Title

The Effect of Double Majoring in Undergraduate on Earnings: Using Propensity Score Matching

Author
Zeng, Zihao

Publication Date
2021-09-07

Undergraduate

The Effect of Double Majoring in Undergraduate on Earnings: Using Propensity Score Matching

Zihao Zeng
Econ 196 Thesis
Professor Shelly Lundberg
Advisor Gonzalo Vazquez-Bare

Abstract
This study provides an estimation and methodology update on previous paper that studies the effect of having a second major in undergraduate on future earnings. Using 2019 National Survey of College Graduates (NSCG) data and Propensity Score Matching (PSM) method, I find that double majoring increases earnings by around 3% for the general population, and this earnings premium is more remarkable for people under the age of 40, which amounts to about 4%. While the proportion of double majors in the population drops from over 20% in 2003 to slightly above 13% in 2019, the returns to double majoring increase from around 2.5% since 2003. I also compare results from OLS regressions and PSM and argue that PSM can relax some of the parametric assumptions imposed by OLS regressions and hence reduce misspecification and extrapolation bias from OLS regressions, which previous literature on this topic relies on.

1. Introduction

Double majoring refers to pursuing two fields of study and is one of the most important decisions undergraduate students make while in college. Del Rossi and Hersch (2008) estimate that around one quarter of college graduates have more than one undergraduate major. Students have various reasons for deciding to have more than one major. Some choose a second major because that major is their personal interest which might not necessarily be related to their first major; others double-major because they hope to satisfy their parents' expectations of studying a particular subject; while others do so because they deem that the combination of majors leads to a promising future in terms of career success. Although the payoff of double majoring to immediate pleasure or reduced parental pressure is hard to quantify, the payoff to better labor market outcomes such as earnings is measurable and can be empirically studied given the various sources of census survey data on college graduates. In addition, the understanding of how double majoring affects earnings will provide insights for educational planners and is thus a meaningful topic to study. This paper intends to provide an update on the estimations of the effect of double majoring in undergraduate on earnings from previous literature, using a novel approach Propensity Score Matching (PSM). Using the 2019 National Survey of College Students (NSCS), I find that double majoring leads to about 3% higher in earnings for the general population, and this earnings premium becomes about 4% for people younger than 40 years old. Both estimations are higher than those estimated from the 2003 NSCS data[1], indicating an increase in the double majoring earnings premium over the past decades.

[1] The 2003 data yields an about 2.5% earnings premium for double majors. Refer to the results section for a detailed discussion.

Although different institutions have various criteria and requirements for students pursuing double majors, double majors generally must satisfy major requirements for the two areas of study upon graduation. Therefore, as part of human capital investment, double majoring is likely an augmentation to that investment since it adds knowledge and skills from another discipline to a student's portfolio. For example, if a student double-majors in social science and natural science, he/she may be able to bring quantitative skills to social issues (like what economists do), as well as a sense of humanities to natural science studies. The synergies and diversity of skillsets are expected to signal competency and higher productivity in the labor market, which can place them at a higher level on the earning scale beyond graduation. On the other hand, we also doubt whether double-majored students have enough time and spared effort to engage in other activities that are as well crucial to career success while in college. Since most colleges and universities have strict limits on graduation time and double majoring often means a richer set of graduation requirements, double-majored students may spend less time than single majors in for instance doing internships, working on extracurricular lab experiments, or even studying one subject in depth. Therefore, this lack of experience and depth in knowledge at the point of graduation may be disadvantageous for double majors in the labor market. If they are unable to catch up in the early years of career, this effect would be long-lasting, and hence these students would end up earning less than single majors.

Given the rather opposite theoretical expectations about the effect of double majoring on earnings, this paper aims to provide an empirical estimation of the effect using the 2019 NSCG data. Using this dataset maintains consistency with the research by Del Rossi and Hersch (2008), who use the 2003 NSCG data and find that double majoring increases earnings by 2.3% relative to single majors among college graduates who do not earn a graduate-level degree. Since the economy has gone through several phases since 2003, it is reasonable to suspect whether the results found more than a decade ago still hold today and whether beliefs and perception about double-majored college graduates have evolved over the decade. In addition to providing an update on the findings by Del Rossi and Hersch, this paper attempts to explicitly study the outcome differences between comparable units from the double majors and single majors by employing the PSM method, which can relax some of the parametric assumptions imposed by the OLS model. Moreover, although in the data some people have second majors in a higher degree, this paper focuses on double majoring at the undergraduate level, which is more common than double majoring in other levels of study, and thus the results are more relevant to the general population. The results of this paper will provide relevant and valuable information for students contemplating on double majoring and for educational institutions who consider adopting certain policies on double majoring.

2. Literature Review

It is well understood in literature and common sense that college major has a significant impact on earnings. There is considerable research on the returns to different college majors, suggesting a great variation in expected lifetime earnings across different majors (e.g., Avery and Turner, 2012). However, to the best of my knowledge, little research has studied the returns to having more than one major. Del Rossi and Hersch (2008), using the 2003 NSCG dataset, provide the first estimates of the effect of having a second major on earnings. They find that in the general sample, double majoring increases earnings by 1.4% relative to single majors, but double majoring does not have significant effect on earnings for those with graduate degrees. The authors also examine heterogeneity in the effect due to different combinations of majors. In particular, doubling within business or education, or having two majors in two areas not closely related often leads to higher earnings compared to single majoring in respective fields. They conclude from such comparisons that double majoring does causally improve productivity and thus increases earnings to some extent; otherwise, if double majoring only served as a signal to employers, its effect on earnings would not be dramatically different across different double major combinations. However, they do not control for other pre-treatment covariates, which impact selection into double majoring and potentially cause bias in the estimation.

In a subsequent study (2016), the same authors use the 2010 NSCG to implement a cost and benefit analysis on double majoring. They specifically focus on the question of whether the benefits of studying liberal arts can be internalized so that students can be encouraged to double major in these areas so as to broaden liberal arts education in the country, but a rather unpromising answer they find is that combining a liberal arts major with another major in STEM or business results in few monetized benefits which cannot effectively incentivize students to pursue those majors. Hemelt (2010), using the same 2003 NSCG dataset, has consistent findings with Del Rossi and Hersch. He controls for undergraduate institution type, which impacts double majoring decisions. He finds a 3.2% earnings premium for double majors, and this number varies for different types of institutions, and there is insignificant effect for students studying at liberal arts colleges.

Other researchers have studied the effect of double majoring on other labor market outcomes, such as job match, job satisfaction, employability, etc. ever since the pioneering paper on this topic by Del Rossi and Hersch in 2008. Pitt and Tepper (2012) find that on average double majors are less likely to report a close relationship between their job and their major(s). Del Rossi and Hersch (2016) find positive relations between combining a liberal arts major with a business or STEM major and research and development activities and job match. Sivertsen

(2019), with the American Community Survey data, finds that on average a second major does not make it easier but in fact harder for an individual to land a job.

None of the papers mentioned above significantly take into account the general differences between single majors and double majors that may have an effect on earnings. Most of them attempt to reduce biases by controlling for demographics, institutions, and job-related characteristics, or separately performing analyses on subgroups of sample, such as by gender or by whether the individual holds a graduate degree. However, since the choice of double majoring is not randomly assigned to those with and without a second major, the inherent differences between single majors and double majors may not only influence the decision to pursue double majors but also play a role in determining future earnings. Omitting these pre-treatment covariates may cause bias in their estimations. In addition, OLS regressions used in these studies impose strong assumptions about the specification and in theory introduce extrapolation or misspecification bias. Therefore, this paper contributes to the literature by first updating the results found in previous research using a more recent dataset, and second applying PSM to this topic, explicitly comparing outcome differences between similar treated units, i.e., double majors, and untreated units, i.e., single majors. This analysis will provide more flexible estimations of the effect of double majoring on earnings by controlling for observable differences between the treatment group and the control group and estimating the treatment effect using a nonparametric model.

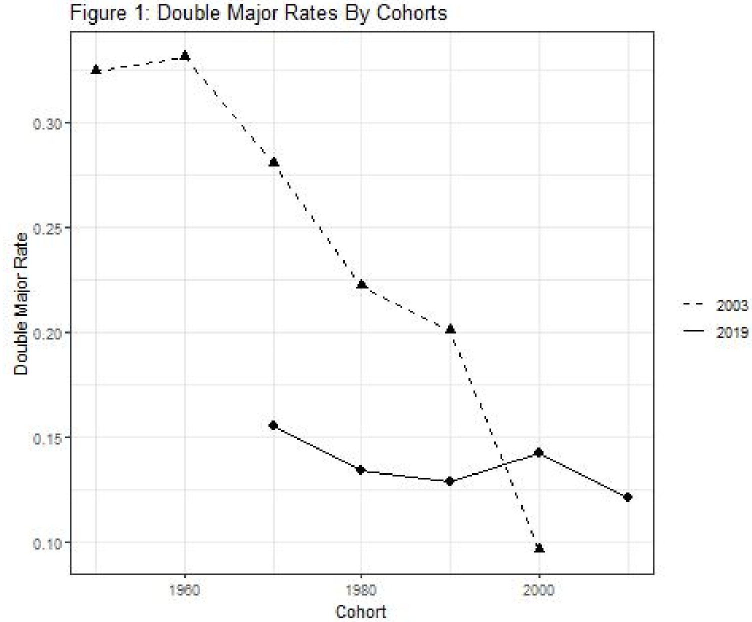
3. Data Description

I use data from the NSCG, which is the same data source many other researchers have conducted related research with. The NSCG is a biennial survey sponsored by SESTAT, the Scientists and Engineers Statistical Data System, and provides holistic characteristics of college graduates. It focuses on science and engineering but also covers all academic disciplines which allows me to extract double major information. In the 2019 cycle, it sampled approximately 147,000 individuals who are under the age of 76, earn at least a bachelor's degree, and are not institutionalized and reside in the US during the survey reference week. Some variables from the dataset that are relevant to this paper are demographics (e.g., age, race, sex, marriage status, and citizenship), educational history (e.g., highest degree, whether attended community college, etc.), employment status, field of degree (primary and secondary fields of study in the first bachelor's degree are of particular interests to this study), and employer type. This wide range of individual characteristics enables me to conduct the PSM analysis, as will be elaborated in the empirical strategy section. Since I aim to compare my results with those from previous studies, I also make use of the 2003 NSCG data, which has a similar survey design with the 2019 data.

The 2019 NSCG yielded 92,537 full observations. The key dummy variable, double major, is derived from the survey response to the question asking about the second field of study in the respondent's first bachelor's degree. If the individual does not skip this question and reports a major different from their primary major, the double major dummy variable takes value 1 to reflect that the individual has double majors; otherwise, this dummy takes value 0. The 2003 NSCG has 100,402 observations, and the double major variable in that dataset is similarly defined.

For the purpose of implementing the empirical strategy and in an attempt to estimate the effect amongst the general full-time working population, I restrict the analysis to those who have a bachelor's degree, who earn a high school diploma before they earn a bachelor's degree, who are younger than 65 years old, who are not taking courses or enrolled in a college during the survey week, whose parents' education level and private/public status of undergraduate institution are identifiable, whose earned income in the year of 2018 is not missing and between $10,000 to $1,000,000, who are classified in the labor force by the surveyor and working at least 35 hours per week during the survey week, and who have not retired before. This gives the final working sample consisting of 45,740 observations, among which 6,077 (13.29%)[2] hold a second major. A similar cleaning process is performed to the 2003 data, yielding a working sample with 53,453 observations, and 12,882 (24.10%) are double majors.

Figure 1: Double Major Rates By Cohorts



As evidenced in Figure 1, which plots the proportion of double major students in the general population by cohorts[3] and by datasets, there has been a constant decline in the popularity of double majoring. The percentage of double majors has dropped from above 30% in the 1980s to slightly above 10% in the 21st century. Aside from the overall decreasing trend, another observation from the figure is that for the 1970, 1980, and 1990 cohorts for which the

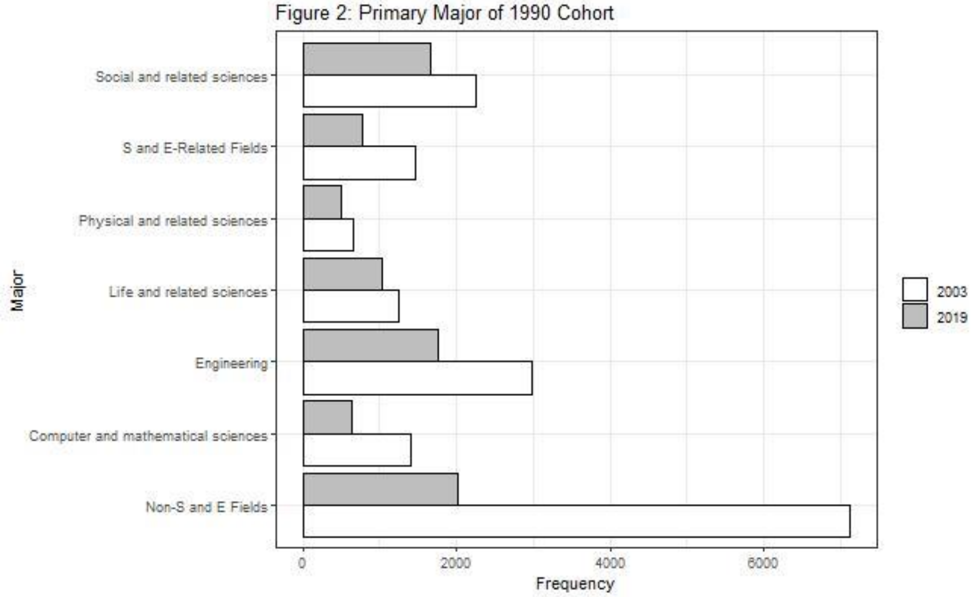
---

[2] The rates here are unweighted, i.e., calculated by dividing the number of double majors by the sample size. The sampling weights provided by the survey could have been used to calculate weighted proportions, but I decide not to present weighted proportions for the sake of consistency throughout the paper since the incorporation of sampling weights complicates the PSM analysis. And there are only small differences between the weighted and unweighted rates here.

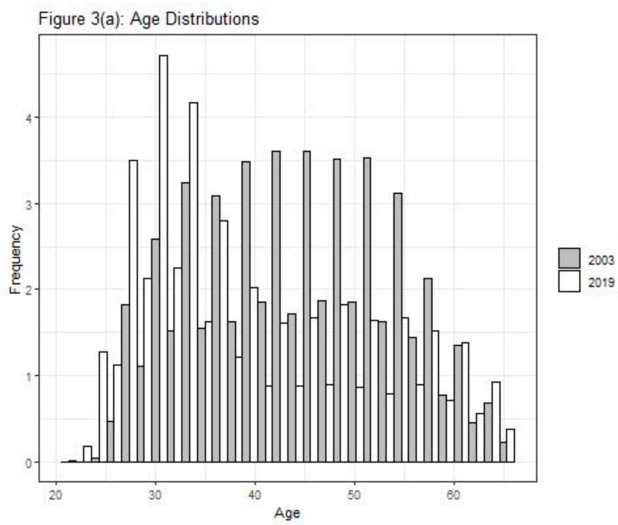
[3] Cohorts are defined as follows, the 1990 cohort, for example, refers to those who got their bachelor's degree between 1990 and 1999.

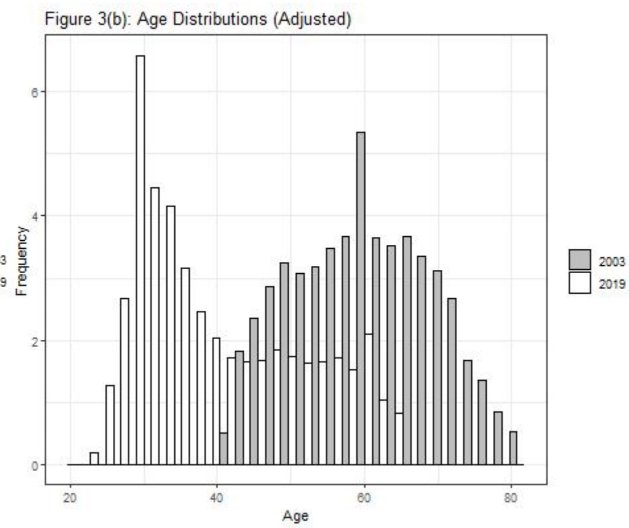
two datasets have significant overlapping observations, the percentages estimated from the 2003 data are much higher than those from the 2019 data. This is probably due to the slightly different sampling focus of the two datasets in terms of fields of study as can be seen in Figure 2. The 2003 data oversamples individuals majoring in fields that are more likely to encourage students to double major, so there is a gap between the double major rates for the two datasets. Nonetheless, a clear overall downward trend of double major rates is evident from the figure. Despite that double majoring has become less attractive, it is still a relevant topic to study because double majoring is time-consuming and amongst the most important educational investment decisions that students make in college. Knowing the implications of double majoring on expected earnings and whether it is worth the costs associated with it is useful for both students and educational institutions.

Figure 2: Primary Major of 1990 Cohort



Since I want to examine the change in the effect from 2003 to 2019, it is crucial to understand the age distributions of the individuals from the two datasets and account for the overlap in cohorts of the two datasets. As Figure 3 shows, there is a significant overlap of sample cohorts because of the large age span of the sample. When ages in both datasets are adjusted to be ages in year 2019 (Figure 3b), people who are older than 40 years old in 2019 are in the population of interests for the 2003 survey. Therefore, I decide to separately estimate the effect on cohorts under 40 years old in each dataset in addition to the full dataset. Furthermore, since changes in labor market conditions and selection into double majoring would have primary effect on new graduate cohorts, it makes sense to estimate the effect for the younger cohorts separately. The 2019 young cohort consists of 26,790 (58.57% of the full sample) observations, and the 2003 young cohort has 20,347 (38.07% of the full sample) observations. This also reflects that the NSCG survey design has shifted focus towards the younger population over the years.





4. Empirical Strategy

I. Earnings Equation and Potential Issues

To elicit the effect of double majoring on earnings, the conventional log-earnings equation with the dummy variable double major can be estimated. This is also the specification that Del Rossi and Hersch use to estimate the treatment effect.

$$\ln(Y_i) = \beta_0 + \beta_1 DM_i + \boldsymbol{\beta X_i} + \epsilon_i$$

Y_i is an individual's earned income in 2018 (or 2002, for the 2003 data), and the dependent variable is the natural log of earnings, i.e., $\ln(Y_i)$. As for the independent variables, DM_i is the double major indicator; $\boldsymbol{X_i}$ is a vector of demographic and job-related characteristics of an individual; $\beta_0,\ \beta_1, \boldsymbol{\beta}$ are the corresponding coefficients to be estimated; ϵ_i is the error term.

$\widehat{\beta_1}$ can be computed using OLS as an estimate for the effect of double majoring on earnings. However, this specification possesses noticeable omitted variable bias as it ignores those pre-treatment covariates captured by the error term and are correlated with DM_i. Some of these factors are unmeasurable, such as motivation, which affects both earnings and the double majoring decision; others are observable, such as undergraduate institution type and parents' education level. College students choose to double-major by themselves, and hence the double major treatment is not assigned randomly across individuals. Consequently, $\widehat{\beta_1}$ is subject to self-selection bias in that the double majors and single majors may be inherently different, thus leading to differences in earnings.

The key differences (both post-treatment and pre-treatment) between the treatment group and the control group of the 2019 NSCG data are shown in Table 1. Our major concerns are differences related to pre-treatment covariates. For example, 53.73% of the double majors are females, compared to 41.50% in the single major counterpart. Double majors have a higher proportion of people of minority (23.35% vs. 21.53%). In terms of their bachelor's degree, a higher proportion of double majors completed the bachelor's degree before 2005 and under the age of 22; double majors are less likely to attend community college before completing their bachelor's degree. What is rather surprising is that it does not take double majors longer to obtain a bachelor's degree after high school even though we expect that double majoring is associated with a heavier workload; in fact, the rate of completing undergraduate within 4 years after high school is much higher for double majors, being 62.46% compared to 53.62% for single majors. A possible explanation might be that double-majored students are generally more motivated and hence are more likely to complete their degree requirement on time. In addition, the undergraduate institutions of the double majors are more likely than those of the single

Table 1: Summary Statistics

Proportion	Full sample	Double major	Single major	Difference
Double major	13.29%			
Annual earning	98,793.49	100,244.68	98,571.15	1,673.54*
Degree:				
Graduate degree	45.05%	54.65%	43.58%	11.07%***
MBA	6.31%	7.34%	6.16%	1.18%***
Medical	1.40%	1.38%	1.40%	-0.02%
Law	2.10%	3.83%	1.84%	2.00%***
Associate	13.91%	11.07%	14.34%	-3.27%***
Employment:				
South	30.61%	29.74%	30.74%	-1.01%
Educational institution	17.46%	22.48%	16.70%	5.78%***
Government	14.24%	14.78%	14.15%	0.62%
Business/Industry	68.30%	62.74%	69.15%	-6.41%***
Demographics:				
Female	43.12%	53.73%	41.50%	12.23%***
Age	40.22	40.28	40.22	0.07
Minority	21.77%	23.35%	21.53%	1.82%***
Asian	11.00%	11.73%	10.89%	0.84%*
Black	6.64%	6.10%	6.72%	-0.62%*
Hispanic	10.62%	11.88%	10.43%	1.45%***
Multiple race	3.74%	4.56%	3.62%	0.94%***
US citizen	98.57%	98.32%	98.61%	-0.29%*
Have physical difficulties (any) while obtaining BA	12.14%	12.28%	12.11%	0.16%
Have physical disabilities	9.29%	9.63%	9.24%	0.38%
Completed BA before 2005	48.43%	49.78%	48.22%	1.56%**
Completed BA under 22	69.02%	74.69%	68.15%	6.54%***
Attended community college	26.07%	23.30%	26.49%	-3.19%***
Completed BA within 4 years after high school graduation	54.79%	62.42%	53.62%	8.80%***
Undergraduate institution:				
Research	43.34%	41.98%	43.55%	-1.57%**
Doctorate granting	13.50%	11.63%	13.78%	-2.15%***
Comprehensive	27.24%	26.07%	27.42%	-1.36%**
Liberal arts	11.95%	18.02%	11.02%	7.00%***
Private	31.17%	40.09%	29.80%	10.29%***
Undergraduate financing:				
Borrowed	61.28%	61.00%	61.32%	-0.32%
Assistantships or work study	18.37%	23.98%	17.51%	6.47%***
From employer	3.95%	3.31%	4.05%	-0.74%***
Parents/relatives, not to be repaid	61.10%	62.65%	60.87%	1.78%***
Tuition waivers,fellowships, grants and scholarships	51.75%	55.97%	51.11%	4.86%***
Primary major:				
Computer and mathematical sciences	8.19%	10.93%	7.78%	3.15%***
Biological, agricultural and environmental life sciences	11.70%	12.19%	11.62%	0.57%
Physical and related sciences	5.38%	6.38%	5.23%	1.16%***
Social and related sciences	21.24%	40.73%	18.25%	22.47%***
Engineering	23.44%	8.06%	25.80%	-17.74%***
SE-Related Fields	10.06%	7.06%	10.52%	-3.46%***
Non SE-Related Fields	19.98%	14.65%	20.80%	-6.16%***
Father's education level:				
Less than high school completed	7.92%	8.43%	7.85%	0.58%
High school	92.08%	91.57%	92.15%	-0.58%
Bachelors	49.86%	53.30%	49.33%	3.97%***
Graduate	24.93%	30.25%	24.12%	6.12%***
Mother's education level:				
Less than high school completed	7.22%	7.21%	7.23%	-0.02%
High school	92.78%	92.79%	92.77%	0.02%
Bachelors	43.60%	47.38%	43.02%	4.36%***
Graduate	17.88%	21.36%	17.35%	4.01%***

Note: *p<0.1; **p<0.05; ***p<0.01.

majors to be private schools and liberal arts schools. In terms of primary major, computer and mathematical sciences and social sciences are particularly popular amongst double majors, while engineering and non-science and engineering (e.g., education, arts, etc.) are more common amongst single majors. We also notice other differences between the two groups in terms of parents' education level and the financing source of undergraduate education. Given such differences between the two groups, it is reasonable to argue that $\widehat{\beta_1}$ estimated from the conventional earnings equation is confounded by and hence biased due to these differences that are not captured by the specification.

One possible solution is to include the pre-treatment covariates that affect the process of selecting into double majoring, i.e., receiving the treatment, into the specification. As a result, $\widehat{\beta_1}$ would be interpreted as the average effect of double majoring on earnings conditioning on all the pre-treatment and post-treatment covariates included by the model. This is also what Hemelt does in his analysis (although he does not include all the pre-treatment variables mentioned in the table above, partly due to the lack of information about some of these covariates in the 2003 data) by including institutional type and parents' education level into the specification. However, even if all available pre-treatment covariates are included in the regression, the issue of misspecification bias arises because the usual OLS regression imposes a strong parametric assumption on the specification model as it dictates the effect of pre-treatment variables on the outcome variable by a pre-specified equation. In addition, it extrapolates the effect of each pre-treatment variables over a range of individuals different in characteristics, adding another layer of bias to the estimation.

Therefore, I consider an alternative approach, namely PSM, which estimates the average treatment effect based on the outcomes of matched pairs that have similar characteristics from the treatment and control group and employs a non-parametric specification. The matching process is done based on the large pool of pre-treatment covariates available in the data. Although PSM would not completely solve the omitted variable bias resulting from unobservable variables which is a common challenge in the literature of educational outcome estimation, it mitigates the misspecification and extrapolation bias arising from applying OLS regressions to all individuals who differ so much in characteristics.

II. Propensity Score Matching (PSM)

PSM is one of the matching techniques that attempt to address the issue that treatment is not assigned randomly amongst individuals (Heinrich, Maffioli, and Vazquez, 2010). It matches a treated unit to one or more nontreated units that are similar over a large set of covariates so that their probabilities of receiving the treatment, i.e., the propensity scores, are similar, and hence their treatment status is “as if” random after matching. The similarity between a treated unit and a nontreated unit is determined by the distance between their propensity scores which are estimated from a participation model. After matching, the difference between the outcome of a treated unit and that of the matched unit from the control group can be used as an estimate of the treatment effect, and averaging the differences across all treatment participants, one can get an estimate of the average treatment effect.

I implement the PSM using the R package **Matching** which performs the matching process and provides standard errors estimations that account for the variance introduced by the matching procedure (Sekhon 2011), and hence it allows for the appropriate interpretation of the estimated treatment effect. Additionally, the package provides range of options on matching algorithms, which allows me to check the robustness of the estimated treatment effect by examining estimations from different matching algorithms.

The first stage of PSM involves estimating a participation model, which is to regress the treatment indicator on pre-treatment variables, and the predicted values of the dependent variable will be the propensity scores. The rich set of variables provided by the 2019 NSCG dataset enables me to control for the relevant covariates that determine the treatment. As a result, I can reasonably make the treated and control group comparable, so that the differences in earnings would be caused by the treatment of double majoring. I use the following specification[4] and run a logistic regression to estimate the propensity scores.

$$\begin{aligned} DoubleMajor_i &= \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 InstitutionType_i + \beta_4 PrivateSchool_i \\ &+ \beta_5 CommunityCollege_i + \beta_6 HSBefore2005 + \beta_7 Loan_i + \beta_8 FirstMajor_i \\ &+ \beta_9 FatherEduLevel_i + \beta_{10} MotherEduLevel_i + \beta_{11} Demographics_i + \epsilon_i \end{aligned}$$

The dependent variable in this specification is the double major indicator. Independent variables include age and age squared and other demographics including gender, races and ethnicities (consisting of 6 dummy variables, American Indian/Alaska Native, Black, Hispanic, White, Native Hawaiian/Other Pacific Islander, and Multiple race), and physical disabilities and difficulties indicators. In terms of undergraduate institution features, it includes 5 dummy variables, research university, comprehensive, liberal arts, doctorate granting, and private school. These covariates likely affect the decision of double majoring due to the varying availability of courses and requirements for graduation in different institutions. A dummy of whether an individual attended community college before getting a bachelor's degree is included because attending community college may reduce their time of enrollment in college which may affect

[4] A similar specification is used to estimate the propensity scores for the 2003 data, but the covariates of community college and loan are not available in that data, so they are left out of the participation model, and the early cohort indicator is graduating from high school before 1989.

their decision of double majoring. Since double majors seem to be more common for the older generations, the model includes an indicator of whether an individual graduated from high school before 2005. Moreover, I include primary major type (Life sciences, Physical and related sciences, Social and related sciences, Engineering, S&E-Related Fields, S&E-Related Fields) as pre-treatments, since college students usually have a first major, which is often the major at matriculation, before declaring a second major, and it is expected that their first major influences the decision to double-major[5]. Finally, socio-economic background, such as undergraduate education financing source (a dummy of whether the individual borrows to finance their undergraduate studies) and parents' education level (each parent has 6 dummies, High school, Some college, Bachelors, Masters, Professional, and Doctorate) might also play a role in double majoring decision, so they are included in the model.

This model is unexhausted as it cannot capture all the potential influencers of the double majoring decision. Firstly, unobservable variables such as motivation and abilities do not enter the equation. Secondly, some nontrivial factors that could affect the probability of having double majors and are observable are omitted because of the lack of information. One of them is the exact institution an individual attended considering that colleges and universities have different policies which impact student behaviors. However, the variables on institution features likely capture most of the institution-specific effects, which have been considered by the model. Other possible factors include citizenship while attending undergraduate, which cannot be unraveled from the data, and the detailed financing source of undergraduate education (e.g., from work

[5] There are many possible mechanisms of how the first major influences the double majoring decision. For example, engineering students may have much more demanding graduation requirements than other majors, which makes double majoring unrealistic. Some other majors tend to generate high expected payoffs if combined with another major, such as adding a mathematics major to an economics major, which makes double majoring more attractive in that case.

study, from scholarships, etc.), which likely impacts double majoring decision but is as well likely impacted by double major status, so it is deliberately left out from the equation. Despite the limitations, this participation model provides the closest form of the participation process given the availability of data.

In the second stage, the predicted values $\widehat{DoubleMajor_i}$, i.e., the propensity scores, are used to detect matched pairs between the treatment and control group based on a matching algorithm, such as nearest neighbor matching. After matching is done, the average treatment effect can be estimated by averaging the differences in log earnings between the matched pairs and the standard error necessary for result interpretation is provided by the **Matching** package. The empirical results are presented in the following section.

5. Results

I. Propensity Score Estimation

I first estimate the selection model mentioned in the empirical strategy section using a logit regression model and calculate the propensity scores on which the matching process is based. Columns 1 to 4 in Table 2 show the regression results for the selection model applied to 2019 full sample, 2019 young cohorts (younger than 40 years old), 2003 full sample, and 2003 young cohorts, respectively. Observations on who is more likely to double major can be made from the direction and significance of the coefficient estimates. For example, males are less likely than females to double-major. There is a mix of double majoring preferences amongst different races and ethnicities. Individuals who attended community college before obtaining a bachelor's degree are less likely to double-major than those did not attend community college, which aligns with the intuition. Double majoring is more common in students whose primary major is computer and mathematical science (the constant term) than students from all other

Table 2: Propensity Scores Estimation

	Double Major (Logit Regression)			
	(1)	(2)	(3)	(4)
Male	-0.261^{***} (0.030)	-0.281^{***} (0.040)	-0.167^{***} (0.022)	-0.198^{***} (0.037)
Age	-0.031^{*} (0.016)	0.135^{**} (0.064)	-0.004 (0.014)	-0.141^{*} (0.080)
Age squared	0.0004^{**} (0.0002)	-0.002^{**} (0.001)	0.0003^{**} (0.0002)	0.002^{*} (0.001)
American Indian/Alaska Native	0.135 (0.194)	0.183 (0.265)	0.565^{***} (0.134)	0.486^{**} (0.231)
Black	-0.324^{***} (0.073)	-0.361^{***} (0.094)	0.151^{**} (0.062)	0.107 (0.095)
Hispanic	0.005 (0.060)	-0.042 (0.074)	0.327^{***} (0.063)	0.297^{***} (0.091)
White	-0.034 (0.048)	-0.018 (0.058)	0.200^{***} (0.051)	0.123^{*} (0.073)
Native Hawaiian/Other Pacific Islander	-0.278 (0.275)	-0.809^{*} (0.422)	-0.194 (0.212)	-0.354 (0.299)
Multiple race	0.044 (0.080)	0.031 (0.093)	0.388^{***} (0.091)	0.321^{**} (0.139)
Physical difficulties while obtaining BA	-0.057 (0.045)	-0.061 (0.058)	0.102^{**} (0.045)	0.120 (0.085)
Physical disabilities	0.061 (0.051)	0.032 (0.078)	0.300^{***} (0.043)	0.355^{***} (0.112)
Attended community college	-0.107^{***} (0.034)	-0.119^{***} (0.044)		
HS diploma before 2005	0.155^{***} (0.058)	0.066 (0.065)		
HS diploma before 1989			-0.071 (0.053)	-0.020 (0.069)
Borrowed to finance BA	0.009 (0.030)	-0.031 (0.040)		
Primary major:				
Life sciences	-0.363^{***} (0.059)	-0.303^{***} (0.079)	-0.170^{***} (0.051)	-0.132 (0.084)
Physical and related sciences	-0.197^{***} (0.071)	-0.097 (0.097)	-0.272^{***} (0.060)	-0.186^{*} (0.102)
Social and related sciences	0.359^{***} (0.050)	0.461^{***} (0.067)	0.025 (0.046)	0.131^{*} (0.073)
Engineering	-1.440^{***} (0.064)	-1.402^{***} (0.083)	-1.043^{***} (0.051)	-1.003^{***} (0.081)
SE-Related Fields	-0.773^{***} (0.069)	-0.730^{***} (0.092)	-0.876^{***} (0.056)	-0.846^{***} (0.098)
Non-SE Fields	-0.741^{***} (0.057)	-0.601^{***} (0.077)	-0.154^{***} (0.040)	-0.034 (0.063)
Undergraduate institution type:				
Research university	0.400^{***} (0.095)	0.565^{***} (0.128)	0.151^{**} (0.063)	0.240^{**} (0.106)
Doctorate granting	0.327^{***} (0.100)	0.396^{***} (0.136)	0.297^{***} (0.066)	0.278^{**} (0.112)
Comprehensive	0.350^{***} (0.095)	0.339^{***} (0.129)	0.389^{***} (0.062)	0.295^{***} (0.106)
Liberal arts	0.530^{***} (0.098)	0.554^{***} (0.133)	0.440^{***} (0.065)	0.446^{***} (0.110)
Private school	0.288^{***} (0.034)	0.354^{***} (0.044)	0.051^{**} (0.025)	0.209^{***} (0.043)
Dad education level:				
Hish school	-0.120^{*} (0.066)	-0.277^{***} (0.097)	-0.083^{**} (0.035)	-0.071 (0.077)
Some college	-0.120^{*} (0.068)	-0.265^{***} (0.099)	0.045 (0.039)	0.012 (0.080)
Bachelors	-0.123^{*} (0.069)	-0.275^{***} (0.100)	-0.086^{**} (0.041)	-0.145^{*} (0.082)
Masters	0.020 (0.073)	-0.153 (0.104)	-0.0001 (0.048)	-0.019 (0.089)
Professional	0.040 (0.087)	-0.081 (0.121)	0.038 (0.058)	0.059 (0.108)
Doctorate	0.049 (0.088)	-0.028 (0.121)	0.062 (0.064)	0.046 (0.109)
Mom education level:				
Hish school	-0.003 (0.069)	0.273^{**} (0.107)	-0.050 (0.037)	-0.104 (0.081)
Some college	0.030 (0.071)	0.309^{***} (0.108)	0.053 (0.041)	0.076 (0.084)
Bachelors	0.091 (0.073)	0.369^{***} (0.110)	-0.038 (0.045)	0.013 (0.088)
Masters	0.121 (0.078)	0.396^{***} (0.115)	0.039 (0.054)	0.049 (0.096)
Professional	0.250^{**} (0.114)	0.508^{***} (0.150)	0.195^{*} (0.114)	0.109 (0.171)
Doctorate	0.255^{**} (0.118)	0.450^{***} (0.156)	-0.019 (0.121)	-0.021 (0.176)
Constant	-1.257^{***} (0.340)	-4.357^{***} (1.058)	-1.650^{***} (0.306)	0.611 (1.372)
Data	2019	2019	2003	2003
Age$\leq$ 40?	No	Yes	No	Yes
Observations	45,740	26,790	53,453	20,347
Log Likelihood	$-16,626.740$	$-9,738.985$	$-28,345.310$	$-9,870.844$
Akaike Inf. Crit.	33,329.470	19,553.970	56,762.610	19,813.690

Note: *p<0.1; **p<0.05; ***p<0.01
Heteroskedastic robust standard errors given in parentheses.

major backgrounds except social sciences. As is expected, undergraduate institution type plays a huge role in selection into double majoring. In particular, liberal arts college students are more likely than students from other types of institution to double-major, and individuals studying at private schools are more likely than those from public schools to double-major. These results confirm the expectation that institutional differences significantly influence double majoring decision. The impacts of parent's education level on double majoring are remarkable in terms of mother's education level and amongst the 2019 young cohorts. The more education an individual's mother receives, the higher the probability that the individual will double-major.

II. Matching

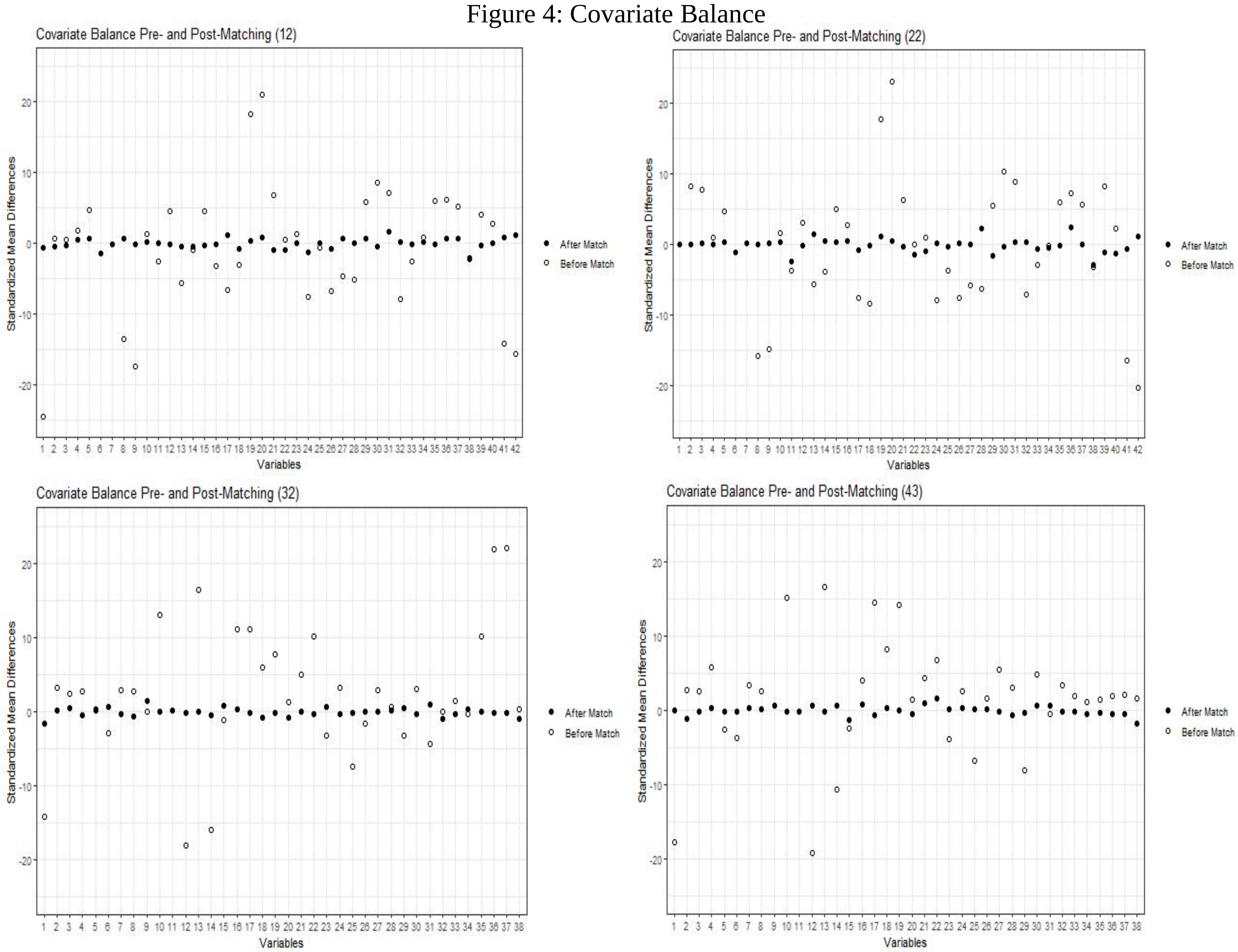
The fitted values of the participation regression models above are used as propensity scores to perform matching. The matching algorithm associated with the treatment effect estimations of the 2019 full data, 2019 young cohort data, and 2013 data presented in the next subsection is nearest 3-1 matching, i.e., each treated unit is matched to 3 closest control units in terms of propensity score, and I allow for the possibility that a control unit is matched to multiple treated units, which is a general practice in PSM studies to reduce bias. The matching algorithm[6] associated with that of the 2003 young sample makes use of the "ties" option in the R Matching package, which finds similar control units to each treated unit based on a default distance of propensity scores. As a result, the number of control units matched to each treated unit is not prespecified.

Before estimating the treatment effect, two assumptions of PSM need to be checked through matching quality assessment. The first assumption is the conditional independence assumption, which states that potential outcomes are independent of the treatment conditional on the covariates included in the participation model. This assumption is to ensure that treatment was as though randomly assigned to the treatment group and the outcome differences could hence be accredited to the treatment status. Although this assumption is untestable, we can assess covariate balance between treated and control units after matching to verify that the matching procedure successfully makes the groups comparable in observable characteristics, which can provide some empirical support for the conditional independence assumption. The pre- and post-

[6] Robustness checks on the estimated treatment effect from different algorithms are presented in section 5 IV.

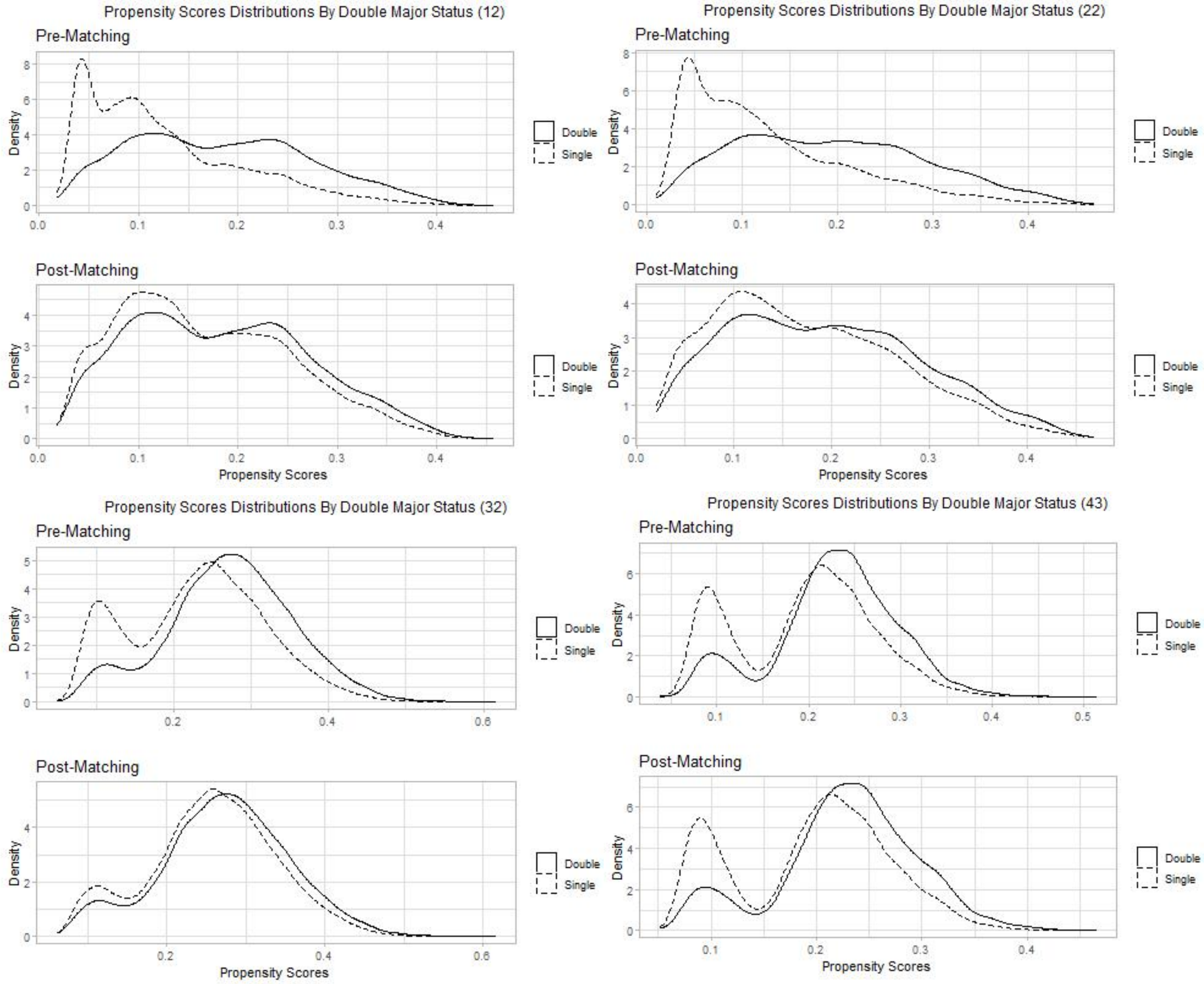
matching balance of pre-treatment covariates for the 4 data are shown in Figure 4 (2019 full data labeled 12, 2019 young data labeled 22, 2013 full data labeled 32, and 2003 young data labeled 43). White dots and black dots represent the standardized mean difference of pre-treatment covariates between the treated and control group before and after the matching procedure, respectively. As can been seen in all 4 graphs, the black dots distribute closer to the 0 horizontal line than the white dots, suggesting that the pre-treatment variables are balanced after matching and ensuring that the first assumption of PSM is reasonably satisfied.

Figure 4: Covariate Balance



Another assumption—the common support assumption—requires that the probability of each treated unit being matched to a control unit is positive so that matched pairs can be successfully generated for each treated unit. The distributions of the propensity scores by

Figure 5: Common support



treatment status before and after matching are reported in Figure 5 (same labels for the 4 data as those of the covariate balance figures). In all 4 data, the distributions of the propensity scores of the treated and control units significantly overlap. Since nearest 3-1 matching is used for the first 3 data, the propensity score distributions of the treated and control units look almost identical after matching. In the fourth data (2003 young) when the "ties" option in the Matching package is used, the propensity score distributions do not change a lot after matching. Since the sample size is large, many control units have "close" propensity scores and thus are matched to the same treated unit. This results in almost all control units being matched, so the propensity score distribution of the single majors after matching does not look very different from that before matching.

Given that the two assumptions of PSM are reasonably satisfied, we can confidently accept the validity of the estimated treatment effect from the PSM analysis.

III. Treatment Effect Estimation

The treatment effects are separately estimated for the four data, 2019 full sample, 2019 young sample, 2003 full sample, and 2003 young sample. In an attempt to sort out the difference in the effect resulting from different empirical methodologies, both OLS regressions (with varying numbers of controls) and PSM are applied to the four data. The results are presented in Table 3.

Table 3: Treatment effect estimation

Data	Working sample size	Double Major	OLS regression	OLS regression (more controls)	PSM estimation
2019 Full Data	45740	6077	0.065***	0.066***	0.031***
		(13.29%)	(0.019)	(0.018)	(0.011)
2019 Data, Age<=40	26790	3620	0.087***	0.077***	0.042***
		(13.51%)	(0.026)	(0.025)	(0.013)
2003 Full Data	53453	12882	0.019**	0.026***	0.025***
		(24.10%)	(0.008)	(0.008)	(0.007)
2003 Data, Age<=40	20347	4112	0.010	0.013	0.024**
		(20.21%)	(0.013)	(0.013)	(0.010)

*p<0.1; **p<0.05; ***p<0.01.

Restricted to young cohorts, double major rate has also dropped over the years from 20.21% to 13.51%, confirming a decrease in the popularity of double majoring. The OLS regression column corresponds to the conventional earnings equation mentioned in the empirical strategy section. This is a similar specification that Del Rossi and Hersch use, with controls for demographics (age, gender, races, and marital status), employment type and location, highest degree information, and primary major in undergraduate. The OLS regression with more controls specification adds more relevant controls to the simple OLS regression and particularly includes pre-treatment covariates to control for selection into double majoring[7]. The estimated

7 The exact specifications and results of the two sets of regressions are given in Appendix A.

coefficients on the double major indicator and the heteroskedastic standard errors are given for these two columns. The last column reports estimation from PSM. The averaged differences in log of earnings between the matched pairs and the appropriate standard errors are reported.

Using the PSM results from the 2019 data, I find that double majoring leads to a 3.1% increase in earnings, and the earnings premium of double majoring is more remarkable amongst individuals under 40 years old, which amounts to 4.2%. Both of these estimates are higher than those from the 2003 data, which are about 2.5%, so the earnings premium of double majoring has increased over the years. This trend is observable in the results from the two OLS regressions as well. The OLS regression with more controls results in a 7.7% earning premium for the young cohort in 2019 and does not find significant effect for the young cohort in 2003. Similarly, the simple OLS regression estimates that double majoring increases earnings by 8.7% for the young cohort in 2019 and does not find significant effect for the young cohort in 2003 either. The estimate for the 2003 full data is 1.9%, which is similar to the estimate by Del Rossi and Hersch[8], indicating that the OLS regression I run here replicates their approach closely.

Comparisons can be made across the different methodologies. In regard to results from 2003 young cohort sample, while OLS regressions do not generate significant results, PSM does give rise to a significant estimate, attesting to the theory that PSM provides a more flexible specification. In addition, OLS regressions overestimate the effect for the 2019 data. The magnitude of the effect estimated by PSM is not as large as that estimated by the OLS regressions, indicating that misspecification or extrapolation does introduce some degree of bias to the OLS regression estimates. Furthermore, comparing results from the OLS regressions with more controls with those from the simple OLS regressions confirms that it is crucial to consider

[8] Their estimate is 1.4%. The small difference is mainly due to the fact my OLS regression additionally controls for fields of study and the working sample is slightly different from what they use.

pre-treatment variables that influence selection into double majoring. This is especially true for the 2003 data. When pre-treatment covariates are included, the estimated effect of double majoring on earnings rises from 1.9% to 2.6%, which are significantly higher than what Del Rossi and Hersch find.

IV. Robustness Checks

This subsection examines whether different matching algorithms applied to the 4 data generate significantly different treatment effect estimates.

The Match function from the Matching package provides some options to alter the matching algorithm. Two of which are the number of control units each treated unit is matched to and whether ties should be maintained for control units close in propensity scores. Table 4 shows the results from altering these two options on the four data. The first algorithm presented is nearest 1-1 matching, i.e., each treated unit is matched to the control unit with the closest propensity score. The second algorithm is nearest 3-1 matching as described above. The default ties algorithm allows for ties, and all the tied control units are weighted according to the number of tied control units for a specific treated unit[9]. The PSM estimation for the 2003 young cohort presented in the previous subsection is the result from this algorithm. The last algorithm presented in the table does not specify the number of matched control unit either, but a distance

Table 4: Results from different matching algorithms

Data	Nearest 1-1	Nearest 3-1	Default Ties	Ties With Small Distance
2019 Full Data	0.032***	0.031***	0.031***	0.028***
	(0.011)	(0.011)	(0.009)	(0.011)
2019 Data, Age<=40	0.043***	0.042***	0.038***	0.045***
	(0.013)	(0.013)	(0.011)	(0.014)
2003 Full Data	0.023***	0.025***	0.026***	0.026***
	(0.007)	(0.007)	(0.006)	(0.007)
2003 Data, Age<=40	0.025**	0.019	0.024**	0.027**
	(0.012)	(0.012)	(0.010)	(0.012)

*p<0.1; **p<0.05; ***p<0.01.

[9] A detailed description can be found in the documentation of the Matching package.

of tolerance is provided manually to allow control units that are close enough to be matched to the same treated unit while avoiding too many control units to be considered close.

The four matching algorithms do not produce very different results for the 2019 full data, 2019 young cohort data, and 2003 full data. Specifically, an about 3% earning premium of double majoring is found in the 2019 full data, around 4% for the 2019 young cohort, and around 2.5% for the 2003 full data. The choice of matching algorithms does seem to affect the results for the 2003 young data. Although the result from nearest 3-1 matching is insignificant, the results from all other algorithms are significant at 5% level and have a magnitude of about 2.5%. Therefore, we can still confidently accept the presence of the effect in the 2003 young cohort data. We can thus conclude that the estimated treatment effect from the PSM is not significantly influenced by the choice of particular matching algorithms, so the results are robust[10].

6. Conclusion

Applying Propensity Score Matching to the 2019 National Survey of College Graduate, I find that double majoring in undergraduate creates an about 3% earnings premium for the general working population and an about 4% earnings premium for people younger than 40 years old. It testifies one of the incentives and rationales behind students deciding to double major in undergraduate, and it provides policy makers the informative empirical evidence necessary to make decisions about whether to promote double majoring in college. Moreover, the effect of double majoring on earnings has increased since 2003, while the proportion of students pursing double majors has dropped almost in half. It might be the case that the requirements for double majoring are now more demanding than they were two decades ago, so double majoring does provide students with a richer skillset that helps lead to a more prosperous career than single

[10] The matching quality of all the algorithms on all four data is presented in Appendix B.

majors. This difficulty in pursuing double majors also explains why the double major rate has dropped. Additionally, the comparison of the results from OLS regression and PSM reveals that PSM does provide a more flexible specification and thus produces more reliable estimation than OLS regressions.

Although the PSM method reduces potential misspecification and extrapolation bias present in OLS regressions, it still relies on the selection-on-observable assumption and cannot control for unobservable factors such as ability and motivation of an individual. After all, these factors are unquantifiable, and we can only get a close estimation of the treatment effect by reducing all noticeable bias from the regression specification. Therefore, one should be cautious about interpreting the average treatment effect estimate as an exact causal effect, for the effect of educational input on labor market outcomes is always convoluted and hence is hardly precisely measured by any model. Moreover, not all observable factors can be taken into account either because of the lack of information or because of the uncertainty about the exact selection into treatment process. Despite these conceivable limitations, the estimated effect of double majoring on earnings using PSM nonetheless provides a close estimation of the causal effect and is theoretically more robust than the estimation provided in previous paper.

PSM can be applied to study the effect of double majoring on other labor market outcomes such as employability, job satisfaction, position in the company, etc., which this study does not examine. A heterogeneity analysis could also be performed to study the effect separately for males and females. Other possible empirical strategies in studying the effect of double majoring on labor market outcomes might be to construct an instrumental variable based on the rate of double majoring at the undergraduate institution the students attend. The reason why this rate may be a legit instrument is that it has little bearing on the student's future earnings and only affects earnings through the process of influencing the willingness of the student to pursue double majors because students tend to emulate what their peers do. However, this requires a specialized and detailed dataset and may be hard to obtain.

Acknowledgement
I would like to thank Professor Shelly Lundberg and Gonzalo Vazquez-Bare for their incessant support for my first research in economics during this 6-month period. I really appreciate their advice on empirical strategy, comments on drafts of this paper, and intellectual stimulation during our conversions. This study cannot be done without their generous help and time commitment, and I have learnt to become an eligible researcher. Thanks again!

References

Avery, C., & Turner, S. (2012). Student Loans: Do College Students Borrow Too Much--Or Not Enough?. *Journal of Economic Perspectives*, 26(1), 165-92.

Del Rossi, A. F., & Hersch, J. (2016). The private and social benefits of double majors. *Journal of Benefit-Cost Analysis*, 7(2), 292-325.

Del Rossi, A. F., & Hersch, J. (2008). Double your major, double your return?. *Economics of Education Review*, 27(4), 375-386.

Heinrich, C., Maffioli, A., & Vazquez, G. (2010). A primer for applying propensity-score matching. *Inter-American Development Bank*.

Hemelt, S. W. (2010). The college double major and subsequent earnings. *Education Economics*, 18(2), 167-189.

Pitt, Richard N. & Tepper, Steven A. (2012). Double Majors: Influences, Identities and Impacts. Prepared for the Teagle Foundation, Curb Center, Vanderbilt University.

Sekhon, J. S. (2008). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software, Forthcoming*.

Sivertsen, D. C. (2019). Does a Second Major Effect Employment?.

Appendix A

Table 5: OLS regression results for 2019 data

	log(Earnings)			
	(1)	(2)	(3)	(4)
Double major	0.065***	0.087***	0.066***	0.077***
	(0.019)	(0.026)	(0.018)	(0.025)
Another bachelors degree	0.005	0.028	−0.007	0.011
	(0.030)	(0.039)	(0.030)	(0.037)
Graduate degree:				
MBA	0.351***	0.311***	0.309***	0.259***
	(0.028)	(0.044)	(0.027)	(0.042)
Other masters	0.115***	0.075***	0.090***	0.049**
	(0.014)	(0.020)	(0.014)	(0.019)
Doctorate	0.367***	0.264***	0.298***	0.182***
	(0.022)	(0.033)	(0.022)	(0.033)
Medical degree	0.715***	0.546***	0.632***	0.457***
	(0.040)	(0.057)	(0.041)	(0.059)
Law degree	0.482***	0.365***	0.393***	0.262***
	(0.033)	(0.049)	(0.034)	(0.047)
Other professional degree	0.137	−0.083	0.066	−0.158
	(0.113)	(0.154)	(0.113)	(0.161)
Employment:				
Government	0.183***	0.164***	0.185***	0.170***
	(0.020)	(0.031)	(0.020)	(0.030)
Business/Industry	0.294***	0.242***	0.277***	0.229***
	(0.016)	(0.023)	(0.015)	(0.022)
South	−0.050***	−0.057***	−0.049***	−0.057***
	(0.014)	(0.020)	(0.013)	(0.020)
Male	0.067***	0.044	0.057**	0.043
	(0.026)	(0.034)	(0.025)	(0.033)
Age	0.077***	0.088***	0.073***	0.099***
	(0.005)	(0.029)	(0.007)	(0.028)
Age squared	−0.001***	−0.001*	−0.001***	−0.001**
	(0.0001)	(0.0004)	(0.0001)	(0.0004)
Married	0.090***	0.085***	0.085***	0.090***
	(0.019)	(0.028)	(0.018)	(0.026)
Male * Married	0.182***	0.144***	0.175***	0.131***
	(0.029)	(0.039)	(0.028)	(0.037)
Veteran	−0.052*	−0.026	−0.008	0.025
	(0.027)	(0.040)	(0.027)	(0.037)
American Indian/Alaska Native	−0.492**	−0.744***	−0.459**	−0.703***
	(0.196)	(0.161)	(0.181)	(0.157)
Black	−0.238***	−0.248***	−0.187***	−0.179***
	(0.032)	(0.044)	(0.032)	(0.044)
Hispanic	−0.166***	−0.199***	−0.125***	−0.148***
	(0.029)	(0.033)	(0.029)	(0.033)
White	−0.105***	−0.113***	−0.101***	−0.096***
	(0.022)	(0.023)	(0.023)	(0.026)
Native Hawaiian/Other Pacific Islander	−0.119	−0.076	−0.105	−0.037

	(0.078)	(0.161)	(0.073)	(0.141)
Multiple race	−0.115**	−0.158***	−0.107**	−0.142**
	(0.049)	(0.059)	(0.049)	(0.060)
Primary major:				
Life sciences	−0.223***	−0.319***	−0.229***	−0.307***
	(0.028)	(0.041)	(0.028)	(0.038)
Physical and related sciences	−0.164***	−0.322***	−0.164***	−0.312***
	(0.031)	(0.039)	(0.031)	(0.038)
Social and related sciences	−0.205***	−0.297***	−0.213***	−0.295***
	(0.021)	(0.029)	(0.021)	(0.027)
Engineering	0.106***	0.057**	0.071***	0.021
	(0.020)	(0.026)	(0.020)	(0.026)
SE-Related Fields	−0.130***	−0.224***	−0.115***	−0.204***
	(0.022)	(0.031)	(0.022)	(0.032)
Non-SE Fields	−0.202***	−0.301***	−0.187***	−0.274***
	(0.021)	(0.029)	(0.021)	(0.028)
Associate degree			−0.075***	−0.046*
			(0.018)	(0.027)
Undergraduate institution:				
Research			0.249***	0.257***
			(0.031)	(0.040)
Doctorate granting			0.191***	0.181***
			(0.031)	(0.041)
Comprehensive			0.136***	0.135***
			(0.030)	(0.039)
Liberal arts			0.034	0.012
			(0.033)	(0.044)
Private school			0.091***	0.114***
			(0.015)	(0.022)
Citizen			0.097*	0.056
			(0.049)	(0.044)
Physical difficulties while obtaining BA			−0.061***	−0.093***
			(0.019)	(0.025)
Physical disabilities			−0.071***	−0.079**
			(0.021)	(0.034)
Attended community college			−0.046***	−0.009
			(0.016)	(0.024)
Attended community college before bachelors			0.022	−0.010
			(0.018)	(0.026)
HS diploma before 2005			0.043	0.023
			(0.028)	(0.032)
Financial support for BA: employer			0.106***	0.122***
			(0.026)	(0.039)
Borrowed to finance BA			−0.070***	−0.102***
Father's education level:				
			(0.013)	(0.020)
High school			−0.047*	−0.062
			(0.026)	(0.044)
Some college			−0.007	−0.003
			(0.027)	(0.044)

Bachelors			0.016	−0.005
			(0.027)	(0.045)
Masters			0.041	0.008
			(0.030)	(0.047)
Professional			0.099**	0.084
			(0.039)	(0.054)
Doctorate			0.006	−0.002
			(0.044)	(0.070)
Mother's education level:				
High school			0.023	0.032
			(0.027)	(0.048)
Some college			0.047*	0.080
			(0.028)	(0.049)
Bachelors			0.035	0.056
			(0.029)	(0.050)
Masters			0.042	0.094*
			(0.032)	(0.053)
Professional			0.001	0.031
			(0.055)	(0.072)
Doctorate			0.017	0.056
			(0.055)	(0.082)
Constant	9.162***	9.109***	9.002***	8.730***
	(0.108)	(0.456)	(0.157)	(0.453)
Age≤40?	No	Yes	No	Yes
Observations	45,740	26,790	45,740	26,790
R^2	0.328	0.309	0.359	0.353
Adjusted R^2	0.328	0.308	0.358	0.351

Note: *p<0.1; **p<0.05; ***p<0.01
Heteroskedastic robust standard errors given in parentheses.

Table 6: OLS regression results for 2003 data

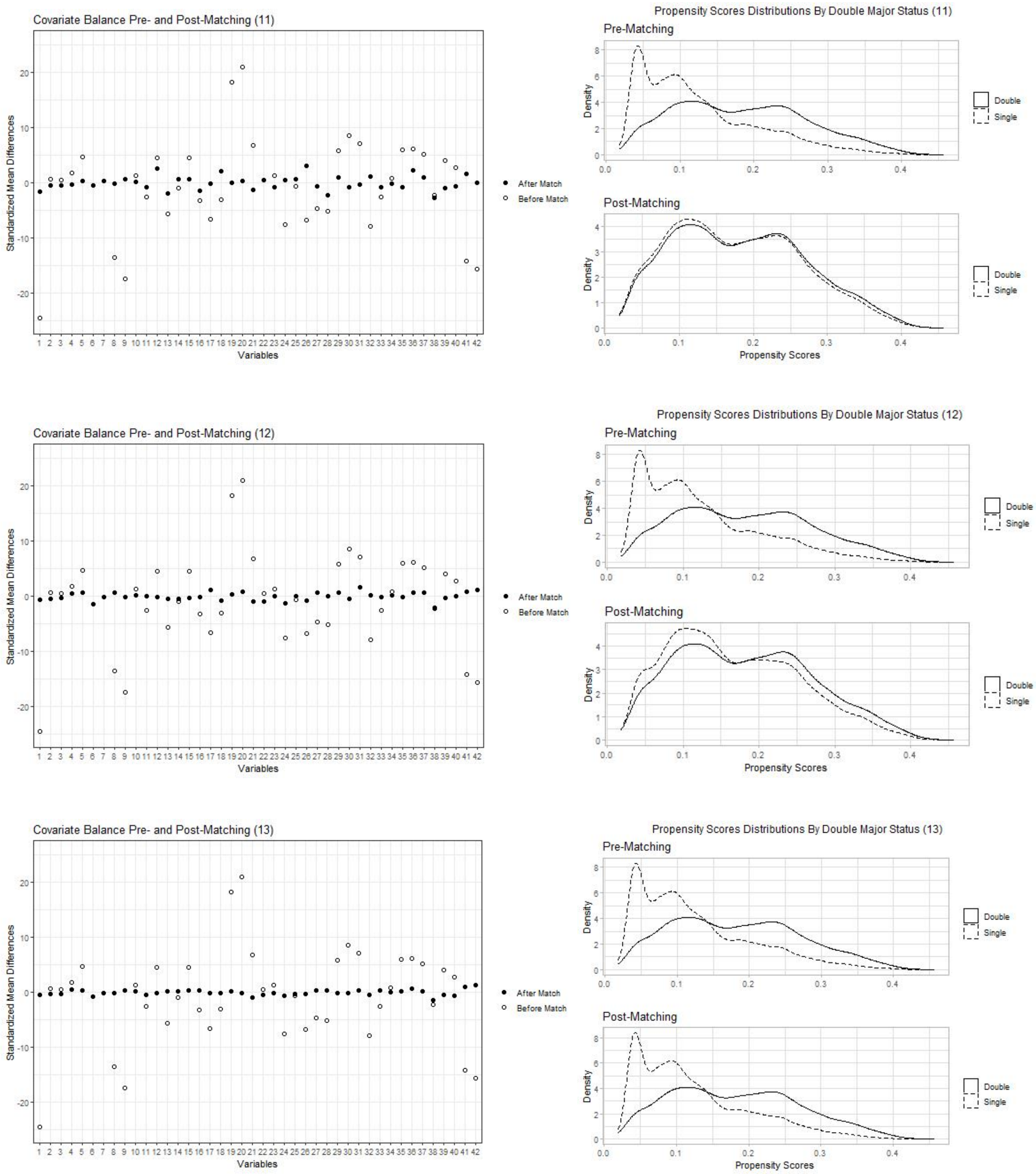
	log(Earnings)			
	(1)	(2)	(3)	(4)
Double major	0.019	0.010	0.026	0.013
	(0.019)	(0.026)	(0.018)	(0.025)
Another bachelors degree	0.014	0.090**	0.016	0.080**
	(0.030)	(0.039)	(0.030)	(0.037)
Graduate degree:				
MBA	0.343***	0.330***	0.322***	0.308***
	(0.028)	(0.044)	(0.027)	(0.042)
Other masters	0.105***	0.080***	0.092***	0.066***
	(0.014)	(0.020)	(0.014)	(0.019)
Doctorate	0.276***	0.179***	0.239***	0.134***
	(0.022)	(0.033)	(0.022)	(0.033)
Medical degree	0.719***	0.518***	0.675***	0.466***
	(0.040)	(0.057)	(0.041)	(0.059)
Law degree	0.466***	0.381***	0.423***	0.326***
	(0.033)	(0.049)	(0.034)	(0.047)
Other professional degree	0.038	0.057	0.011	−0.011
	(0.113)	(0.154)	(0.113)	(0.161)
Employment:				
Tenure	0.223***	0.126	0.214	0.114
	(0.017)	(0.051)	(0.017)	(0.051)
Government	0.123***	0.123***	0.119***	0.123***
	(0.020)	(0.031)	(0.020)	(0.030)
Business/Industry	0.250***	0.278***	0.233***	0.262***
	(0.016)	(0.023)	(0.015)	(0.022)
South	−0.039	−0.047**	−0.035***	−0.042**
	(0.014)	(0.020)	(0.013)	(0.020)
Male	0.081***	0.079**	0.077***	0.073**
	(0.026)	(0.034)	(0.025)	(0.033)
Age	0.073***	0.122***	0.060***	0.122***
	(0.005)	(0.029)	(0.007)	(0.028)
Age squared	−0.001	−0.001***	−0.001***	−0.001***
	(0.0001)	(0.0004)	(0.0001)	(0.0004)
Married	0.037	0.044	0.034*	0.045*
	(0.019)	(0.028)	(0.018)	(0.026)
Male * Married	0.188***	0.147***	0.192***	0.151***
	(0.029)	(0.039)	(0.028)	(0.037)
Veteran	−0.089	−0.086	−0.058	−0.041
	(0.196)	(0.161)	(0.181)	(0.157)
American Indian/Alaska Native	−0.107***	−0.159***	−0.082**	−0.120***
	(0.032)	(0.044)	(0.032)	(0.044)
Black	−0.145***	−0.183***	−0.119***	−0.144***
	(0.029)	(0.033)	(0.029)	(0.033)
Hispanic	−0.038*	−0.086***	−0.033	−0.065**
	(0.022)	(0.023)	(0.023)	(0.026)
White	−0.094	−0.140	−0.079	−0.124

	(0.078)	(0.161)	(0.073)	(0.141)
Native Hawaiian/Other Pacific Islander	−0.082*	−0.096	−0.061	−0.075
	(0.049)	(0.059)	(0.049)	(0.060)
Multiple race	−0.200***	−0.214***	−0.204***	−0.216***
	(0.028)	(0.041)	(0.028)	(0.038)
Primary major:				
Life sciences	−0.097***	−0.160***	−0.102***	−0.170***
	(0.031)	(0.039)	(0.031)	(0.038)
Physical and related sciences	−0.172***	−0.151***	−0.179***	−0.164***
	(0.021)	(0.029)	(0.021)	(0.027)
Social and related sciences	0.049**	0.036	0.030	0.008
	(0.020)	(0.026)	(0.020)	(0.026)
Engineering	−0.111***	−0.107***	−0.105***	−0.102***
	(0.022)	(0.031)	(0.022)	(0.032)
SE-Related Fields	−0.198***	−0.192***	−0.195***	−0.194***
	(0.021)	(0.029)	(0.021)	(0.028)
Non-SE Fields			0.075	0.021
HS diploma before 1989			−0.070***	−0.070**
			(0.018)	(0.027)
Associate degree			0.139***	0.094**
			(0.031)	(0.040)
Undergraduate institution:				
Research			0.070**	0.013
			(0.031)	(0.041)
Doctorate granting			0.062**	−0.0002
			(0.030)	(0.039)
Comprehensive			−0.010	−0.073*
			(0.033)	(0.044)
Liberal arts			0.069***	0.078***
			(0.015)	(0.022)
Private			−0.013	−0.035
			(0.049)	(0.044)
Citizen			−0.046**	−0.047*
			(0.019)	(0.025)
Physical difficulties while obtaining BA			−0.081***	−0.074**
			(0.021)	(0.034)
Physical disabilities			0.015	−0.009
			(0.026)	(0.044)
Father's education level:				
High school			0.026	0.012
			(0.027)	(0.044)
Some college			0.069**	0.069
			(0.027)	(0.045)
Bachelors			0.050*	0.052
			(0.030)	(0.047)
Masters			0.048	0.045
			(0.039)	(0.054)
Professional			−0.010	−0.008
			(0.044)	(0.070)

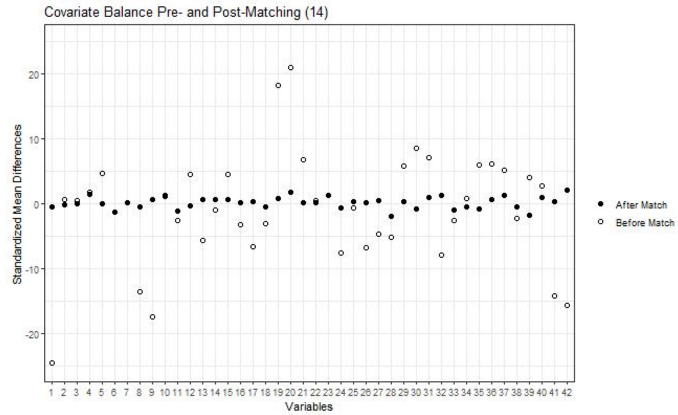
Doctorate			0.003	−0.010
			(0.027)	(0.048)
Mother's education level:				
High school			0.003	−0.008
			(0.028)	(0.049)
Some college			0.010	0.010
			(0.029)	(0.050)
Bachelors			0.009	0.008
			(0.032)	(0.053)
Masters			0.001	0.056
			(0.055)	(0.072)
Professional			0.031	0.030
			(0.055)	(0.082)
Doctorate	9.020***	8.168***	9.145***	8.141***
	(0.108)	(0.456)	(0.157)	(0.453)
Age¡=40?	No	Yes	No	Yes
Observations	53,453	20,347	53,453	20,347
R^2	0.265	0.260	0.278	0.278
Adjusted R^2	0.265	0.259	0.278	0.276
Note:	*p<0.1; **p<0.05; ***p<0.01			
	Heteroskedastic robust standard errors given in parentheses.			

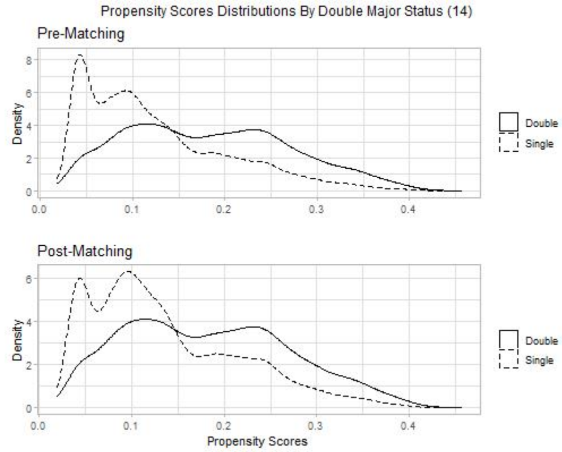
Appendix B[11]

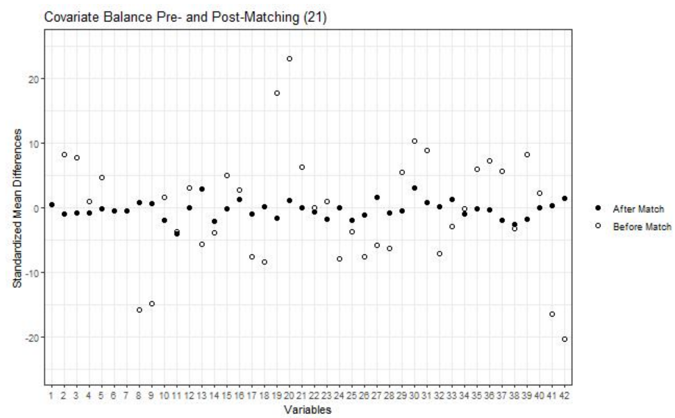
Figure 6: Matching Quality

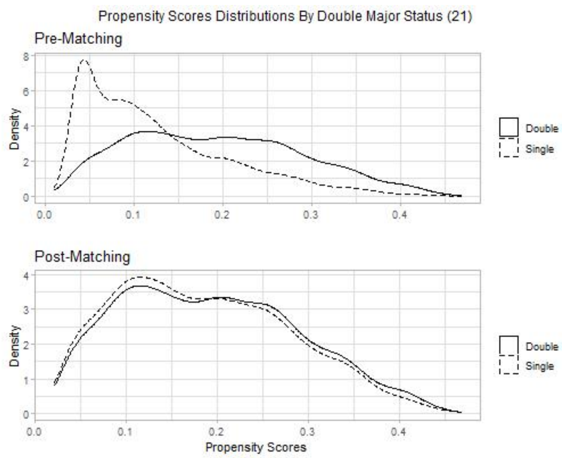


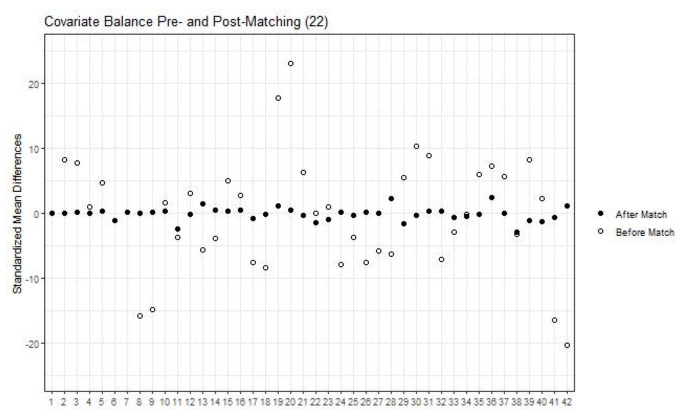
[11] The figure label (ij) corresponds to the model in the ith row and jth column of Table 4. For example, (31) means nearest 1-1 matching for the 2003 full data.

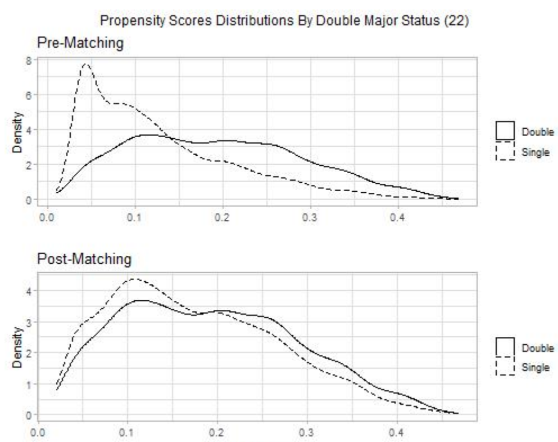

Covariate Balance Pre- and Post-Matching (14)
Standardized Mean Differences
Variables
After Match
Before Match

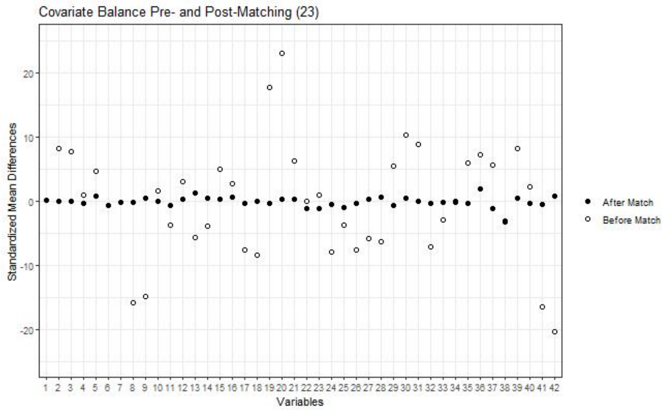

Propensity Scores Distributions By Double Major Status (14)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

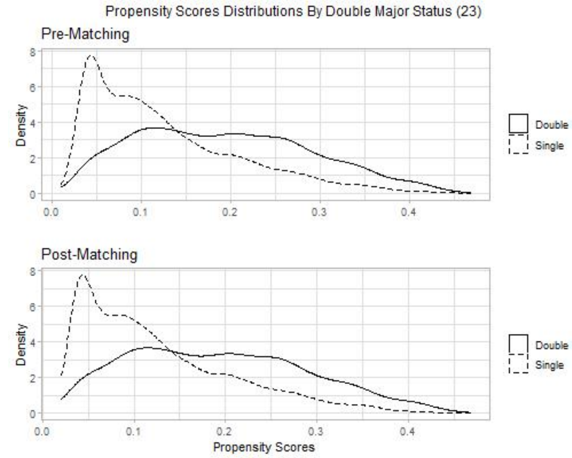

Covariate Balance Pre- and Post-Matching (21)
Standardized Mean Differences
Variables
After Match
Before Match

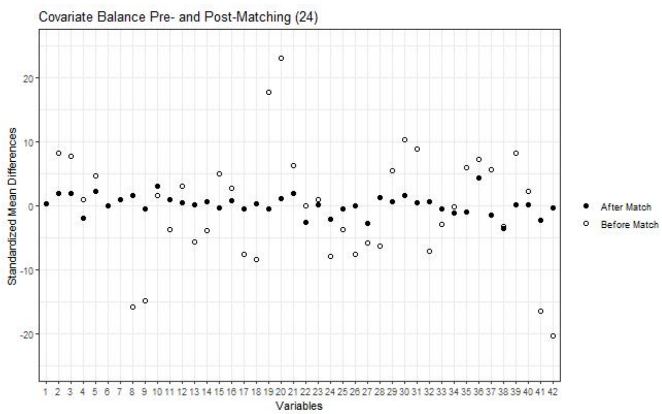

Propensity Scores Distributions By Double Major Status (21)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

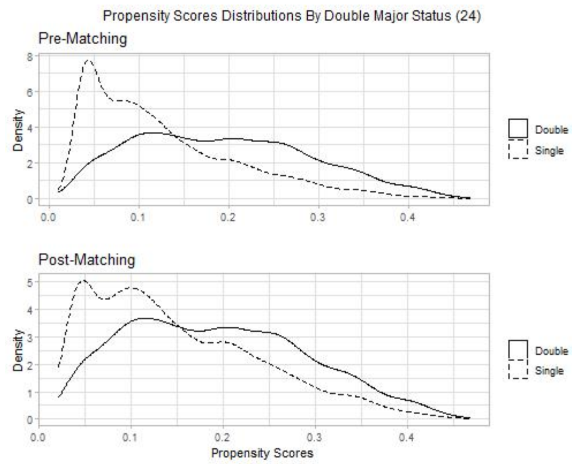

Covariate Balance Pre- and Post-Matching (22)
Standardized Mean Differences
Variables
After Match
Before Match

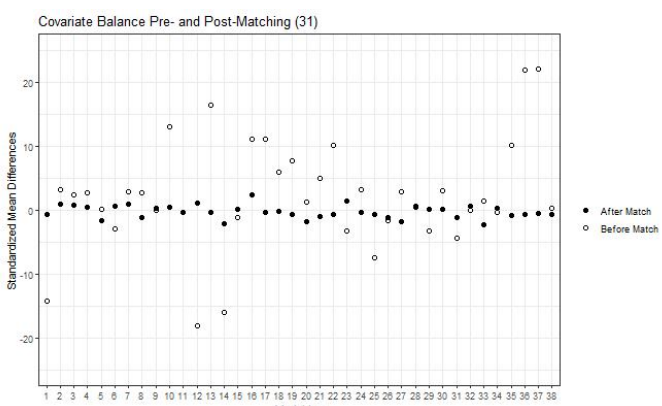

Propensity Scores Distributions By Double Major Status (22)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

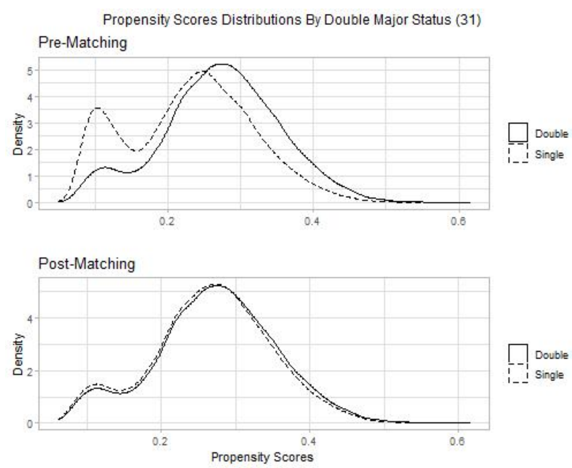

Covariate Balance Pre- and Post-Matching (23)
Standardized Mean Differences
Variables
After Match
Before Match

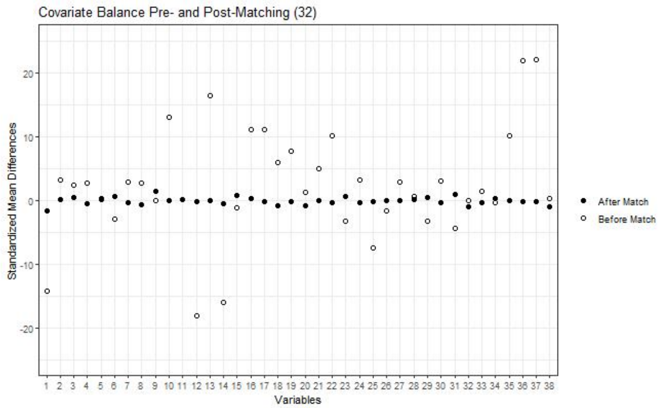

Propensity Scores Distributions By Double Major Status (23)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

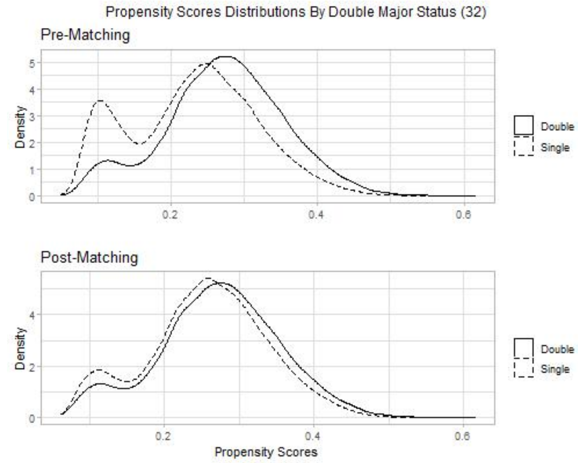

Covariate Balance Pre- and Post-Matching (24)
Standardized Mean Differences
Variables
After Match
Before Match

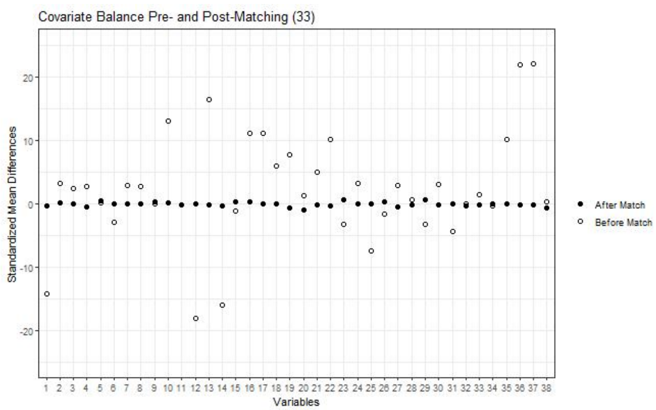

Propensity Scores Distributions By Double Major Status (24)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

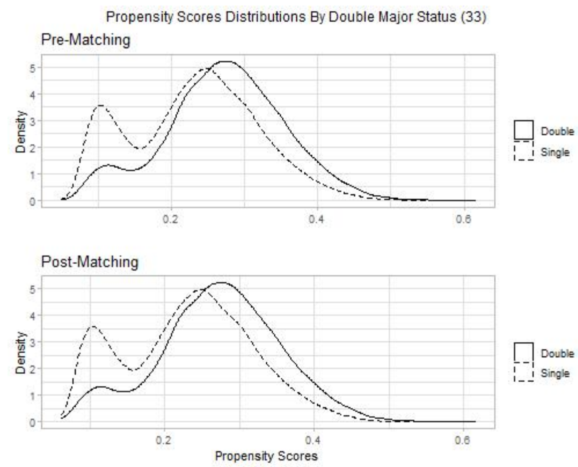

Covariate Balance Pre- and Post-Matching (31)
Standardized Mean Differences
Variables
After Match
Before Match

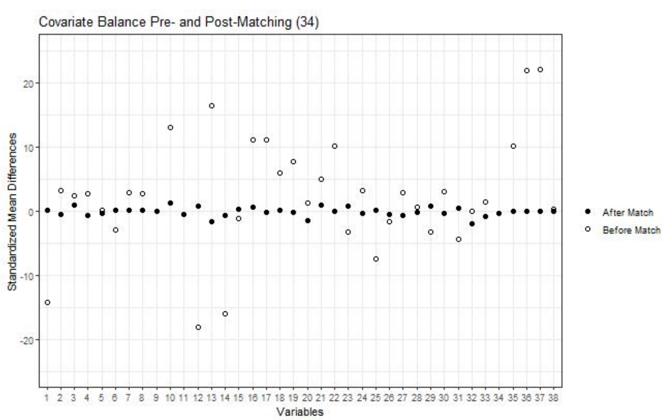

Propensity Scores Distributions By Double Major Status (31)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

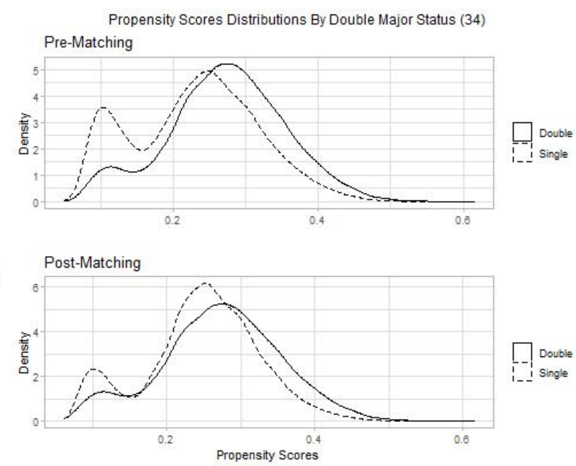

Covariate Balance Pre- and Post-Matching (32)
Standardized Mean Differences
Variables
After Match
Before Match

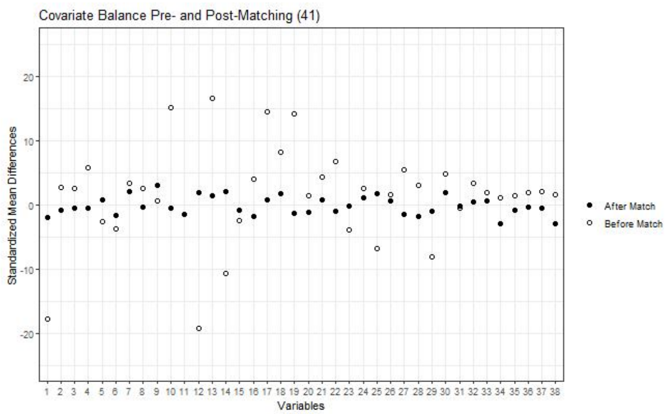

Propensity Scores Distributions By Double Major Status (32)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

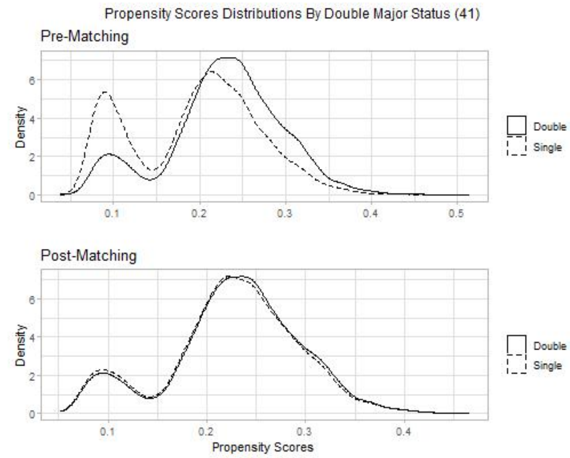

Covariate Balance Pre- and Post-Matching (33)
Standardized Mean Differences
Variables
After Match
Before Match

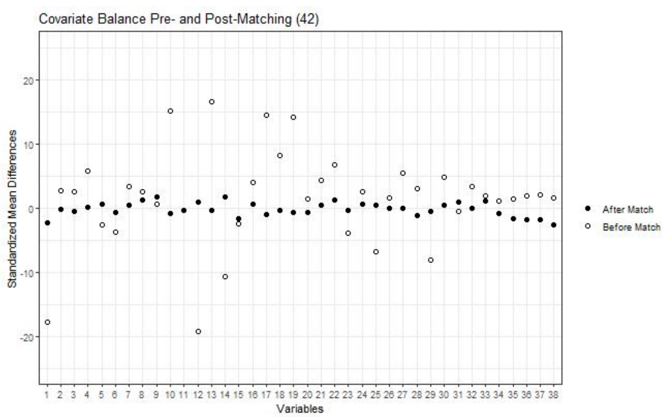

Propensity Scores Distributions By Double Major Status (33)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

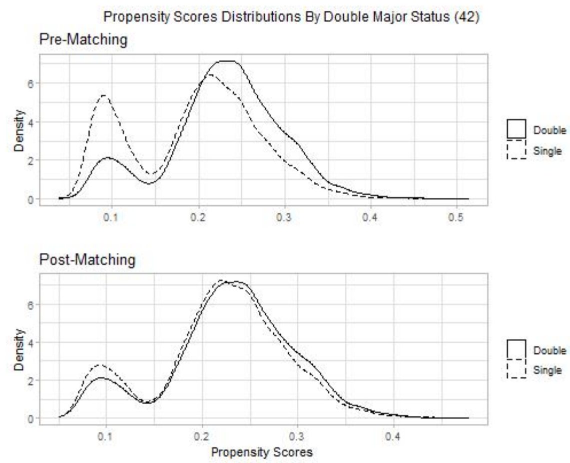

Covariate Balance Pre- and Post-Matching (34)
Standardized Mean Differences
Variables
After Match
Before Match

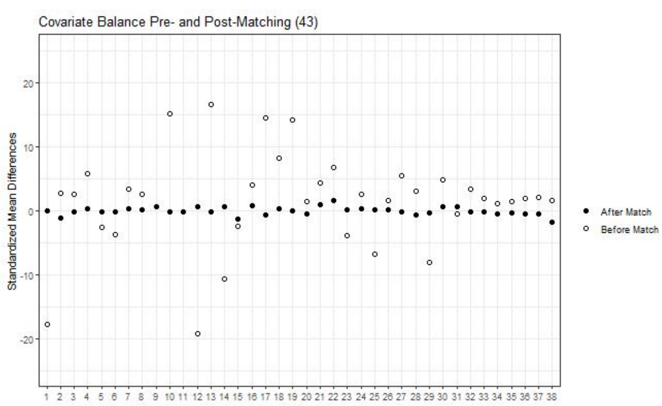

Propensity Scores Distributions By Double Major Status (34)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

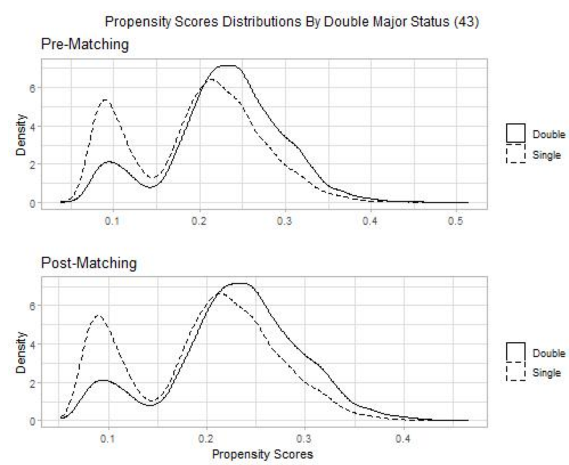

Covariate Balance Pre- and Post-Matching (41)
Standardized Mean Differences
Variables
After Match
Before Match

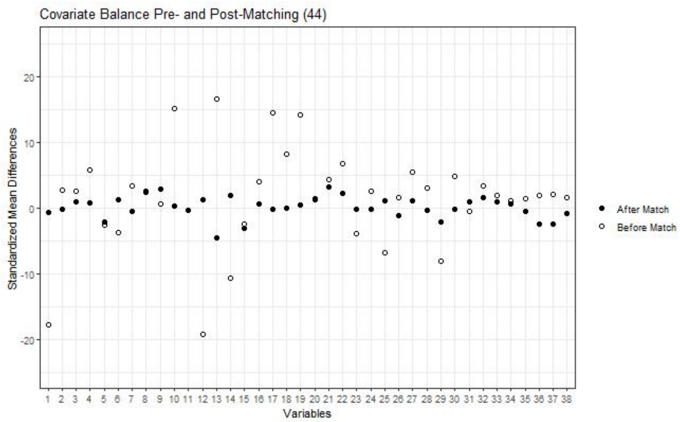

Propensity Scores Distributions By Double Major Status (41)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single

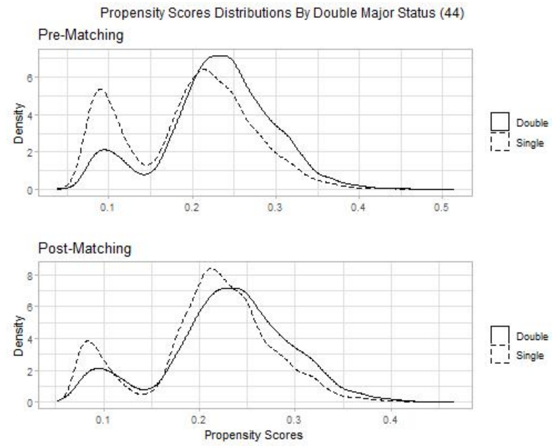

Covariate Balance Pre- and Post-Matching (42)
Standardized Mean Differences
Variables
After Match
Before Match


Propensity Scores Distributions By Double Major Status (42)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single


Covariate Balance Pre- and Post-Matching (43)
Standardized Mean Differences
Variables
After Match
Before Match


Propensity Scores Distributions By Double Major Status (43)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single


Covariate Balance Pre- and Post-Matching (44)
Standardized Mean Differences
Variables
After Match
Before Match


Propensity Scores Distributions By Double Major Status (44)
Pre-Matching
Post-Matching
Density
Propensity Scores
Double
Single