. FIGURE 3 redisplays the relationship between overall covariate balance and

the proportion of middle-valued propensities in the 766,642 models, yet stratified into quarters based on the number of variables used to estimate the probabilities. The data essentially show that as more variables initially get included in estimation, balance typically improves until a certain point, after which balance declines precipitously and the ATT coefficients take on any possible value.

In general, this might justify limiting the set of possible propensity models to just those using between 30 and 50 of the original 81 covariates. Yet there is no guarantee that any particular propensity model will actually reduce overt bias. Recall that 42% of covariates are balanced before matching. In the first quarter, 64.5% of the samples have levels of balance higher than this 'baseline'. Yet, this improvement shrinks to 58% in the second quarter and 17.1% in the third quarter, and bottoms out at 1.1% in the fourth. Although using fewer variables in estimating the propensity is actually 'better' in this narrow sense, even here roughly half of all the models produce matches that reduce overall balance from the unmatched starting point.

5 Genetic Matching to Reduce Overt Bias

Propensity score matching, as a general approach, falls short as a way to control selection confounders that affect both participatory and educational outcomes, at least in the Y-PSPS data. Yet, perhaps we can exploit the advantages afforded by genetic matching to deal with the complexity associated with selection into college. In this vein, we employ genetic matching on two different populations. First, we conduct 1 to 3 genetic matching to estimate ATT for the 1973 and 1982 outcomes in the same population defined above. In a second run, we use 1 to 1 genetic matching to estimate the ATT in a sample stratified by middle values on a parental political knowledge variable.[17] Again we check

[17]The number of neighbors, n, that are matched to each treated observation (i.e. 1 to n) will generally influence the level of covariate balance that is achieved after matching. For example, when matching on a propensity score, S, as n increases, the level of balance on the propensity score must weakly decrease,

for balance and sensitivity in each of these matched populations to compare the genetic matching estimators against the previous propensity score ones.

In the first genetically matched sample, balance in general and on important covariates makes dramatic improvements. After matching, 48% of the covariates are balanced at the $p \geq 0.1$ level. TABLE 3 shows that balance on important variables is much better for the 1973 and 1982 populations when compared to that which is achieved by matching on Kam and Palmer's propensity score. However, significant imbalances still remain.

[TABLE 3 about here]

Given these imbalances, we next match on a sample stratified by parent political knowledge to reduce some of the selection effects that may be confounding the matching results. In particular, we stratify the data by including only treated and control individuals with mid-level scores on the PARENT POLITICAL KNOWLEDGE INDEX (i.e. 2 – 4).[18] We believe that parent knowledge of political affairs is a strong predictor of both college attendance and later political participation. However, it remains one of the most poorly balanced variables in almost every matching sample investigated in this analysis. Thus, we suspect that this imbalance is a great source of bias.

Balance begins to improve even before matching, with 55% of covariates balanced simply due to the stratification. After genetic matching, we obtain balance on 53% of all of the dichotomized covariates. More importantly, as indicated in TABLE 3, some of the most persistently imbalanced key covariates make serious leaps in balance, with 63% of the 16 balanced. Looking specifically at the PARENT POLITICAL KNOWLEDGE, balance

since the distance on S between the treated observation and the nth next control must be at least as far as, and possibly farther away than the $(n-1)$th control. Thus, in the absence of a strong theoretical justification, 1 to 1 matching is most widely recommended.

[18] Each of the political knowledge indices measures how many of six politically-related questions respondents were able to answer correctly in the survey. For instance, one question asks for the number of Supreme Court justices.

on this covariate improves from 0.000 in the full sample to 0.568 after genetic matching on the stratified sample. Other stubborn covariates make similar improvements, including Student College Plans, Parental College Education, and Student GPA.

[Table 4 about here]

Although genetic matching recovers significant effects in both the stratified and unstratified samples, as seen in Table 4, imbalances between the college and non-college populations remain.[19] Furthermore, both sets of estimates fail sensitivity analyses indicating that overt and hidden bias remains a serious problem.[20] Thus, despite the consistency in matching estimators across multiple specifications and despite improvements in balance, sensitivity tests reveal the findings to be quite fragile. This leads us to reject these estimates as well, as the correct causal inference of college on participation.

6 Conclusion

College is by definition selective. Attending college demands the capacity to afford its costs, withstand its rigors, and value its purposes. Naturally then, people who are born into favorable circumstances and further develop certain attitudes, abilities, and inclinations throughout pre-adulthood will possess great advantages over others when deciding to apply and being admitted to college. It is hard to deny that these differences play a major role in deciding who attends college.

On the other hand, participation in democracy is inclusive. In the U.S., with the

[19]The top four most-used controls get matched to 95, 72, 49, and 62 treateds respectively in our 1 to 3 genetically matched population. Dropping these observations does not effect the significance of the estimate.

[20]We use the sensitivity test from Rosenbaum (2002), which examines how large the difference in the probability of receiving treatment would have to be between treated and control observations in order to change the results of a matching estimator

exception of felony restrictions, suffrage is universal for all adult citizens. And most other forms of participation, like writing letters or protesting, are open to all who reside in the country. Nonetheless, stark differences between participants and non-participants persist, and these correlate strongly with differences in college education. Certainly many of the skills, resources, and values necessary to navigate the political arena are developed before reaching college-age. But does going to college really contribute no additional resources relevant to the calculus of costly participation?

Kam and Palmer argue that it does not. In their view, college is simply a proxy that reflects the different paths people take early in life. Yet, for those who do attend, few things in life are as extensive, immersive, and transformative as going to college. Thus, college education may be subject to considerable selection pressure and may also confer some improvement in certain relevant political resources.

Adjudicating between these competing claims, however, requires causal inferences that are largely free from bias. But, as we have shown, finding persuasive causal estimates of the effect of college education on political participation is a very difficult problem. The selection pressures that pervade social life strongly influence the likelihood that people will pursue certain educational and political opportunities over others. Furthermore, the many unmeasured (and perhaps unmeasurable) characteristics that differentiate people, may also drive their unique educational and participatory paths in life. The difficulty in estimating this causal effect then lies in separating out the selection pressures and unmeasured qualities that motivate both educational attainment and political participation.

Kam and Palmer (2008) significantly advance this conversation by reigniting the debate over selection confounders in estimating the effects of education on participation. The authors apply an interesting matching method on well-structured data and recover a novel and surprising result. However, inherent limitations in the propensity score approach, namely that achieving good matches is extremely difficult and the resulting

estimates are highly fragile, lead us to conclude that this particular null finding simply cannot be supported.

Moreover, we provide considerable evidence to suggest that matching on any propensity score model to estimate college treatment effects is highly tenuous. More often than not, matching on a propensity score induces greater bias. We cannot rule out that this finding is unique to the Y-PSPS sample or its particular choice of measures.[21] Yet the limitations in matching on propensity scores remain whenever selection is an issue and the true propensity for treatment is unknown, both of which are generally true for the study of educational outcomes.

Genetic matching significantly reduces, but does not eliminate the overt bias in estimation. Thus, even the most advanced matching approaches are unable to fully resolve the problem of selection in the data. Given enough time, genetic matching will eventually converge to the lowest bias estimator, though there is no guarantee that this estimate will be free from observed and unobserved confounders (Diamond and Sekhon 2005).

Therefore, interpreting any of these proposed estimates as the correct causal inference requires strong assumptions, most significantly, that persistent imbalances in the covariates are uncorrelated with the true effects of college on participation. Such assumptions are difficult to sustain in light of our theories and expectations about college education. Indeed, the shortcomings in these sophisticated matching techniques suggest a return to research design and theory-building. Developing new empirical implications and devising novel tests may sharpen future analyses and provide greater leverage in using observational methods to reduce bias in assessing the causal effects of education on political participation.

[21] See Sekhon (2008) and Ho et al. (2007) on the bias reduction properties of matching estimators.

Appendix A: Baseline Covariates (from Kam and Palmer 2008)

Student Baseline Covariates

Cognitive Ability:
- Number of public affairs courses taken
- Frequency of reading the newspaper about politics
- Frequency of listening to the radio about politics
- Frequency of reading magazines about politics
- Amount of political conversation with family
- Amount of political conversation with friends
- Amount of political conversation with adults

External Efficacy:
- Does government listen to your opinion
- Is government crooked
- Is government wasteful
- Trust government to do what is right
- Are representatives in government smart
- Is government for everyone

Personality Characteristics:
- Does the respondent's life go as they wish
- Does the respondent have good luck
- Does the respondent finish their plans
- Does the respondent typically win arguments
- Does the respondent have strong opinions
- Is the respondent's mind easily changed
- Does the respondent trust others
- Does the respondent see others as helpful
- Can others be trusted to be fair

Civic Participation:
- Officer in a school organization
- Participation in school publications
- Participation in hobby clubs
- Participation in school subject clubs
- Participation in occupational clubs
- Participation in neighborhood clubs
- Participation in religious clubs
- Participation in youth service organizations
- Participation in miscellaneous clubs

Other:
- Strength of partisanship
- Youth knowledge index
- Respondent's plans for school next year
- Does the respondent have a phone at home
- Gender
- Race
- Student GPA

Parent Baseline Covariates

Cognitive Ability:
- Frequency of reading the newspaper about politics
- Frequency of listening to the radio about politics
- Frequency of watching TV about politics
- Frequency of reading magazines about politics

External Efficacy:
- Does government listen to your opinion
- Is government crooked
- Is government wasteful
- Trust government to do what is right
- Are representatives in government smart
- Is government for everyone

Personality Characteristics:
- Does the respondent's life go as they wish
- Does the respondent have good luck
- Does the respondent finish their plans
- Does the respondent typically win arguments
- Does the respondent have strong opinions
- Is the respondent's mind easily changed
- Does the respondent trust others
- Does the respondent see others as helpful
- Can others be trusted to be fair

Civic Participation:
- Participation in church or religious organizations
- Participation in fraternal organizations
- Participation in professional organizations
- Participation in civic organizations
- Participation in civil liberties organizations
- Participation in neighborhood clubs
- Participation in sports clubs
- Participation in informal clubs
- Participation in farm groups
- Participation in women's clubs
- Participation in miscellaneous clubs

Political Participation:
- Vote in 1964 Presidential election
- Attempt to persuade anyone in 1964 Presidential elec tion
- Attend any political rallies
- Other political acts on behalf of candidates or parties
- Member of any political clubs
- Wear a button or post any campaign signs
- Donate money to parties or candidates

Other (Parent):
- Strength of partisanship
- Parental knowledge index
- Employment status
- Level of education
- Household income
- Homeownership

Additional Covariates Added to Our Model

Kam and Palmer did not include the additional covariates that we include in genmatch. They included the education of only one parent, while we chose to include the education level of both parents.

- Did student vote in school elections
- Rural/Non-rural area
- Father's education level
- Mother's education level

Works Cited

Abramson, Paul R. and John H. Aldrich. 1982. "The Decline of Electoral Participation in America," *American Political Science Review* 76.3: 502-521.

Almond, Gabriel A. and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*, Princeton, NJ; Princeton University Press.

Arrow, Kenneth. 1973. "Higher Education as a Filter," *Journal of Public Economics* 2.3: 193-216.

Ashenfelter, Orley and Stanley Kelley, Jr. 1975. "Determinants of Participation in Presidential Elections," *The Journal of Law and Economics* 18.3: 695.

Baker, Therese and William Velez. 1996. "Access to and Opportunity in Postsecondary Education in the U.S.: A Review," *Sociology of Education* 69: 82-101.

Brady, Henry, Sidney Verba, and Kay Schlozman. 1995a. "Beyond SES: A Resource Model of Political Participation," *American Political Science Review* 89.2: 271-294.

Brady, Henry, Sidney Verba, and Kay Schlozman. 1995b. *Voice and Equality: Civic Voluntarism in American Politics*, Cambridge; Harvard University Press.

Brady, Henry, Sidney Verba, and Kay Schlozman. 1999. "Prospecting for Participants: Rational Expectations and the Recruitment of Political Activists," *American Political Science Review* 93.1: 153-168.

Brady, Henry, James Fishkin, and Robert Luskin. 2003. "Informed Public Opinion about Foreign Policy: The Uses of Deliberative Polling," *The Brookings Review* 21.3: 16-19.

Brody, Richard. 1978. "The Puzzle of Political Participation in America." In Anthony King, ed., *The New American Political System*, pp. 287-324. Washington D.C; American Enterprise Institute.

Cameron, Stephen and James Heckman. 1993. "The Nonequivalence of High School Equivalents," *Journal of Labor Economics* 11.1: 1-47.

Campbell, Angus, Philip Converse, Warren Miller, and Donald Stokes. 1960. *The American Voter*, Chicago; University of Chicago Press.

Cacioppo, John, Richard Petty, Jeffrey Feinstein, and W. Blair Jarvis. 1996. "Dispositional Differences in Cognitive Motivaton," *Psychological Bulletin* 119: 197-253.

Converse, Phillip. 1972. "Change in the American Electorate." In A. Campbell and P.E. Converse, ed., *The Human Meaning of Social Change*, pp. 263-337. New York; Russell Sage.

Cox, Gary W. and Michael Munger. 1989. "Closeness, Expenditures, and Turnout in the 1982 U.S. House Elections," *American Political Science Review* 83.1: 217-231.

Dalton, Russell J. 1988. *Citizen Politics in Western Democracies*, Chatham, NJ; Chatham House.

Dehejia, Rajeev H. and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association* 94: 1053-1062.

Diamond, Alexis and Jasjeet Sekhon. 2005. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies," Version 1.2. Working Paper. Available at http://sekhon.berkeley.edu.

Downs, Anthony. 1957. *An Economic Theory of Democracy*, New York; Harper and Row.

Duncan, Greg, Ariel Kalil, Susan Mayer. 2005. "The Apple Does Note Fall Far From the Tree," In *Unequal Chances: Family Background and Economic Success*, Samuel Bowles, Herbert Gintis, and Melissa Osbourne Groves, eds. New York; Russell Sage Foundation.

Entwisle, Doris et. al. 2005. "First Grade and Educational Attainment by Age 22: A New Story," *American Journal of Sociology* 110.5: 1458-1502.

Erikson, Robert S. 1981. "Why Do People Vote? Because They Are Registered," *American Politics Quarterly* 9.3: 259-276.

Galiani, Sebastian et. al. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality," *Journal of Political Economy* 113.1: 83-120.

Galston, William. 2001. "Political Knowledge, Political Engagement, and Civic Education," *Annual Review of Political Science* 4: 217-34.

Gerber, Alan and Donald Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment," *American Political Science Review* 94.3: 653-663.

Gerber, Alan, Donald Green, and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," *American Political Science Review* 102.1: 33-48.

Grusky, David, ed. 2001. *Social Stratification: Class, Race, and Gender in Sociological Perspective,* Boulder: Westview Press.

Hillygus, D. Sunshine. 2005. "The Missing Link: Exploring the Relationship Between Higher Education and Political Engagement," *Political Behavior* 27.1: 25-47.

Ho, Daniel E., Kosuke Imai, Gary King, Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," *Political Analysis* 15.3: 199-236.

Jackson, Robert. 1995. "Clarifying the Relationship between Education and Turnout," *American Politics Research* 23: 279-299.

Jackson, Robert. 1996. "A Reassessment of Voter Mobilization," *Political Research Quarterly* 49: 331-349.

Jencks, Christopher, Marshall Smith, Henry Acland, Mary Jo Bane. 1972. *Inequality: A Reassessment of the Effect of Family and Schooling in America*. New York; Basic Books.

Jennings, M. Kent and Richard Niemi. 1968. "The Transmission of Political Values from Parent to Child," *American Political Science Review* 62.1: 169-184.

Jennings, M. Kent and Richard Niemi. 1981. *Generations and Politics: A Panel Study of Young Adults and Their Parents*, Princeton; Princeton University Press.

Jennings, M. Kent, Gregory B. Markus, Richard G. Niemi, and Laura Stoker. 2004. "Youth-Parent Socialization Panel Study, 1965-1997: Four Waves," Ann Arbor, MI: University of Michigan, Center for Political Studies/Survey Research Center. Available at http://www.icpsr.umich.edu/.

Jennings, M. Kent and Laura Stoker. 2008. "Another and Longer Look at the Impact of Higher Education on Political Involvement and Attitudes." Paper presented at the annual meeting of the MPSA Annual National Conference, Palmer House Hotel, Hilton, Chicago, IL.

Kam, Cindy and Carl Palmer. 2008. "Reconsidering the Effects of Education on Political Participation." *Journal of Politics* 70: 612-631.

LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review* 76.4: 604-620.

Leighley, Jan. 1995. "Attitudes, Opportunities and Incentives: A Field Essay on Political Participation," *Political Research Quarterly* 48.1: 181-209.

Luster, Tom and Harriette McAdoo. 1996. "Family and Child Influences on Educational Attainment: A Secondary Analysis of the High/Scope Perry Preschool Data," *Developmental Psychology* 32.1: 26-39.

Miller, Warren, and J. Merrill Shanks. 1996. *The New American Voter*, Cambridge; Harvard University Press.

Neyman, Jerzy. 1990. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," *Statistical Science* 5.4: 465-472.

Nie, Norman, Jane Junn, and Kenneth Stehlik-Barry. 1996. *Education and Democratic Citizenship in America*, Chicago; University of Chicago Press.

Niemi, Richard and Jane Junn. 1998. *Civic Education: What Makes Students Learn*, New Haven; Yale University Press.

Olson, Mancur. 1965. *The Logic of Collective Action*, Cambridge; Cambridge University Press.

Ranney Austin and Leon D. Epstein. 1966. "The Two Electorates: Voters and Non-Voters in a Wisconsin Primary," *Journal of Politics* 28.3: 598-616.

Rosenbaum, Paul R. 2002. *Observational Studies*, 2nd ed. New York: Springer.

Rosenbaum, Paul and Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70.1: 41-55.

Rosenstone, Steven, and John Hansen. 1993. *Mobilization, Participation, and Democracy in America*, New York; Longman Publishing.

Rubin, Donald B. 2006. *Matched Sampling for Causal Effects,* Cambridge; Cambridge University Press.

Saunders, Peter. 1990. *Social Class and Stratification,* London; Routledge.

Schlozman, Kay. 2002. "Citizen Participation in America," in *Political Science: State of the Discipline*, Ira Katznelson and Helen Milner, eds., New York; W.W. Norton.

Sears, David and Sheri Levy. 2003. "Childhood and Adult Political Development," In *Oxford Handbook of Political Psychology*, David Sears, Leonie Huddy, and Robert Jarvis, eds. Oxford; Oxford University Press.

Sekhon, Jasjeet. 2008."The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods", in *The Oxford Handbook of Political Methodology,* Janet Box-Steffensmeier, Henry Brady, and David Collier, eds., Oxford; Oxford University Press.

Sekhon, Jasjeet. 2009. "Opiates for the Matches," *Annual Review of Political Science* 12.

Sekhon, Jasjeet. Forthcoming. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R," *Journal of Statistical Software*, http://sekhon.berkeley.edu.

Sekhon, Jasjeet and Walter R. Mebane. 1998. "Genetic Optimization Using Derivatives: Theory and Application to Nonlinear Models," *Political Analysis* 7: 189-213.

Smith, Jeffrey A. and Petra E. Todd. 2001. "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods," *The American Economic Review* 91.2: 112-118.

Tomlinson-Keasey, Carol and Todd Little. 1990. "Predicting Educational Attainment, Occupational Achievement, Intellectual Skill, and Personal Adjustment Among Gifted Men and Women," *Journal of Educational Psychology* 82.3: 442-455.

Verba, Sidney, Kay Lehman Schlozman, Henry Brady, and Norman H. Nie. 1993. "Citizen Activity: Who Participates? What Do They Say?" *American Political Science Review* 87.2: 303-318.

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

Wilson, James. 1995. *Political Organizations*, New York; Basic Books.

Wolfinger, Raymond, and Steven Rosenstone. 1980. *Who Votes?*, New Haven; Yale University Press.

Figure 1: KP Propensity Score Density

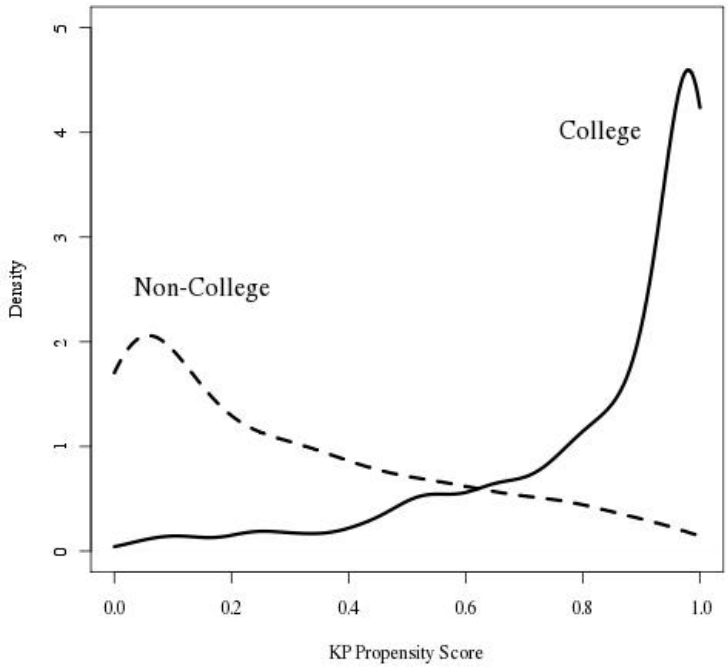


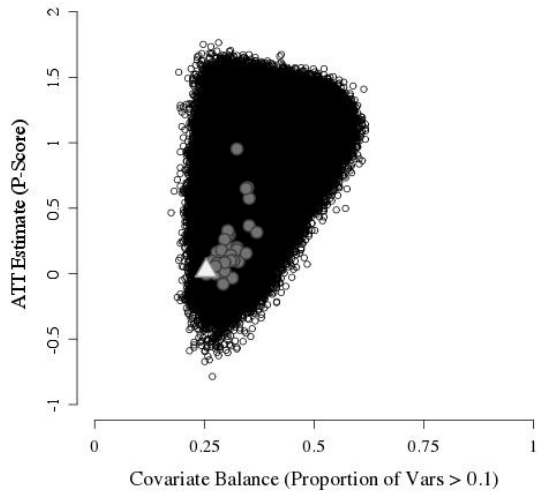
Table 1: KP Propensity Score Covariate Balance: 1973 and 1982

	Before Matching	KP P-Score: 1 to 3	
	–	1973	1982
Student GPA	0.000	0.000	0.000
Student Gender	0.000	0.000	0.000
Student Race	0.134	0.326	0.000
Student Replublican Party ID	0.000	0.000	0.016
Student Knowledge Index	0.000	0.000	0.000
Student College Plans	0.000	0.306	0.000
Parent Vote Participation	0.000	0.148	0.004
Parent Political Persuasion	0.000	0.833	0.888
Parent Participation Index	0.000	0.000	0.000
Parent Employment	0.003	0.079	0.421
Head of Household Education	0.000	0.000	0.000
Wife's Education	0.000	0.000	0.000
Parent Income	0.000	0.000	0.000
Parent Homeownership	0.000	0.000	0.508
Parent Republican Party ID	0.000	0.000	0.000
Parent Knowledge Index	0.000	0.000	0.000

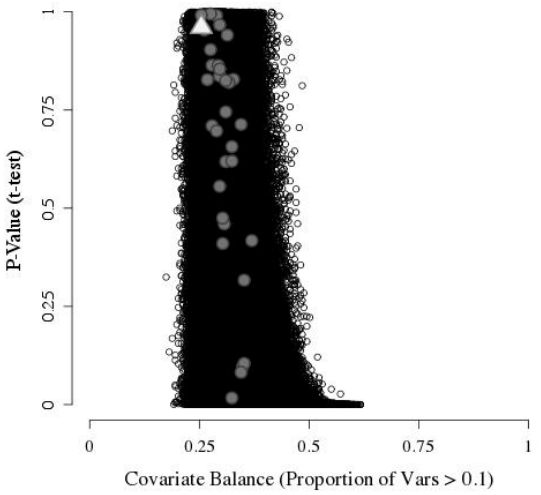
Table 2: KP Propensity Score ATT: 1973 and 1982

	KP P-Score: 1 to 3		Outliers Dropped	
	1973	1982	1973	1982
Estimate	0.023	0.071	1.340	1.105
AI Standard Error	0.469	0.452	0.519	0.565
T-Statistic	0.049	0.158	2.691	1.956
P-Value	0.961	0.875	0.007	0.051

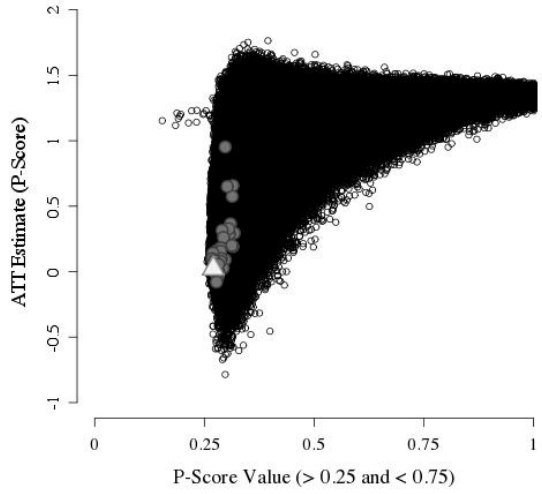
Figure 2: Propensity Score Sampling Distribution of Covariate Balance and ATT Estimate Statistics: 1973 Participation Outcome



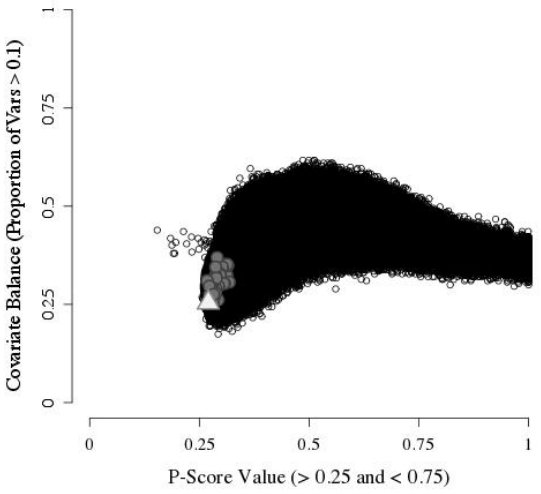
(a) Covariate Balance and ATT Convergence



(b) Covariate Balance and P-Value Convergence



(c) Middle P-Score Values and ATT Convergence



(d) Middle P-Score Values and Covariate Balance

Figure 3: Effect of Adding Variables to the Propensity Model on the Covariate Balance and Middle P-Score Distributions: 1973 Participation Outcome

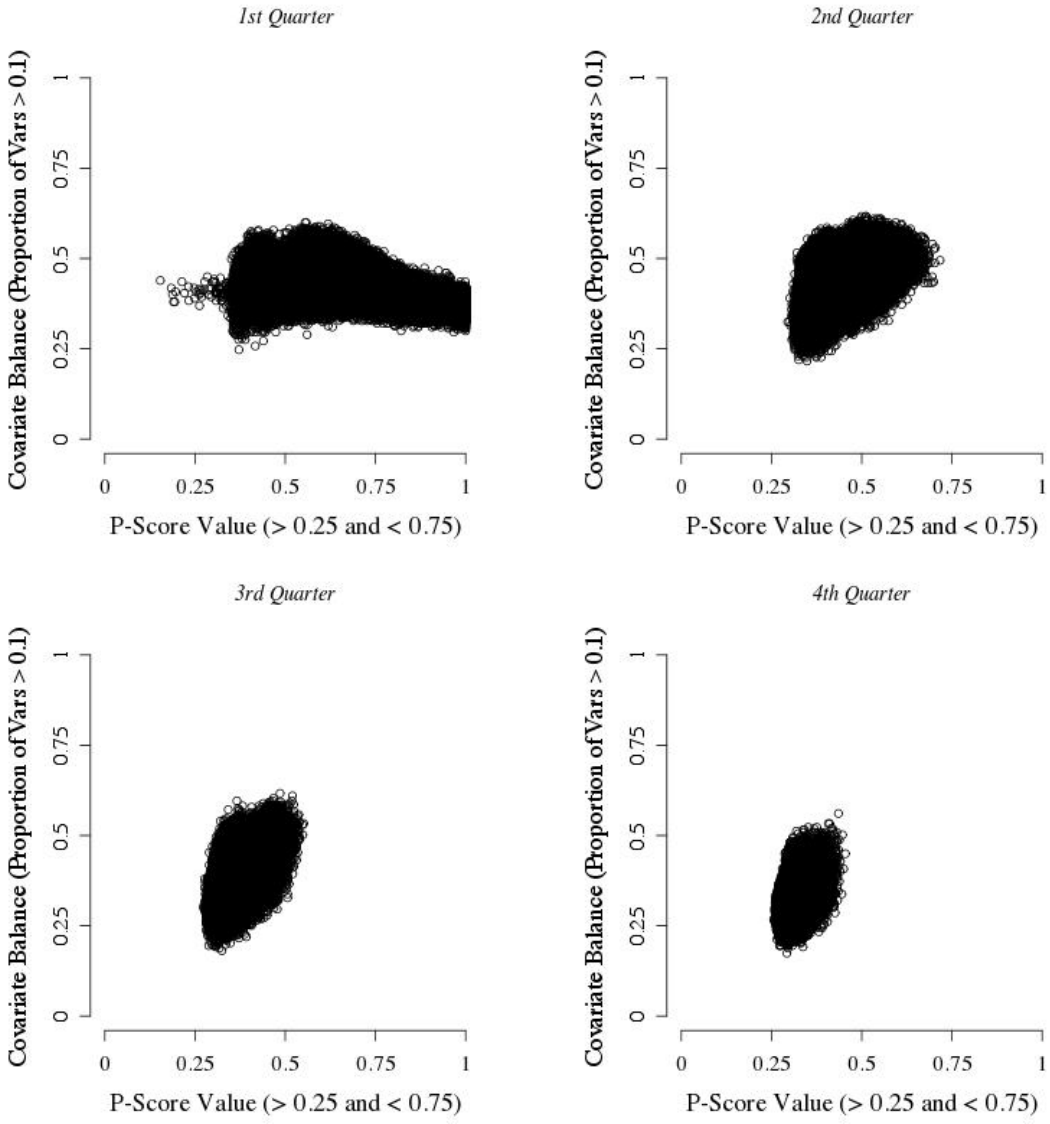


Table 3: Genetic Matching Covariate Balance: 1973 and 1982

	Before Matching	Genetic: 1 to 3		Stratified: 1 to 1	
	–	1973	1982	1973	1982
Student GPA	0.000	0.000	0.500	0.628	0.920
Student Gender	0.000	1.000	0.000	0.104	0.169
Student Race	0.134	0.177	0.216	0.115	0.492
Student Replublican Party ID	0.000	0.000	0.619	0.000	0.000
Student Knowledge Index	0.000	0.000	0.000	0.000	0.000
Student College Plans	0.000	0.564	0.001	0.157	0.157
Parent Vote Participation	0.000	0.783	0.481	0.009	0.029
Parent Political Persuasion	0.000	1.000	0.003	0.108	0.025
Parent Participation Index	0.000	0.000	0.000	0.000	0.004
Parent Employment	0.003	0.297	0.280	0.782	0.644
Head of Household Education	0.000	0.000	0.000	0.324	0.208
Wife's Education	0.000	0.000	0.000	0.156	0.082
Parent Income	0.000	0.000	0.000	0.000	0.002
Parent Homeownership	0.000	0.193	0.384	0.907	0.606
Parent Republican Party ID	0.000	0.000	0.000	0.005	0.014
Parent Knowledge Index	0.000	0.000	0.000	0.568	0.558

Table 4: Genetic Matching ATT: 1973 and 1982

	Genetic: 1 to 3		Stratified: 1 to 1	
	1973	1982	1973	1982
Estimate	1.020	0.823	0.986	0.633
AI Standard Error	0.203	0.212	0.253	0.297
T-Statistic	5.014	3.869	3.889	2.133
P-Value	0.000	0.000	0.000	0.033