



Should Student Employment Be Subsidized? Conditional Counterfactuals and the Outcomes of Work-Study Participation

A CAPSEE Working Paper

Judith Scott-Clayton

Veronica Minaya

Community College Research Center
Teachers College, Columbia University

September 2014

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C110011 to Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Appendices A and B are available in a separate document available at capseecenter.org

For information about authors and CAPSEE, visit capseecenter.org

Abstract

Student employment subsidies are one of the largest types of federal employment subsidies, and one of the oldest forms of student aid. Yet it is unclear whether they help or harm students' long term outcomes. We present a framework that decomposes overall effects into a weighted average of effects for marginal and inframarginal workers. We then use an application of propensity scores, which we call conditional-counterfactual matching, in which we estimate the overall impact, and the impact under two distinct counterfactuals: working at an unsubsidized job, or not working at all. Finally, we estimate the effects of the largest student employment subsidy program—Federal Work-Study (FWS)—for a broad range of participants and outcomes. Our results suggest that about half of FWS participants are inframarginal workers, for whom FWS reduces hours worked and improves academic outcomes, but has little impact on future employment. For students who would not have worked otherwise, the pattern of effects reverses. With the exception of first-year GPA, we find scant evidence of negative effects of FWS for any outcome or subgroup. However, positive effects are largest for lower-income and lower-SAT subgroups, suggesting there may be gains to improved targeting of funds.

Table of Contents

1. Introduction	1
2. Conceptual Framework	3
Wage Subsidies in Partial Equilibrium Analysis	3
Estimating the Impact of a Student Employment Subsidy	5
3. Program Details and Prior Literature	7
The Federal Work-Study Program	7
Prior Research Relating to Federal Work-Study	9
4. Empirical Methodology	11
Propensity Scores and Conditional Counterfactual Matching	11
Implementation	13
Data and Sample	14
5. Results	16
Modeling the Propensity Score: Who Participates in FWS?	16
Common Support and Match Quality	18
Main Results: Estimated Effects of FWS Participation	22
Robustness to Alternative Specifications	26
Sensitivity to Unobservable Selection	28
6. Subgroup Analysis	29
Parental Income and Student SAT Score	29
Local Unemployment Rates and Urbanicity	31
Gender	32
7. Discussion	32
References	34

1. Introduction

Student employment subsidies—also known as work-study programs—are one of the largest types of targeted employment subsidies funded by the federal government, and are also one of the oldest policy mechanisms intended to promote college access and persistence for low-income students. Since 1964, the Federal Work-Study program has provided approximately \$1 billion annually to cover up to 75 percent of the wages of student employees, who typically work on campus for 10 to 15 hours per week (The College Board, 2012; U.S. Department of Education, 2009). Federal Work-Study has an extensive reach, serving nearly 700,000 students per year in recent decades (for a total of 30 million student-years since its inception), including one out of every 10 full-time first-year undergraduates (and three out of 10 at private nonprofit four-year colleges).¹ Some states and institutions also subsidize student employment.

Despite the durability of student employment subsidies in student aid policy, their economic justification has never been fully articulated, and the presumed impacts of such subsidies are sometimes conflated with the impacts of student employment itself. Prior literature on student employment at both the college and high school level suggests that combining work and school reduces students' academic performance (Stinebrickner & Stinebrickner, 2003; Tyler, 2003), even while it may improve later labor market outcomes (Light, 2001; Ruhm, 1997). But knowing whether or not student employment is beneficial is insufficient to infer the effectiveness of student employment subsidies as a policy tool. Indeed, student employment subsidies may be most effective when students' unsubsidized employment options are harmful, because they may enable working students to access jobs with better amenities (or fewer disamenities).

As in other wage subsidy contexts, the employment and (hedonic) wage effects of a student employment subsidy will depend upon the elasticities of labor supply and demand, as well as on the design features of the program (Katz, 1996). In the absence of underlying market imperfections, wage subsidies can generate windfalls to employers (in this case, colleges) and deadweight loss to society, and are thus an inefficient means of transferring money to students. However, market imperfections that could justify student employment subsidies include students' information constraints regarding the value of work experience, structural barriers to employment for low-income and/or minority students, or statutory barriers (e.g., minimum wage laws) constraining employers' ability to offer the types of part-time, flexible positions most complementary with college enrollment. Whatever the underlying rationale, policymakers may want to know to what extent a student employment subsidy actually increases students' access to productive employment, to what extent it raises their wages or improves the non-wage aspects of employment, and to what extent it simply subsidizes institutional employers to hire students that would have been working in similar jobs anyway.

¹ Online computation by NCES QuickStats on 8/29/2013 using BPS: 2009 Beginning Postsecondary Students.

Despite the scale and longevity of the Federal Work-Study program, scant research has been conducted regarding its impacts. This is particularly surprising given the dramatically increasing prevalence of college student employment over the past 40 years: average weekly hours of work (including zeros) rose from six hours per week in 1970 to 10 hours per week just before the Great Recession (Perna, Cooper, & Li, 2007; Scott-Clayton, 2012). The available rigorous research examining effects of work-study has been limited to single-institution or single-small-state data, and the findings have been inconsistent across studies and across subgroups within studies (DesJardins, Ahlburg, & McCall, 2002; Scott-Clayton, 2011; Stinebrickner & Stinebrickner, 2003). Prior findings have been limited to academic outcomes, despite the fact that potential returns in the labor market are an important motivation for the policy.

We extend the literature in several ways. First, we present a conceptual framework linking work-study programs to the broader wage-subsidy literature, and illustrating how the impact of student employment subsidies operates through two distinct mechanisms. Just as any subsidy may affect both quantity and price, we show how a work-study program may influence both students' likelihood of working, as well as the types of jobs students hold conditional on working (wage and non-wage aspects including hours, location, job content). Some work-study participants who would not have worked may do so because of the program, but other students will replace outside work with a work-study job and may even *reduce* their overall hours worked as a result. The impacts of work-study participation in these two cases could plausibly go in opposite directions. As a result, not only are program effects likely to be heterogeneous across individuals, but the average effect is unlikely to apply to *any* individual.

Second, our conceptual framework leads us to use an application of propensity scores, which we call *conditional-counterfactual matching*. This approach is related to, but different from, both the principal stratification (Barnard, Frangakis, Hill, & Rubin, 2003; Frangakis & Rubin, 2002) and principal score modeling (Hill, Waldfogel, & Brooks-Gunn, 2002; Jo & Stuart, 2009) approaches in the context of randomized experiments. In this approach, we utilize propensity score matching to estimate not only the overall impact of participating in work-study, but also to estimate the impact under two highly distinct conditional counterfactuals: what would have happened if the participant had worked at a non-work-study job, and what would have happened if the participant had not worked at all. In all three analyses, we perform extensive diagnostics and sensitivity analyses to help us evaluate the plausibility of the selection-on-observables assumption, the quality of matches, and the robustness of our findings to alternative specifications. Moreover, our conditional counterfactual analysis provides multiple opportunities to test implications from the theoretical framework. While unobservable selection cannot be ruled out in this non-experimental context, our key findings are reasonably robust to analyses that place bounds on unobservable bias (Becker & Caliendo, 2007; Oster, 2014).

Finally, this study provides estimates of the effect of Federal Work-Study participation for a much broader range of participants and outcomes than have been considered by the prior literature. We utilize two waves of the nationally representative Beginning Postsecondary

Student (BPS) survey to look at the effects of work-study participation not only on first-year GPA, but also on number of months enrolled, bachelor's degree completion, employment, and earnings six years after college entry. This breadth of analysis is particularly valuable because our theoretical framework implies heterogeneous impacts depending upon the context, the counterfactual and the type of outcome (academic versus labor market). We view our results as complementing the more causally rigorous, but highly localized estimates from prior studies.

We find that six years after entering college, Federal Work-Study participants are 3.2 percentage points more likely to earn a BA and 2.4 percentage points more likely to be employed after finishing school than similar non-participants ($p < 0.05$ in both cases). But these overall impacts mask important patterns of heterogeneity. Less than half of work-study participants are induced into employment; a slight majority are students who would have worked in their first year anyway. For these students, participation appears to have positive academic impacts, but no future employment impacts. In contrast, compared against not working at all, participation appears to have null or negative effects on academics but a positive effect on post-college employment. With the exception of first-year GPA, we find little evidence of negative effects for any outcome or subgroup. However, positive effects appear substantially larger for lower-income and lower-SAT students versus their higher-income, higher-SAT counterparts.

In the following section, we present our conceptual framework. In Section III, we describe the Federal Work-Study program in more detail and summarize the relevant literature. Section IV presents our methodology, Section V presents our main results, Section VI presents results from subgroup analyses, and Section VII provides a discussion with implications for policy and future research.

2. Conceptual Framework

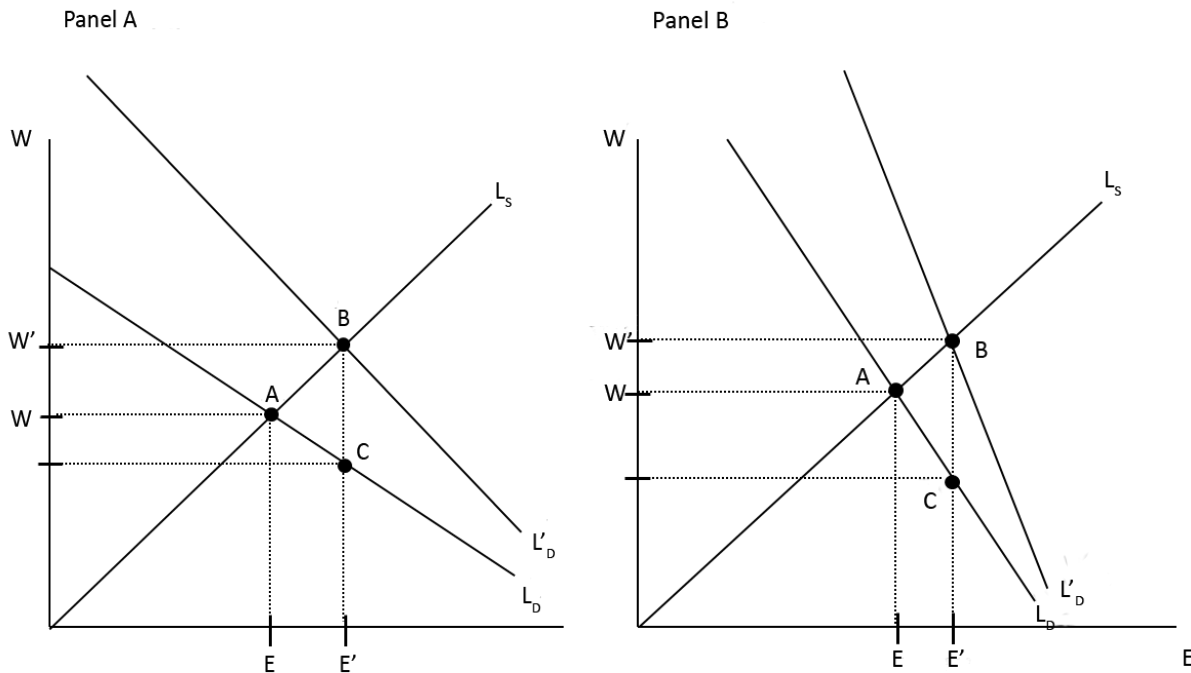
Wage Subsidies in Partial Equilibrium Analysis

Katz (1996) provides a useful overview of the economics of wage subsidies. If labor supply were infinitely elastic, proportional wage subsidies would have no effect on wages but employers would hire more individuals. Figure 1 illustrates two more general partial-equilibrium cases in which labor supply is less than infinitely elastic, with Panel A showing a case of more elastic labor demand and Panel B showing a case of less elastic labor demand. In both cases, the wage subsidy increases both employment and wages, but the effects on both margins are larger when demand is more elastic. Though not shown in Figure 1, one can also imagine a case of very

inelastic labor supply (for example, if credit constrained students must work a certain amount to fund their studies), which would imply smaller employment effects but larger wage effects.²

In the absence of underlying market imperfections, any shifts in equilibrium employment induced by a wage subsidy generate deadweight loss. However, if the underlying market is flawed, such subsidies may increase total social surplus. For example, government subsidies could be justified if students undervalue the returns to work experience, and thus either work less than they optimally should or over-prioritize wages rather than experience when choosing among job options. Alternatively, low-income and/or minority students may face structural barriers to employment that limit their options. Finally, minimum wage laws and other statutory constraints may limit employers' ability to offer the types of part-time, flexible positions that are less convenient for employers but most complementary with college enrollment.

Figure 1. Partial Equilibrium Analysis of Effects of a Proportional Wage Subsidy



² Katz notes that while in theory it should not matter whether the subsidy nominally goes to employers or workers, in practice this and other design features may affect a program's impact. For example, a relevant feature in the case of FWS is that if funds go unused by a college in one year, their funds for the next award year will be reduced.

One basic implication from Figure 1 is that there are two important margins of possible impact for a wage subsidy: employment and wages (or rather, hedonic wages which include the value of non-monetary job amenities such as scheduling flexibility, intensity, and connection to interests). A second basic implication is that the magnitude of effects on either margin will depend upon the elasticities of labor supply and demand, which may well vary from context to context, subgroup to subgroup.

Unlike a typical wage subsidy, however, in the case of work-study programs, student employment and wages are not the outcomes of ultimate interest. They are only means to other ends, including improved academic and future labor market outcomes. Importantly, the effect on these later outcomes may be very different—potentially even opposite sign—depending upon whether the recipient is one of the “new” workers induced into employment by the subsidy, or whether the recipient is an “inframarginal” worker who would have worked anyway, but now receives a higher hedonic wage.

Estimating the Impact of a Student Employment Subsidy

Econometrically, the simplest way to estimate the impact of an employment subsidy on future academic and labor market outcomes would be to randomly assign students access to the subsidy. Then one could estimate the equation:

$$(1) \quad y_{ij} = \alpha + \beta(\text{Subsidy}_i) + \delta\mathbf{X}_i + \gamma\mathbf{Z}_j + v_{ij}$$

where y_{ij} is a student outcome (indexed for the i th individual in the j th institution) such as first-year GPA, bachelor’s degree completion, or later-life employment or earnings; Subsidy_i is an indicator of a subsidy offer, \mathbf{X}_i is a vector of individual-level controls; \mathbf{Z}_j is a vector of institution-level controls, and v_{ij} is an error term (possibly clustered within institutions). Because of random assignment, β can be interpreted as the causal impact of a subsidy offer (\mathbf{X}_i and \mathbf{Z}_j here serve only to increase the precision of the estimates). If every student offered the subsidy actually takes it up, β can be further interpreted as the causal effect of subsidy receipt.

Even in the random-assignment with full take-up scenario, however, simply knowing β may be unsatisfying, because it represents a combination of effects on two margins: the employment margin and the job characteristic (or hedonic wage) margin. The subsidy may induce changes in the likelihood that students work while in school, or in the types of jobs they hold conditional on employment, or both. If we make the assumption that work-study operates only through these two channels, we can model student outcomes as:

$$(2) \quad y_{ij} = \alpha + \beta_1(\text{Emp}_i) + \beta_2(\text{SubsJob}_i) + \delta\mathbf{X}_i + \gamma\mathbf{Z}_j + v_{ij}$$

where Emp_i is an indicator of any type of student employment, and SubsJob_i is an indicator of subsidized employment, such that β_2 represents the differential effect of a subsidized job over and above any employment, due to potential differences in job characteristics such as location, wage, relationship to major, weekly hours, flexibility of scheduling, and so on.

Allowing the impact of work-study to vary across individuals, and using Rubin's (1974) potential outcomes notation, we can decompose the individual treatment effect as:³

$$(3) \quad \beta_i = y(1)_i - y(0)_i = \begin{cases} \beta_1 + \beta_2 & \text{for those induced into work (a)} \\ \beta_2 & \text{for always - workers (b)} \\ 0 & \text{for never - workers (c)} \end{cases}$$

where $y(1)_i$ represents an individual's potential outcome if assigned to the subsidized group and $y(0)_i$ represents an individual's potential outcome if assigned to the control group. Always-workers are those who would work regardless of treatment assignment, and never workers are those who would not work, regardless of treatment assignment. The average treatment effect for the treated group can then be expressed as:

$$(4) \quad \beta = (1 - \text{Pr}(Emp)_c) * (\beta_1 + \beta_2) + \text{Pr}(Emp)_c * \beta_2$$

This equation illustrates how impacts on the employment margin (the first term in the above equation) combine with impacts on the job characteristics margin (the second term) to generate the overall impact. It is possible that β_1 (the effect of working at a typical job) and β_2 (the additional effects of working at a subsidized job) may have opposite signs, so an overall null result may mask important underlying impacts. Moreover, different samples may generate wildly different estimates of the impact of work-study if there is heterogeneity in control group employment rates or job characteristics. Finally, the equations above indicate that if the impact of work-study operates primarily via changes in employment status, β may be informative about the effects of student employment more broadly (and vice versa); but if it operates primarily by changing the types of jobs students hold, it would be inappropriate to generalize to other types of student work.

How then might this important heterogeneity be explored in the context of a randomized experiment? Even in an RCT, it would not be straightforward to identify β_1 and β_2 , since we cannot identify who was induced into employment, and who would have worked anyway. One option might be to stratify the sample prior to randomization into groups with higher and lower predicted propensities of employment⁴. As we will show, however, students' employment decisions can be hard to predict based upon background characteristics, making it difficult to create highly distinct subgroups.

Ultimately, to obtain estimates of β , β_1 , and β_2 , three distinct experiments would be needed: one in which the control group was unrestricted, one in which the control group was expected to work at a non-FWS job, and one in which the control group was expected not to

³ In this framework $y(1)$ represents an individual's potential outcome if assigned to the treatment group and $y(0)$ is the same individual's potential outcome if assigned to the control group. Note that we are reluctant to use the typical "complier" terminology here because even among those who complied with a theoretical random assignment to FWS treatment (in the sense that they would not have participated in FWS if assigned to a control group), there could be always-workers.

⁴ This would be a straightforward application of the type of analysis conducted by Frangakis and Rubin (2002), Barnard et al. (2003), Jo and Stuart (2009), or Hill et al. (2002).

work at all. We use this “ideal experiment” to guide our empirical approach in Section IV. We will not have the benefit of randomization, meaning that we will need to confront issues of selection bias in addition to these general conceptual issues.

3. Program Details and Prior Literature

The Federal Work-Study Program

Student eligibility for Federal Work-Study (FWS) assistance is based on financial need (U.S. Department of Education, 2009). In order to establish FWS eligibility, a student must file a Free Application for Federal Student Aid (FAFSA). Using a complex formula based primarily on family income, but also on assets, family size, and number of family members in college, an “expected family contribution” (EFC) is computed. If the student’s total cost of attendance (including living expenses, for full-time students) exceeds the EFC, the student is deemed to have financial need. Students are eligible for, but not necessarily offered nor guaranteed, FWS funds to fill this need as long as the gap has not already been filled by other sources of aid.

Unlike Pell Grants or Stafford Loans, the FWS program is a “campus-based” federal program, meaning that program funds are allocated as a lump sum to institutions, which then have great discretion on how funds may be distributed among eligible students. The formula that determines institutional allocations is based primarily upon prior-year allocations rather than upon any measure of students’ financial need, so longstanding institutional participants receive significantly more per student than do recent entrants (Smole, 2005). Selective four-year institutions receive highly disproportionate awards relative to the size of their student bodies.⁵ At schools with any FWS funds, only 16 percent of institutions award FWS to every eligible student.⁶

Schools may use work-study funds to cover up to 75 percent of an eligible student’s wages, or they may choose to cover a lower fraction of wages for a greater number of students. Despite the large wage subsidy, FWS jobs appear to provide lower hourly wages than other jobs, which may indicate that FWS jobs have attractive compensating characteristics.⁷ Approximately

⁵ For example, in 2009–10 Columbia University received an allocation of \$6.75 million while the University of Florida, despite having more than four times as many undergraduates, received just \$2.97 million (U.S. Department of Education, 2010b).

⁶ Of the remainder, 82 percent give preference to students with exceptional financial need, and 78 percent give preference to students who actively request FWS or indicate their interest on the FAFSA (U.S. Department of Education, 2000a, p. 58). The timing of the FWS request, the student’s year in school, attendance status, and previous academic performance are other factors that institutions report using to determine FWS offers (U.S. Department of Education, 2000b, p. C9). While awards may go to graduate students, in practice 93 percent of participants are undergraduates (U.S. Department of Education, 2010a).

⁷ The BPS does not ask about wages, only about typical hours per week and total earnings from jobs during the school year. Assuming that on average working students work 28 weeks during the school year, the hourly wage of

one-third of recipients are paid at the minimum wage; the average wage is about 18 percent above minimum wage (U.S. Department of Education, 2000a). Work-study funds can be used at most on-campus jobs as well as some types of off-campus community-service jobs.

Table 1 describes the types of jobs held by working first-year students inside and outside of the FWS program, as measured in two waves of the nationally representative Beginning Postsecondary Students (BPS) dataset (National Center for Education Statistics [NCES], 2003; 2011). More than 80 percent of FWS participants work on campus, compared with only 8 percent of working non-participants. Among students who held only one job, FWS participants averaged 11 hours of work per week compared with 18 hours per week for working non-participants. Although the majority of jobs are unrelated to students' majors, FWS jobs are 6 percentage points more likely to be related than non-FWS jobs. Compared to other working first-year students, FWS participants are much more likely to work in clerical occupations, and less likely to work in sales, labor or service. Finally, while the majority of all working students report their job has no effect on their academic performance, FWS participants are somewhat more likely to report that their job has a positive effect and somewhat less likely to report negative effects.

Table 1: Student Employment

Variable	BPS 1996–2001 & BPS 2004–2009 Sample Characteristics		
	Total	FWS Recipients	Non-FWS working students
Full Sample			
Any employment during school year (in %)	69.5		
Working Students (BPS 96 and BPS 04 Combined)			
Total hours worked per week in year 1 (zeros excluded)	26.5	17.8	27.5
Number of jobs held during year 1	1.4	1.7	1.4
Total hours worked per week in year 1, of those who only held one job	16.9	11.4	18.4
Any on-campus (include on/both) employment (in %)	15.3	83.5	8.2
Any off-campus (include off/both) employment (in %)	86.0	44.7	90.4
Any job is related to major (in %)	22.8	28.3	22.3
Estimated hourly wages, 2004 dollars (2003–04 only)		\$7.30	\$6.30

FWS participants was approximately \$6.30 in 2003–04, while working non-participants earn a higher hourly wage (approximately \$7.30). The FWS estimate is in line with a survey from 2000 which found average wages of \$6.10 for FWS recipients (U.S. Department of Education, 2000a).

Sample Size (unweighted)	16,420	2,640	13,790
Working Students (BPS 96 Only)			
Job has a positive effect on academic performance (in %)*	11.8	15.9	11.4
Job has a negative effect on academic performance (in %)*	31.9	23.6	32.8
Job has no effect on academic performance (in %)*	56.3	60.5	55.8
First year occupation*:			
Clerical	29.4	46.0	28.0
Manager	7.7	5.1	7.9
Professional	5.3	9.3	5.0
Service Occupations	20.5	18.3	20.7
Technical	2.3	2.9	2.2
Laborer	6.7	3.1	7.0
Sales	9.6	4.6	10.0
Other	18.5	10.6	19.2
Sample Size (unweighted)	5,360	880	4,480

Notes: Authors' calculations based on BPS:96/01 and 04/09 restricted use data files. The sample only includes full-time and part-time and dependent and independent students entering a two-year and a four-year institution in their first academic year, 1995–96 and 2003–04. The BPS tabulations use BPS longitudinal sampling weights (wta000). Sample sizes may vary for certain variables due to missing values. Unweighted sample sizes are rounded to the nearest 10. Estimated wages are computed based on total school year earnings and typical weekly hours of work, assuming that students work 28 weeks during the school year.

Prior Research Relating to Federal Work-Study

Despite having as long a history as Pell Grants, empirical research on the consequences of the FWS program is limited, and it has never been evaluated using a randomized study. In fact, causal evidence on college student employment of any kind is limited, due to the difficulty of identifying plausibly exogenous variation in work decisions. Recent reviews of the non-experimental student employment research have highlighted the distinction between on- and off-campus jobs, concluding that while off-campus employment appears to have negative effects, on-campus jobs may positively influence persistence and completion (Pascarella & Terenzini, 2005; Hossler, Ziskin, Kim, Cekic, & Gross, 2008; Perna, 2010).

A handful of studies have attempted to identify sources of plausibly exogenous variation in student employment from which a truly causal effect may be estimated, although none of these studies separately estimates the effect of on-campus employment. Darolia (2014) uses a student fixed-effects approach with data from the 1997 National Longitudinal Survey of Youth, and finds no evidence that students have lower GPAs in terms when they work more intensively, but full-time students accumulate fewer credits. DeSimone (2008), Kalenkoski and Pabilonia (2010), and Dadgar (2012) instrument for student employment (using parental schooling and religion,

and parental transfers and net schooling costs, and sector-season of employment, respectively) and find small negative effects on students' grades (a reduction of 0.011 to 0.03 in GPA per hour of work, though Dadgar's estimates are not statistically significant).⁸

Stinebrickner and Stinebrickner (2003) provide the one of the most compelling causal analyses to date of the effect of on-campus student employment. The authors analyze a small private college in Kentucky at which all students are required to work at a campus job for 10 hours per week, but some jobs offer students the possibility to work more. Although the study does not separate out federal work-study employment, it is likely that many of these jobs are at least partially funded by FWS. Students at the college are randomly assigned by administrators to on-campus jobs, and those who are assigned to a job with additional hours available end up working more than those for whom this is not an option. The authors find that students who work more because they were assigned to a high-availability job earn significantly lower first-year GPAs, a decline of about 0.162 points per additional hour of work.

Only two prior studies examine FWS specifically. Scott-Clayton (2011) uses an instrumental-variables difference-in-difference (IV-DID) approach with administrative data on public two- and four-year college enrollees in West Virginia.⁹ She uses institution-level federal FWS allocations (which are determined by an historical federal formula) to instrument for actual FWS participation, with ineligible students serving as a control group to account for unobserved fixed differences between high- and low-allocation institutions. Scott-Clayton finds strong impacts of FWS participation on the likelihood of school-year employment. However, she finds no statistically significant effects of FWS participation on academic outcomes for the full sample (though point estimates are generally negative). She also finds heterogeneity by gender, with strong negative academic effects for women, and some significant positive effects for men. Long and Soliz (2014). replicate this approach using data on public four-year enrollees in Ohio, and find a small but statistically significant negative effect of FWS on first and second semester GPAs, but a statistically significant positive effect on first year cumulative credits earned (they find no significant gender differences).

This body of non-experimental and quasi-experimental evidence remains inconclusive regarding the impact of FWS. Those studies with stronger causal identification strategies (Scott-Clayton, 2011; Stinebrickner & Stinebrickner, 2003) have tended to find more negative results. But given the likelihood of heterogeneous effects depending upon the counterfactual and the

⁸ Although Kalenkoski and Pabilonia (2010) tried to use variation in state work-study assistance as an instrument for employment, they found that no statistically significant effect of the state programs on hours of work could be detected in their data.

⁹ We attempted to implement a similar approach using the BPS data; unfortunately, the first stage estimate was extremely small in magnitude after controlling for institutional characteristics known to correlate with FWS allocations, and the subsequent IV estimates were too noisy to detect any consistent pattern of findings. The failure of the IV-DID strategy in the national data could be due to the highly skewed distribution of FWS funds nationally; noise in the national FWS-by-institution allocation data; or due to the very small number of FWS recipients per institution in the sample (1 or 2 on average).

outcome of interest, the highly localized nature of quasi-experimental estimates is a non-trivial limitation. The propensity score approach we describe below complements prior quasi-experimental work with nationally representative estimates from propensity score estimation, and allows for a deeper exploration of heterogeneity.

4. Empirical Methodology

Propensity Scores and Conditional Counterfactual Matching

We describe an application of propensity score matching (PSM) that we term *conditional counterfactual matching*. This approach is related to, but distinct from, both the principal stratification (Barnard et al., 2003; Frangakis & Rubin, 2002) and principal score modeling (Jo & Stuart, 2009; Hill et al., 2002) approaches used in the context of randomized experiments. These studies share the goal of examining heterogeneity in treatment effects depending upon the counterfactual state or principal strata (e.g., whether an individual is a complier or an always-taker; in our setting, depending on whether a student would have worked or not in the absence of FWS). Hill et al. (2002), for example, use pre-treatment characteristics to predict principal strata membership in the context of a high-quality child care experiment, and then examine heterogeneity in impacts for subgroups defined by predicted care choices in the absence of treatment.

We take a different approach for two reasons. First, even with the benefit of randomization, principal strata membership can be difficult to predict using pre-treatment characteristics. As we will show below, it is much more difficult to predict the likelihood of working in general than it is to predict FWS participation. This means that we cannot define highly distinct subgroups based upon prior predicted probability of employment. Second, we do not have the benefit of randomized assignment to FWS. This is a limitation, but also a release in the sense that the assumptions needed to justify a comparison of FWS participants to matched non-participants are the same as those needed to justify a comparison of FWS participants to specific types of matched non-participants. Brand, Pfeffer, and Goldrick-Rab (2012) provide an example of how propensity scores, in a non-experimental setting, can be used to examine heterogeneity in treatment effects depending upon the counterfactual.¹⁰ In a sense, our approach combines the conceptual motivation underlying principal stratification with the techniques of propensity score matching with multi-valued treatments (Imbens, 2000).

¹⁰ Brand et al. (2012) examine outcomes for those who do and do not enroll in community college, then split the analysis to compare community college enrollees against matched four-year enrollees and against non-enrollees separately. Our setting is different because we examine the effect of a specific program (FWS), where the overall program effect is a weighted average of the effects estimated in each comparison.

We first estimate the average effect of FWS for all participants using standard PSM methods. Then, we repeat the analysis but condition the control group on the distinct counterfactuals of interest, as indicated by our conceptual framework: first, we re-estimate the PSM model comparing FWS participants with the most similar students who also worked during their first year, but at a non-FWS job; then we re-estimate the model comparing FWS participants with the most similar students who did not work at all.

Like the concept of mechanism experiments in RCTs, the goal of conditional counterfactual matching is to help uncover the mechanisms driving the observed treatment effects. Understanding mechanisms can help explain patterns of heterogeneity across different populations and contexts, and makes it easier to interpret disparate findings in prior literature. It can also help suggest implications for policy beyond the particular program under study (Ludwig, Kling, & Mullainathan, 2011). Finally, in non-experimental research, exploring mechanisms is a particularly important way of building a credible causal argument because true causal effects should operate through the predicted channels, while spurious correlations may or may not.

Incorporating PSM in this approach confers several advantages above simple ordinary least squares (OLS) regression. First, PSM can potentially control much more effectively for observable differences between participants and non-participants (Rosenbaum & Rubin, 1983; Rubin, 1973). In our context, FWS participants are a relatively small fraction of college enrollees, and the PSM approach ensures that they are compared only with those non-participants who are most observably similar. OLS regression, in contrast, could compare FWS participants with very different non-participants and thus might excessively rely on linear extrapolation to estimate treatment effects by (in other words, OLS can hide situations in which there is a lack of common support between the treatment and control group).

A second benefit of PSM is that it focuses attention on the selection process: who are the students who participate in FWS and how different are they from non-participants? To what extent is there overlap in the types of students that do and do not participate? The diagnostic value of PSM analysis can illuminate whether causal interpretations of the resulting estimates are justified, and can indicate in which direction any remaining biases are likely to go.

Of course, illuminating the problem of selection bias is not the same as solving it. To identify the causal effect of FWS, PSM is similar to OLS in that it relies on a selection-on-observables assumption. Specifically, any factors that jointly affect both FWS participation and subsequent outcomes must be included in the propensity score model.

While it is impossible to rule out unobservable biases, our analysis is well-suited to PSM for several reasons. First, the survey data we use includes a very rich set of background characteristics—much richer than is available in the administrative datasets often used in quasi-experimental analyses. Second, several important factors in the selection process are explicitly known (e.g., the role of family income, institution type, and unmet need) and can be included. Third, the fact that so many more students qualify for FWS than receive it creates a large pool of

potential control observations for matching. Finally and perhaps most crucially, the idiosyncratic nature of the FWS offer and take-up process mean that we can compare individuals with similar propensity scores and argue that the remaining variation in FWS status is plausibly random.

Implementation

To implement the PSM, we first model student participation in FWS using a probit regression:

$$(5) \quad \Pr(FWS = 1) = \Phi(X'\beta)$$

in which Φ is the cumulative distribution function of a standard normal distribution, and X is a vector of individual and institutional characteristics including gender, age, race/ethnicity, distance from home, parental education, high school GPA, SAT score (or equivalent), the student's EFC, a flag for Pell receipt, college selectivity, tuition, an interaction between EFC and tuition, the amount of institutional grants received by the student (a proxy of unobservable need/merit), enrollment size, and urbanicity.¹¹ For a complete list of variables included, see Table 3 in the results section. We follow Stuart (2010) in our liberal inclusion of predictor variables, given that there is little cost to including something that is not truly relevant, while the omission of an important factor could cause bias. Also note that by controlling for Pell Grant receipt and institutional grants, we are holding these aspects of the aid package constant based on the understanding that FWS is packaged after these other components; we do not control for student loans in our main specification, as this is potentially endogenous to FWS receipt.

The resulting parameters are used to construct each individual's propensity score. We then utilize the PSMATCH2 command in Stata, with a radius caliper of 0.01, with replacement and excluding observations with no common support, to evaluate the quality of the matching and estimate average treatment effects for the treated (ATT). Using a radius caliper of 0.01 allows us to match each FWS participant with any control observations within ± 0.01 of the participant's propensity score. This avoids the idiosyncratic matches that might result from simple one-to-one nearest neighbor matching. The outcomes we examine include first-year employment and hours, GPA, and student loan debt; persistence into the second year; on-time (four-year) BA completion; and graduate school enrollment, employment, earnings, and student loan debt measured at the six year follow-up.

We implement the PSM process first comparing FWS participants with all similar non-participants (we will refer to this as Model 1). We then estimate two separate conditional counterfactual models. In Model 2, we limit the sample to working students and re-run the FWS participation equation as well as the subsequent diagnostic, matching and outcomes analyses. Model 2 thus matches each FWS participant to the most similar working students. In Model 3,

¹¹ If a student did not file a FAFSA, the BPS estimates the EFC based on student and parent survey responses.

we limit the sample to FWS participants and non-working non-participants and repeat all steps of the PSM. Model 3 thus matches each FWS participant to the most similar non-working students.

Data and Sample

We use the two most recent cohorts of the Beginning Postsecondary Student (BPS) longitudinal study, BPS:96/01 and BPS:04/09, which together follow a nationally-representative sample of 30,545 U.S. college entrants who first began their post-secondary education in 1995–96 or 2003–04. These students were followed for six years after they first enrolled (through 2001 and 2009, respectively).

We limit the sample to 14,064 full-time dependent students who began at four-year baccalaureate-granting institutions, including 6,668 students from the BPS: 96/01 and 7,396 students from the BPS:04/09. We focus on these students because they represent the majority of FWS recipients (results including part-time, independent, and two-year attendees are similar and are included in Appendix Table A.3).¹² We further excluded 257 individuals missing information on first-year employment, 1,466 records with missing data on BA attainment after six years, and 146 records with missing institutional data. These filters yield a final sample (rounded to the nearest 10, per NCES requirements) of 12,200 individual student cases, 5,030 from the BPS: 96/01 and 7,170 from the BPS: 04/09. Roughly 2,070 students in this sample (17 percent) received FWS during their first year of enrollment.

Descriptive statistics on background characteristics and academic and labor market outcomes for this sample of 12,200 students are provided in Table 2. The first column represents the overall sample, while the subsequent columns provide the same statistics separately for FWS recipients, working non-recipients, and non-working students respectively. The table indicates that FWS recipients are quite different from other students. They have substantially lower incomes, and higher rates of financial aid receipt than non-recipients. Academically, they have higher high school grade point averages (GPAs) and SAT scores than other working students, but about the same GPAs and somewhat lower SATs than non-working students. They are nearly twice as likely as other students (working or non-working) to attend private institutions. Academic and labor market outcomes also differ across the three groups, with FWS students generally outperforming both other groups of students. These differences will be explored more systematically in our propensity score model below.

¹² Appendices A and B are available in a separate document available at capseecenter.org.

Table 2: Descriptive Statistics, Full-Time Dependent Students Entering Four-Year Institutions

Variable	BPS 1996–2001 and BPS 2004–2009 Sample Characteristics			
	Total	FWS Recipients	Working Non-Recipients	Non-Working Students
Student background and institutional characteristics				
Female (in %)	55.4	62.0	55.1	53.4
White, non-Hispanic (in %)	72.6	69.4	74.6	71.6
Age at entry (in years)	18.4	18.3	18.5	18.3
Parental income (in 2003 constant prices)	\$76,216	\$52,513	\$74,311	\$86,678
High school GPA \geq 3.00 (in %)	81.8	85.1	78.9	83.8
Derived SAT score	1026	1033	1002	1051
Received any Pell Grant (in %)	26.0	46.1	24.0	21.2
Average Pell Grant (zeros excluded)	\$2,256	\$2,315	\$2,128	\$2,371
Financial need after all grants	\$5,874	\$9,564	\$5,156	\$5,368
Any Student loan through year 1 (in %)	53.4	85.3	52.8	42.9
Student loan amount borrowed through year 1 (zeros excluded)	\$4,539	\$5,085	\$4,296	\$4,491
First institution: Public institution (in %)	61.8	33.3	67.5	65.5
First institution: Non/least selective (in %)	33.6	29.7	40.1	27.7
First institution: Selective (in %)	39.1	39.6	40.0	38.0
First institution: Very selective (in %)	27.3	30.7	20.0	34.3
First institution: Tuition and fees	\$9,562	\$14,478	\$7,940	\$9,621
First institution, total enrollment	14006	9279	14297	15338
Academic and labor market outcomes				
GPA in year 1	2.8	2.9	2.8	2.9
Still enrolled or attained during year 2 (in %)	93.2	96.2	90.7	94.9
Earned a BA within 4 years (in %)	36.7	46.0	29.8	41.0
Earned a BA within 6 years (in %)	65.0	72.1	58.2	70.1
Enrollment in graduate school in year 6	16.6	19.7	14.3	18.1
Any cumulative undergrad student loan through year 6 (in %)	62.9	90.7	63.4	52.5
Cumulative total student loan amount borrowed through year 6 (zeros excluded)	\$20,988	\$25,306	\$19,448	\$20,334
Enrolled or employed in year 6 (in %)	92.5	93.1	92.9	91.8
Employed in year 6, of those not enrolled (in %)	88.5	89.9	89.5	86.8
Total earnings from current job in year 6 (zeros excluded)	\$36,317	\$35,247	\$35,846	\$37,277
Sample size (unweighted)	12,200	2,070	5,240	4,890

Note. Authors' calculations using BPS:96/01 and 04/09 restricted-use data files. The sample includes only full-time and dependent students entering a four-year institution in their first academic year for 1995–96 and 2003–04. The BPS tabulations use BPS longitudinal sampling weights (wta000). Sample sizes may vary for certain variables due to missing values. *Financial need after grants* refers to the student budget minus EFC, federal, state and institutional grants. *Derived SAT* refers to either the sum of SAT verbal and math scores or to the ACT composite score

converted to an estimated SAT combined score. 1995 figures are in constant 2003 prices. Unweighted sample sizes are rounded to the nearest 10.

5. Results

Modeling the Propensity Score: Who Participates in FWS?

Table 3 displays the marginal effects from a probit regression of FWS participation (equation [5] above).¹³ To highlight how the selection process into FWS may differ from the selection process into student employment more broadly, we also show results from a probit regression of any student employment on the same set of independent variables. Being male, having higher family income (as measured by the EFC) and attending a public institution all significantly reduce the likelihood of both FWS participation and any student employment. However, selection into FWS and any employment significantly vary across many other dimensions. For example, minority students are less likely to work in general, but no more or less likely to receive FWS.¹⁴ Students at very selective institutions are less likely to work in general, but more likely to receive FWS. Students who receive Pell Grants or institutional grants—which we interpret as a proxy for unobservable need and merit—are no more or less likely to work in general, but are more likely to receive FWS. Interestingly, when we run equivalent OLS regressions we find that it is much easier to predict FWS employment than it is to predict student employment generally: the r-squared statistics on the equivalent OLS regressions were 0.21 and 0.08, respectively.¹⁵

¹³ OLS regression produced very similar results.

¹⁴ This is consistent with the hypothesis that minority students face structural barriers to employment, though it is not the only possible explanation.

¹⁵ With binary dependent variables, the R-squared from OLS regression can be interpreted as the difference in average *predicted* FWS participation (or employment) rates between those who actually participated (or worked) and those who actually did not (Gronau, 1998).

Table 3. Federal Work Study Participation and Any Employment

Variable	Dprobit: P(FWS)		Dprobit: P(Emp)	
	B	(SE)	B	(SE)
Female	0.009	(0.003) ***	0.031	(0.010) ***
Age	-0.035	(0.056)	-0.167	(0.189)
Age squared	0.001	(0.001)	0.005	(0.005)
Black, non-Hispanic	-0.000	(0.007)	-0.109	(0.022) ***
Hispanic	-0.002	(0.006)	-0.033	(0.019) *
Other race	-0.007	(0.005)	-0.065	(0.017) ***
Distance from first institution	-0.000	(0.000)	-0.000	(0.000) **
Mother is a HS grad only	0.008	(0.008)	0.026	(0.025)
Mother has some college	0.005	(0.008)	0.007	(0.025)
Mother has a BA	0.002	(0.008)	-0.010	(0.026)
Father is a HS grad only	0.003	(0.006)	0.005	(0.022)
Father has some college	-0.003	(0.007)	-0.014	(0.023)
Father has a BA	-0.004	(0.007)	-0.040	(0.023) *
pmomed_m	-0.016	(0.011)	-0.012	(0.043)
pdaded_m	0.007	(0.013)	-0.046	(0.042)
HS GPA (2.5–2.9)	0.013	(0.008)	0.023	(0.023)
HS GPA (3.0–3.4)	0.017	(0.007) **	0.010	(0.020)
HS GPA (3.5–4.0)	0.015	(0.008) **	0.004	(0.021)
hsgpacat_m	-0.009	(0.009)	-0.029	(0.027)
SAT score	-0.000	(0.000)	-0.000	(0.000)
Enrollment delay	-0.008	(0.008)	0.070	(0.025) ***
enrdelay_m	0.013	(0.065)	-0.002	(0.186)
EFC	-0.005	(0.001) ***	-0.008	(0.001) ***
EFC squared	-0.000	(0.000) ***	0.000	(0.000) ***
Financial need after all grants	0.000	(0.001)	-0.006	(0.002) ***
Received any Pell Grant	0.028	(0.006) ***	0.001	(0.014)
Institutional grants	0.004	(0.001) ***	0.002	(0.002)
EFC × tuition	0.000	(0.000) ***	-0.000	(0.000) ***
First institution: Selective	0.012	(0.006) **	-0.004	(0.017)
First institution: Very selective	0.015	(0.008) *	-0.094	(0.020) ***
First institution: Public institution	-0.018	(0.009) *	-0.051	(0.025) **
First institution: Tuition and fees	0.008	(0.001) ***	0.004	(0.003)
Tuition squared	-0.000	(0.000) ***	0.000	(0.000) **
First institution, total enrollment	-0.001	(0.001)	-0.004	(0.002) **
Total enrollment squared	0.000	(0.000)	0.000	(0.000) *
Inst in large city	-0.024	(0.007) ***	0.049	(0.020) **
Inst in mid-sized city	-0.013	(0.006) **	0.032	(0.018) *
Inst in urban fringe of large city	-0.017	(0.007) ***	0.087	(0.021) ***
insturb_m	-0.025	(0.011) **	-0.075	(0.048)
BPS2003	0.003	(0.005)	-0.027	(0.014) **
Sample size (unweighted)	12200		12200	
R-squared from OLS	0.208		0.074	

Note: Authors' calculations using BPS:96/01 and 04/09 restricted-use data files. The dependent variable is if

the student received any Federal Work Study (FWS) grant during his first academic year participation. Standard errors are in parenthesis. The sample covers only full-time and dependent students entering a four-year institution in their first academic year for 1995–96 and 2003–04. Model 1 reports OLS coefficient estimates with standard errors clustered by institution. Model 2 reports dprobit estimates with standard errors clustered by institution. 1995 figures are in constant 2003 prices. Unweighted sample sizes are rounded to the nearest 10.

Common Support and Match Quality

Figures 2.1 through 2.3 provide evidence of common support. We first show the distributions of treated (FWS) and control group observations across the range of estimated propensity scores for our main model (matching FWS participants to all non-participants), and then for models 2 and 3 (matching FWS participants against other working students or non-working students, respectively). For each model, only a few observations with high propensity scores lie outside the range of common support. Because the control group is roughly five times larger than the treatment group (for our main model), even at the high range of propensity scores we still have numerous control observations available for matching (e.g., we have 32 control observations with p -scores of 0.75 or higher). Thus, the common support condition appears to pose no barrier to the analysis.

Fig 2.1: Common Support-Model 1

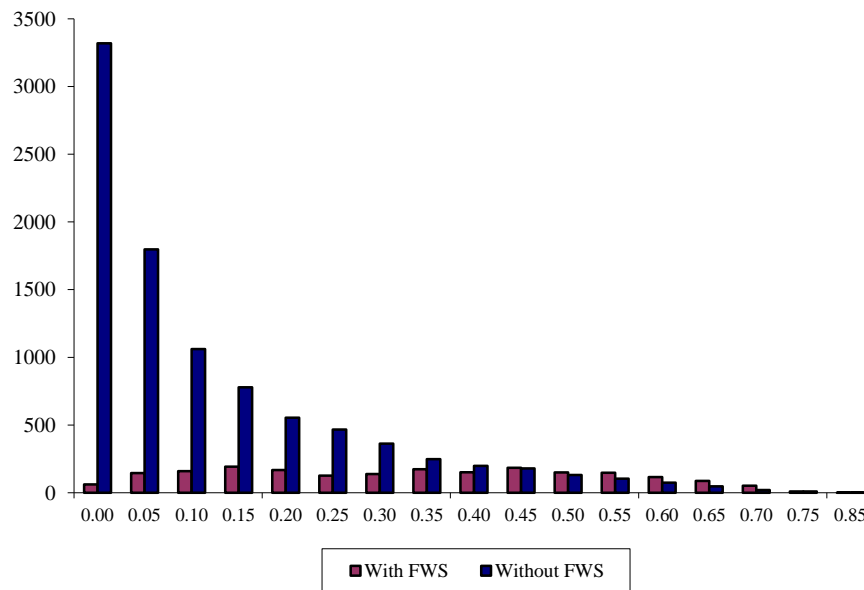


Fig 2.2: Common Support-Model 2

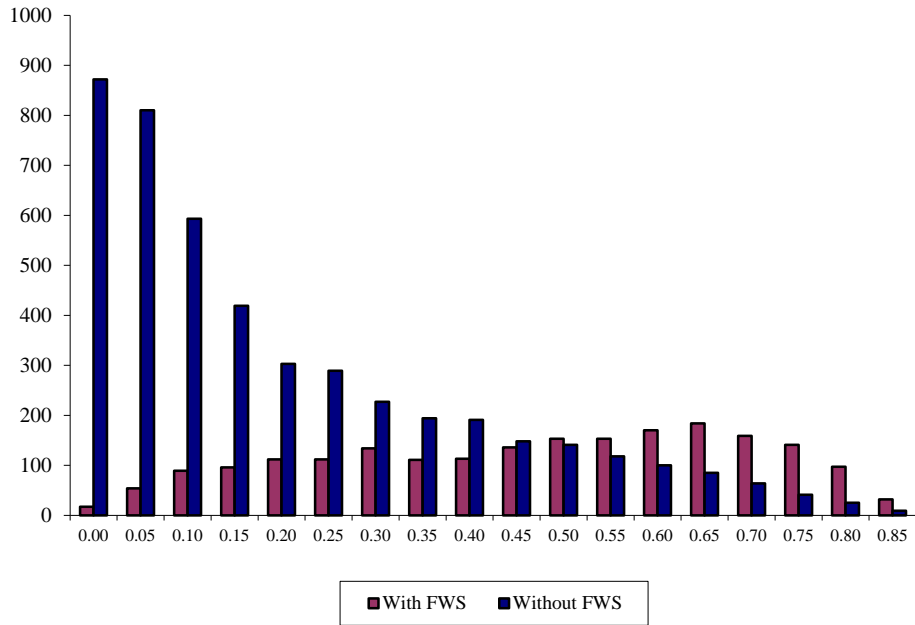
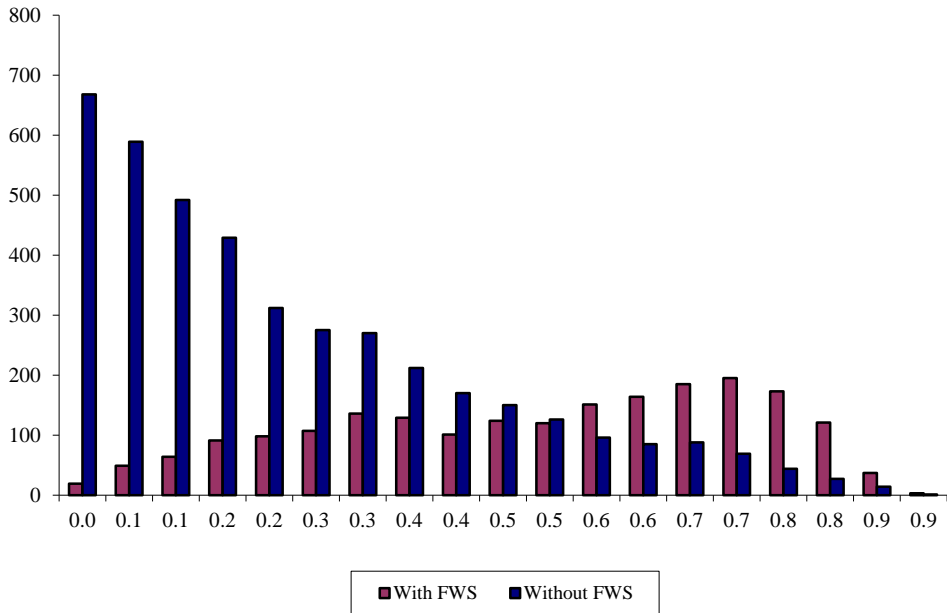


Fig 2.3: Common Support-Model 3



Figures 3.1 through 3.3 provide evidence that after matching, our treated and control group observations are well-balanced on a range of observable characteristics. Each line in the

graph represents one of the background characteristics from Table 3. The circles indicate the standardized bias on that characteristic that exists in the unmatched sample (i.e., the difference in means between treated and untreated observations, expressed in terms of standard deviations). The x's indicate the standardized bias in the matched sample. For all three models, standardized bias for virtually all characteristics is reduced to below 5 percent, and in most cases below 3 percent, levels that are considered acceptable in the matching literature (Caliendo & Kopeinig, 2008). The sole exception is that in model 3, financial need after all grants remains about 6 percent bigger among FWS participants than among working students (the difference was 67 percent prior to matching). Given the large number of characteristics evaluated and the dramatic reductions in bias for key characteristics, we consider all models to be very well balanced. Full results of these balance tests are provided in Appendix Tables A1.1–A1.3.

Fig 3.1: Bias Reduction Graph-Model 1

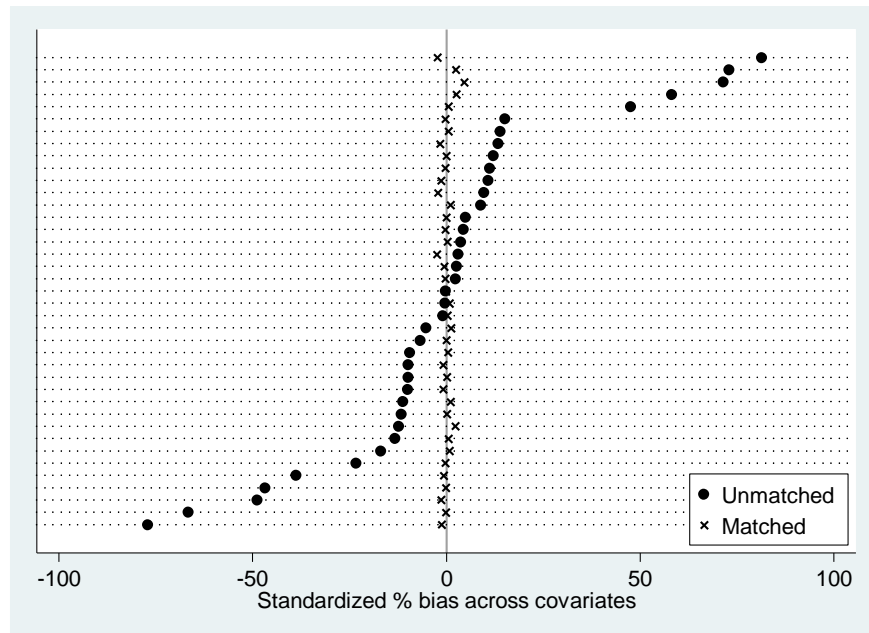


Fig 3.2: Bias Reduction Graph-Model 2

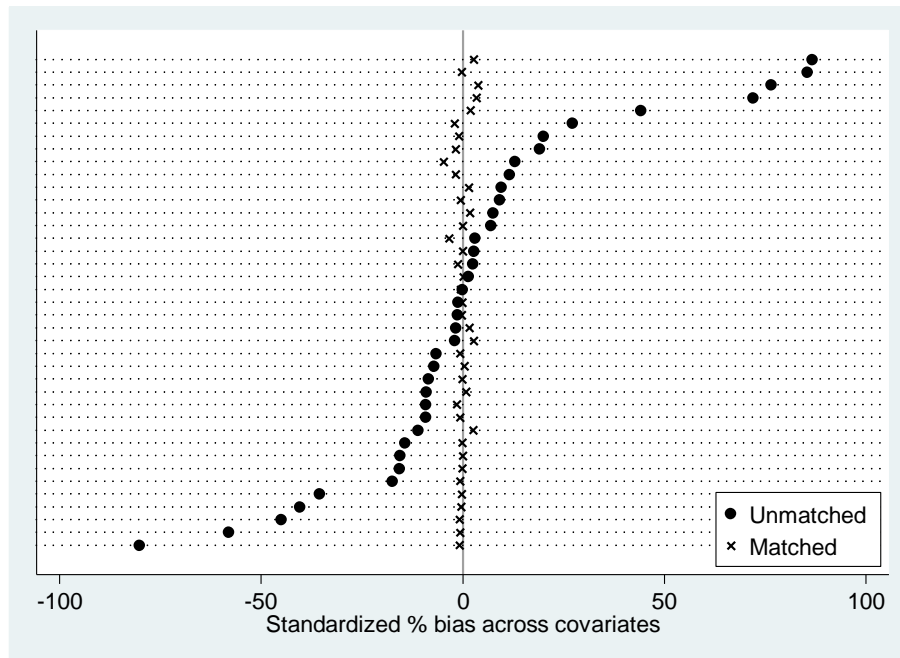
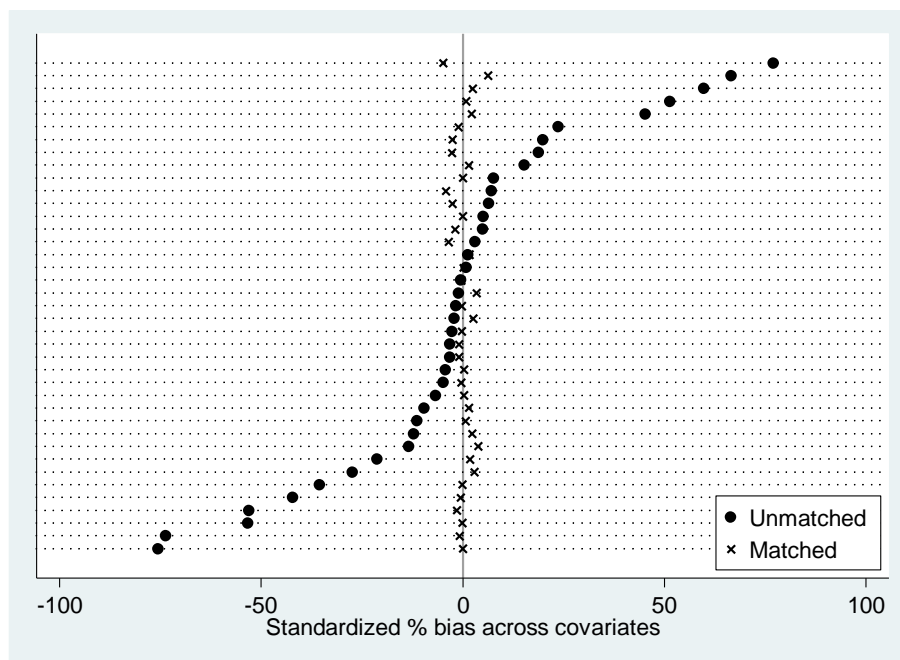


Fig 3.3: Bias Reduction Graph-Model 3



Main Results: Estimated Effects of FWS Participation

Our main results are presented in Table 4. The first notable result is that overall, FWS participants are “only” 48 percentage points more likely to be employed during the school year than non-participants. While this is a large and statistically significant impact, it also indicates that a slight majority of participants would have worked at another job even if they had not received FWS. Overall, recipients work 6 hours more per week than non-recipients; but this is a combination of a 1.5-hour reduction in weekly hours for those participants who would have worked at another job anyway (Model 2), and a 15-hour increase in hours for those who would not have worked at all in the absence of FWS (Model 3).

The second intriguing pattern is that for the full sample, there are no significant negative effects, and several significant positive effects on academic and labor market outcomes. However, there is a clear pattern of the positive academic effects being most concentrated among those who would have worked at another job in the absence of FWS (Model 2), while the labor market effects are more pronounced among those who would not have worked otherwise (Model 3). For example, FWS participants were 5.2 percentage points more likely than similar working students to earn a BA within six years, but the difference was only 1.4 percentage points and not significant when FWS students were compared against non-working students. FWS participants were 3.7 percentage points more likely than similar non-working students to be employed after six years, but the difference was an insignificant 1.4 percentage points when FWS students were compared against similar working students.¹⁶ Interestingly, the effects of FWS participation on first-year GPA—one of the only outcomes examined by prior research—do appear to be significantly negative when compared against not working at all (-0.055 GPA points). But this does not carry over into negative effects on longer-term academic outcomes such as persistence and graduation. Earnings effects are not significant in any of the models, though these are still measured very early in the career trajectory.

It seems unlikely that the large effects on BA completion in Model 2 could be driven solely by the 1.5-hour reduction in weekly hours. Thus, we ran additional regressions using Model 2 to examine impacts on available job characteristics for FWS versus non-FWS working students (not shown in Table 4). First, we find a significant negative impact on estimated hourly wages (of $-\$0.61$)—an unusual finding for a wage subsidy. For a 75-percent wage subsidy to result in students receiving lower wages strongly suggests that FWS jobs must offer significant compensatory amenities. As could be predicted from the descriptive differences in Table 1, we find large impacts on the probability that a student works on-campus (a 52 percentage point increase) as well as a 7.4 percentage point impact on the likelihood that the job is related to the student’s major. Both impacts are highly significant.

¹⁶ We focus here on the employment effects for students who were not enrolled in school in Year 6 (there is no impact on enrollment in Year 6). Unconditional estimates are also provided in Table 4.

Table 4: Impact of Federal Work Study (FWS) on Academic and Labor Market Outcomes, Overall and By Counterfactual

Variable	Model 1		Model 2: FWS vs. Working Students		Model 3: FWS vs. Non-Working Students	
	B	S.E.	B	S.E.	B	S.E.
Any employment during school year	0.478	(.01) ***				
Total hours worked per week in year 1	6.236	(.336) ***	-1.539	(.375) ***	15.041	(.24) ***
GPA in year 1	-0.027	(.022)	-0.027	(.027)	-0.055	(.028) **
Still enrolled or attained during year 2	0.011	(.006) *	0.012	(.008)	0.005	(.007)
Number of months enrolled through year 6	0.503	(.321)	0.852	(.405) **	0.003	(.387)
Earned a BA within 4 years	0.011	(.014)	0.029	(.016) *	-0.005	(.018)
Earned a BA within 6 years	0.032	(.013) **	0.052	(.016) ***	0.014	(.016)
Enrolled or employed in year 6	0.015	(.008) **	0.011	(.009)	0.023	(.01) **
Enrolled (graduate or undergraduate) in year 6	-0.006	(.013)	-0.010	(.016)	-0.008	(.017)
Employed in year 6, of those not enrolled	0.024	(.011) **	0.014	(.013)	0.037	(.014) **
Employed in year 6	0.021	(.014)	0.021	(.016)	0.031	(.018) *
Log of total earnings from current job in year 6	-0.023	(.021)	-0.028	(.024)	-0.018	(.027)
Enrollment in graduate school in year 6	0.007	(.011)	0.007	(.013)	0.001	(.014)
Any undergraduate student loan, cumulative through year 6	0.162	(.011) ***	0.142	(.014) ***	0.186	(.016) ***
Undergraduate student loan amount, cumulative through year 6	\$6,263	(561) ***	\$6,003	(639) ***	\$6,761	(690) ***
Any student loan in year 1	0.209	(.012) ***	0.180	(.015) ***	0.240	(.016) ***
First year student loan amount	\$1,270	(118) ***	\$1,047	(138) ***	\$1,501	(145) ***
First year FWS amount	\$1,521	(18) ***	\$1,521	(18) ***	\$1,521	(18) ***
First year Pell Grant amount	\$16	(38)	\$28	(43)	-\$24	(46)
First year total aid amount	\$2,814	(227) ***	\$2,687	(259) ***	\$2,896	(277) ***

Note. Authors' calculations using BPS:96/01 and 04/09 restricted-use data files. This table shows results for our baseline model as well as the first two alternative specifications, Models 2 and 3. Model 2 compares FWS recipients with working students (non-recipients), while Model 3 compares them with non-working students. All models use a two-step process in which logit command is used to estimate the propensity score and then PSMATCH2 command with radius caliper of 0.01, with replacement and excluding observations outside of common support, is used to execute the matching and estimate ATT. No weights. Continuous variables are imputed to the mean if missing; dummy variables are imputed to zero; missing data flags are included for all variables with missing data. No clustering. The same covariates used in Table 3 are used for these models. ATT estimates. Propensity scores are re-estimated separately for Model 2 and Model 3, which use restricted samples. Sample is restricted to full-time, dependent students enrolled at four-year institutions with non-missing data on institution ID, FWS receipt (first year), first-year employment, and BA attainment within six years. The unweighted sample size, rounded to the nearest 10, for total hours worked per week in year 1 and for Model 1 is 11,420 for Model 2 is 6,700 and for Model 3 is 6,190. This sample size is different for the variables conditioned on employment and cumulative debt.

One persistent and unexpected finding is that FWS recipients appear to accumulate substantially higher debt burdens than either comparison group of students. Focusing on Model 1, FWS participants are 21 percentage points more likely to borrow in their first year and this persists as a 16 percentage point increase in the likelihood of having borrowed after six years. Cumulative undergraduate debt is \$6,263 higher (including zeros) for FWS participants than similar non-participants. These patterns do not vary much depending upon whether we compare participants with other working or non-working students. Note that the impact on first year total aid is approximately equal to the sum of the impacts on first year FWS and loan aid.

It is striking that such strong debt effects remain even after matching students on individual characteristics including EFC (the primary determinant of eligibility for student aid including loans) and financial need after grants, as well as on institutional characteristics such as selectivity, tuition, and tuition squared. To ensure that these differences are not simply a flag for unobservable bias, we perform a robustness check in which we add a first-year borrowing dummy and first-year loan amounts to our matching model.¹⁷ The results of this check, shown in Table 5, indicate that it makes little difference. The overall pattern of results is qualitatively similar to that shown in Table 4. By construction there is no longer any effect on first-year loan amounts, but the impacts on longer-term debt accumulation remain statistically significant (though they are smaller). The magnitude and persistence of these results suggest that student loans, rather than being a substitute for student employment, may be packaged with FWS funds by aid officers in a formulaic way. This can be partly, but not fully explained by the fact that students must file a FAFSA to access either program. When we limit the analysis to students who file a FAFSA, the magnitude of the loan impacts is cut approximately in half, but still remain large and highly significant (the pattern of academic and labor market results remain unchanged; see Appendix Table A8).

¹⁷ First year loan amounts are properly considered as an outcome variable and thus generally should not be included in the matching model; however, if this variable is also a proxy for unobservable bias it may be preferable to control for it.

Table 5: Impact of Federal Work Study (FWS), with Loans Included as a Control Variable

Variable	Model 1		Model 2: FWS vs. Working Students		Model 3: FWS vs. Non-Working Students	
	B	S.E.	B	S.E.	B	S.E.
Any employment during school year	0.458	(.011) ***				
Total hours worked per week in year 1	5.850	(.352) ***	-1.663	(.396) ***	15.036	(.24) ***
GPA in year 1	-0.009	(.023)	0.002	(.028)	-0.024	(.029)
Still enrolled or attained during year 2	0.010	(.006)	0.013	(.009)	0.004	(.008)
Number of months enrolled through year 6	0.558	(.335) *	0.885	(.43) **	0.208	(.407)
Earned a BA within 4 years	0.021	(.015)	0.042	(.017) **	0.013	(.019)
Earned a BA within 6 years	0.039	(.014) ***	0.059	(.017) ***	0.032	(.017) *
Enrolled or employed in year 6	0.013	(.008) *	0.010	(.009)	0.028	(.01) ***
Enrolled (graduate or undergraduate) in year 6	-0.004	(.014)	0.005	(.016)	-0.006	(.018)
Employed in year 6, of those not enrolled	0.021	(.012) *	0.013	(.013)	0.046	(.015) ***
Employed in year 6	0.017	(.015)	0.004	(.017)	0.033	(.019) *
Log of total earnings from current job in year 6	-0.023	(.022)	-0.025	(.025)	-0.004	(.029)
Enrollment in graduate school in year 6	0.015	(.012)	0.025	(.013) *	0.013	(.015)
Any undergraduate student loan, cumulative through year 6	0.043	(.012) ***	0.036	(.015) **	0.049	(.018) ***
Undergraduate student loan amount, cumulative through year 6	\$2,307	(581) ***	\$2,521	(671) ***	\$2,148	(730) ***
Any student loan in year 1	0.004	(.013)	0.006	(.016)	0.003	(.017)
First year FWS amount	\$1,522	(18) ***	\$1,521	(18) ***	\$1,521	(18) ***
First year Pell Grant amount	-\$25	(39)	-\$17	(45)	-\$85	(47) *
First year total aid amount	\$1,610	(235) ***	\$1,638	(271) ***	\$1,478	(293) ***

Note. Authors' calculations using BPS:96/01 and 04/09 restricted-use data files. These models include amount borrowed and any loan taken during the first year of enrollment in the first stage or logit model. This table shows results for our baseline model as well as the first two alternative specifications, Models 2 and 3. Model 2 compares FWS recipients with working students (non-recipients) while Model 3 compares them with non-working students. All models use a two-step process in which logit command is used to estimate the propensity score and then PSMATCH2 command with radius caliper of 0.01, with replacement and excluding observations outside of common support, is used to execute the matching and estimate ATT. No weights. Continuous variables are imputed to the mean if missing; dummy variables are imputed to zero; missing data flags are included for all variables with missing data. No clustering. The same covariates used in Table 3 are used for these models. ATT estimates. Propensity scores are re-estimated separately for Model 1 and Model 2, which use restricted samples. Sample is restricted to full-time, dependent students enrolled at four-year institutions with non-missing data on institution ID, FWS receipt (first year), first-year employment, and BA attainment within six years. The

unweighted sample size, rounded to the nearest 10, for total hours worked per week in year 1 and for Model 1 is 12,190, for Model 2 is 6,960 and for Model 3 is 6,950. This sample size is different for the variables conditioned on employment and cumulative debt.

Robustness to Alternative Specifications

One concern with propensity score estimation is that if two individuals have such similar propensities to participate, what explains why some do and some do not? Among students with low propensity scores, factors exogenous to students' own potential outcomes may plausibly determine who gets access and who does not (e.g., idiosyncratic variation in institutional funding or disbursement policies). We may worry more about why some of those with high propensity scores do not participate; in this case the difference is more likely to be the student's own choice. For this reason, we run a robustness check limiting the sample to the "thick support" region, including only observations with p -scores less than 0.60. The results are presented in Table 6. The pattern of findings is similar to the main results in Table 4, but generally more positive and more statistically significant.

Results from additional robustness checks can be found in Appendix A. Propensity score estimation requires numerous decisions to be made, often without much clear guidance from the literature. Reassuringly, our results were highly consistent across a range of alternative specifications, including:

- Using a 0.2 caliper instead of 0.01, using nearest neighbor and five-nearest-neighbors with replacement (Appendix Tables A2.1–2.3)
- Including independent, part-time, and two-year enrollees (Appendix Table A3)
- Discarding observations with missing data (Appendix Table A4)
- Clustering standard errors by institution after matching (Appendix Table A5)
- Abadie and Imbens (2012) standard error adjustments (Appendix Tables A6.1–A6.2)
- Using OLS instead of PSM (Appendix Table A7)

This is not to say that the choice of specification makes no difference; indeed there are small differences in the magnitudes of given effects in some cases (for example, FWS impacts appear somewhat more positive when estimated using OLS). Overall, however, the big picture remains the same across these specifications. This should give readers confidence that our findings are not the result of cherry-picking results or massaging the data.

Table 6: Federal Work Study (FWS) Impact on Student Outcomes for Model 1, Model 2, and Model 3 for Full-Time Dependent Students Entering Four-Year Institutions With a Thicker Common Support Area

Variable	Model 1		Model 2: FWS vs. Working Students		Model 3: FWS vs. Non-Working Students	
	B	S.E.	B	S.E.	B	S.E.
Any employment during school year	0.474	(.09) ***				
Total hours worked per week in year 1	6.067	(.342) ***	-2.410	(.397) ***	15.693	(.325) ***
GPA in year 1	-0.029	(.023)	-0.034	(.028)	-0.004	(.029)
Still enrolled or attained during year 2	0.013	(.006) **	0.022	(.009) ***	0.011	(.008)
Number of months enrolled through year 6	0.646	(.332) *	1.230	(.437) ***	0.301	(.419)
Earned a BA within 4 years	0.018	(.015)	0.031	(.017) *	-0.016	(.018)
Earned a BA within 6 years	0.042	(.013) ***	0.076	(.017) ***	0.022	(.017)
Enrolled or employed in year 6	0.015	(.008) **	0.015	(.009) *	0.031	(.01) ***
Enrolled (graduate or undergraduate) in year 6	0.003	(.014)	-0.002	(.016)	0.005	(.017)
Employed in year 6, of those not enrolled	0.025	(.011) **	0.022	(.013) *	0.049	(.015) ***
Employed in year 6	0.012	(.014)	0.017	(.017)	0.026	(.018)
Log of total earnings from current job in year 6	-0.017	(.021)	-0.025	(.026)	-0.012	(.028)
Enrollment in graduate school in year 6	0.016	(.012)	0.026	(.013) *	0.012	(.015)
Any undergraduate student loan, cumulative through year 6	0.170	(.011) ***	0.151	(.014) ***	0.246	(.015) ***
Undergraduate student loan amount, cumulative through year 6	\$6,124	(580) ***	\$5,811	(685) ***	\$7,526	(734) ***
Any student loan in year 1	0.207	(.012) ***	0.190	(.015) ***	0.278	(.016) ***
First year student loan amount	\$1,141	(120) ***	\$1,006	(142) ***	\$1,383	(150) ***
First year FWS amount	\$1,516	(19) ***	\$1,480	(24) ***	\$1,499	(25) ***
First year Pell Grant amount	-\$1	(39)	-\$46	(46)	-\$71	(47)
First year total aid amount	\$3,022	(220) ***	\$2,958	(227) ***	\$3,687	(254) ***

Note. Authors' calculations using BPS:96/01 and 04/09 restricted-use data files. This table shows results for our baseline model as well as the first two alternative specifications, Models 2 and 3. Model 2 compares FWS recipients with working students (non-recipients) while Model 3 compares them with non-working students. All models use a two-step process in which logit command is used to estimate the propensity score, then PSMATCH2 command with radius caliper 0.01, with replacement and excluding observations outside of common support, is used to execute the matching and estimate ATT. No weights. Continuous variables are imputed to the mean if missing; dummy variables are imputed to zero; missing data flags are included for all variables with missing data. No clustering. The same covariates used in Table 3 are used for

these models. ATT estimates. Propensity scores are reestimated separately for Model 2 and Model 3, which use restricted samples. Sample is restricted to full-time, dependent students enrolled at four-year institutions with non-missing data on institution ID, FWS receipt (first year), first-year employment, and BA attainment within six years. Sample is also restricted to a thicker common support area, only observations with pcores greater than zero and less than or equal to 0.6. The unweighted sample size, rounded to the nearest 10, for total hours worked per week in year 1 and for Model 1 is 11,000, for Model 2 is 5,490 and for Model 3 is 4,890. This sample size is different for the variables conditioned on employment and cumulative debt.

Sensitivity to Unobservable Selection

As with any non-experimental analysis, it is impossible to rule out the possibility that students who select into FWS differ in unobservable ways from those that do not, even after accounting for the extensive controls in our model. Prior work has proposed several ways to nonetheless quantify how “big” this unobservable selection would have to be in order to completely explain the observed pattern of effects. For example, using Rosenbaum’s (2002) approach to quantify the potential for hidden bias (as described in Becker & Caliendo [2007]), we estimate that the overall effect on BA completion after six years could lose significance if those who were going to earn a BA anyway are 50 percent more likely to participate in FWS than those who were not, after controlling for observable differences. The results from our conditional counterfactual models make this degree of hidden bias seem unlikely, because the overall BA effect is driven by those who would have worked anyway. For this comparison (Model 2), the difference in FWS participation would have to be 93 percent to reduce the effect to insignificance.

We also implemented the bounding approach proposed by Oster (2014), which makes the assumption that unobserved selection is proportional to observed selection. This analysis requires r-squared statistics from OLS regression, so we ran OLS linear probability models on our matched sample with the same set of control variables as our main propensity score model. Oster shows that the sensitivity of coefficients to unobservables can, under certain assumptions, be estimated using two pieces of information: (1) the difference in coefficients between an unadjusted model and a model including observable control variables, and (2) the difference in r-squared statistics between the adjusted and unadjusted models. The intuition is that coefficient stability across models can be indicative of robustness when combined with information on r-squared movements (if the r-squared moves little, then observed controls are not very informative). We find that our overall impact on BA completion within six years is robust to unobserved factors that are up to 75 percent as valuable as the entire set of controls actually included; in Model 2 (FWS vs. other working students) the impact is robust to unobserved factors up to 2.1 times as important as the included set of controls.¹⁸ For employment outcomes,

¹⁸ We assumed a maximum possible r-squared of 2.2 times the full-model r-squared, or about 0.40 ($= 2.2 \times 0.18$) in the case of BA completion after six years. Note the application of Oster’s method to a binary outcome is somewhat speculative. In a personal communication, Oster indicated that because her procedure is defined in terms of variances, it should be possible to apply to the binary-variable case although the interpretation of r-squareds in this case is somewhat different.

adding controls makes the estimates more positive, so unobservable bias would have to go in the opposite direction to explain any of the observed effects.

6. Subgroup Analysis

Our conceptual framework implies that the effects of FWS participation may vary across subgroups depending upon the counterfactual (i.e., the probability of working in the absence of FWS, and the characteristics of counterfactual jobs). We examined results for subgroups defined by income, SAT score, local unemployment rate, urbanicity, and gender. While in many cases standard errors were too large to draw strong conclusions about subgroup differences, highlights of these analyses are discussed below. Full subgroup results are available in the appendix.

Parental Income and Student SAT Score

We split our sample into subgroups by family income and SAT score, using the median values among FWS participants. Our hypothesis is that lower-income and lower-SAT students may be more likely to work even in the absence of FWS, and thus may experience more positive (or less negative) academic impacts than higher-income and higher-SAT students. Our results are supportive of this hypothesis, and the differences in subgroup estimates are in many cases large enough to be statistically significant. Focusing on the main Model 1, Table 7 shows that high SAT students who receive FWS experience a larger impact than low SAT students on both the probability of working (an increase of 50.3 percentage points versus 42.9) and average weekly hours (6.4 versus 5.6). High SAT recipients receive no academic benefit overall, while lower-SAT recipients persist longer and have 6.8 percentage point higher rates of BA completion after six years than low-SAT non-recipients. Lower-SAT participants also appear to have better post-college outcomes, though the differences here are not as large.

Similarly, results in Table 8 for higher versus lower-income participants indicate almost uniformly more positive effects for lower-income participants. For high-income students there is also a significant negative impact on first-year GPA (-0.06). It is also notable that the student debt impacts are twice as big for the high-income and high-SAT groups as for their low-income and low-SAT counterparts; it is possible that the sole positive effect for high-income and high-SAT students—on employment in year 6—may be related to loan debt rather than being a productive impact of the program. (See Appendix Tables B.1 and B.2 for additional results on Models 2 and 3.)

Table 7. Impact of Federal Work-Study (FWS), by SAT Subgroup (Model 1 Only)

Outcome	Model 1: Higher SAT		Model 1: Lower SAT	
	B	S.E.	B	S.E.
Any employment during school year	0.503	(.015) ***	0.429	(.014) ***
Total hours worked per week in year 1	6.396	(.438) ***	5.551	(.537) ***
GPA in year 1	-0.015	(.027)	-0.033	(.034)
Still enrolled or attained during year 2	0.008	(.007)	0.015	(.011)
Number of months enrolled through year 6	0.180	(.368)	1.090	(.565) *
Earned a BA within 4 years	0.005	(.02)	0.022	(.02)
Earned a BA within 6 years	0.013	(.016)	0.068	(.021) ***
Enrolled or employed in year 6	0.017	(.011)	0.023	(.011) **
Enrolled (graduate or undergraduate) in year 6	0.003	(.019)	-0.020	(.019)
Employed in year 6, of those not enrolled	0.033	(.016) **	0.026	(.016)
Employed in year 6	0.014	(.02)	0.043	(.021) **
Log of total earnings from current job in year 6	-0.022	(.031)	-0.028	(.03)
Enrollment in graduate school in year 6	0.007	(.018)	0.014	(.014)
Any undergraduate student loan, cumulative through year 6	0.189	(.017) ***	0.102	(.017) ***
Undergraduate student loan amount, cumulative through year 6	\$7,778	(796) ***	\$4,371	(830) ***
Any student loan in year 1	0.243	(.018) ***	0.144	(.018) ***
First year student loan amount	\$1,455	(172) ***	\$910	(174) ***
First year FWS amount	\$1,594	(23) ***	\$1,457	(28) ***
First year Pell Grant amount	\$33.3	(47)	-\$7	(62)
First year total aid amount	\$3,046	(344) ***	\$2,357	(296) ***

Note. Authors' calculations using from BPS:96/01 and 04/09 restricted-use data files. This table shows results for our baseline model comparing FWS recipients against all matched non-recipients. The same covariates used in Table 3 were used for these models. ATT estimates. Sample is restricted to full-time, dependent students enrolled at four-year institutions. Higher-SAT students were identified as those who scored at or above the median SAT for FWS students, 1030 points. The unweighted sample size, rounded to the nearest 10, for total hours worked per week in year 1 and for Model 1 is 6,140 for higher-SAT sample and 5,680 for the lower-SAT sample. This sample size is smaller for the variables conditioned on employment and cumulative debt.

Table 8. Federal Work Study (FWS) Impact on Student Outcomes for Full-Time Dependent High/Low Income Students Entering Four-Year Institutions

Variable	Model 1: High Income		Model 1: Low Income	
	B	S.E.	B	S.E.
Any employment during school year	0.502	(.013) ***	0.447	(.015) ***
Total hours worked per week in year 1	5.964	(.439) ***	6.529	(.543) ***
GPA in year 1	-0.055	(.03) *	0.007	(.035)
Still enrolled or attained during year 2	0.006	(.007)	0.013	(.011)
Number of months enrolled through year 6	-0.087	(.405)	0.873	(.551)
Earned a BA within 4 years	0.003	(.02)	0.020	(.021)
Earned a BA within 6 years	0.020	(.017)	0.048	(.021) **
Enrolled or employed in year 6	0.015	(.01)	0.023	(.012) *
Enrolled (graduate or undergraduate) in year 6	-0.019	(.019)	0.010	(.02)
Employed in year 6, of those not enrolled	0.018	(.015)	0.030	(.018) *
Employed in year 6	0.034	(.02) *	0.013	(.022)
Log of total earnings from current job in year 6	-0.033	(.03)	-0.019	(.033)
Enrollment in graduate school in year 6	-0.014	(.017)	0.036	(.016) **
Any undergraduate student loan, cumulative through year 6	0.198	(.016) ***	0.100	(.017) ***
Undergraduate student loan amount, cumulative through year 6	\$7,640	(834) ***	\$3,928	(780) ***
Any student loan in year 1	0.242	(.017) ***	0.148	(.019) ***
First year student loan amount	\$1,347	(180) ***	\$1,004	(162) ***
First year FWS amount	\$1,508	(24) ***	\$1,564	(27) ***
First year Pell grant amount	-\$12.9	(12)	\$90	(67)
First year total aid amount	\$2,910	(329) ***	\$2,498	(327) ***

Note. Authors' calculations using BPS:96/01 and 04/09 restricted-use data files. This table shows results for our baseline model comparing FWS recipients to all matched non-recipients. The same covariates used in Table 3 were used for these models. ATT estimates. Sample is restricted to full-time, dependent low-income students enrolled at four-year institutions. Low-income students were identified as those whose annual household income is at or below the median income of FWS recipients, approximately \$49,000. The unweighted sample size, rounded to the nearest 10, for total hours worked per week in year 1 is 7,740 for high-income students and 4,060 for low-income students. This sample size is smaller for the variables conditioned on employment and cumulative debt.

Local Unemployment Rates and Urbanicity

The motivation behind this analysis is that students in weak labor markets (as defined by unemployment rates, using the median in our sample as a cutpoint), or more rural areas might have less access to employment opportunities outside FWS. We might thus expect to see higher impacts on employment and hours in our main model (Model 1) for these students, and this in turn might be associated with less positive academic impacts and more positive employment impacts. Our results for local labor markets are broadly consistent with this hypothesis, though

the differences are not large (see Appendix Table B.3). However, there is little consistent pattern in the urban versus rural comparison (Appendix Table B.4), with results generally being less positive (or more negative) for students in rural areas, but with a stronger positive effect on of BA completion.

Gender

Gender is a relevant subgroup to examine because Scott-Clayton (2011) found the impact of FWS to be more negative for women than for men at public institutions in West Virginia, perhaps due to differences in job quality and/or differences in the academic opportunity cost. However, in this analysis the consequences of FWS participation appear to be slightly better for women than for men, although the differences are small and generally insignificant (see Appendix Table B.5). Women who receive FWS are less likely than men to have been induced into employment (46 versus 50 percentage points; see Model 1), but women who would not have worked otherwise appear to derive particularly large labor market benefits from FWS participation (e.g., 6.2 percentage point effect on employment in Year 6 for those not currently in school; see Model 3).

7. Discussion

In this paper, we develop a conceptual framework to illustrate the mechanisms underlying the potential impact of subsidizing student employment. We then use this framework to describe an application of propensity scores that we call conditional counterfactual matching. This method may be useful for exploring treatment mechanisms in other non-experimental contexts. For example, it could be used to decompose the effects of preschool subsidies for those who would or would not have attended preschool in the absence of the subsidy. Finally, we apply our framework and method using two waves of nationally representative student survey data, providing the first rigorous national estimates of the effects of the Federal Work-Study program.

On average, our results suggest generally positive effects of Federal Work-Study participation on long-term outcomes, including a 3.2 percentage point impact on BA completion six years after enrollment, and a 2.4 percentage point impact on employment after six years for students who are not still enrolled. But our analysis of distinct conditional counterfactuals highlights why these results do not extend to the impact of student employment more broadly. First, only about half of FWS participants increase their labor supply as a result of the program; the other half actually reduce hours worked. Indeed, all of the academic effects of FWS appear driven by changes in job characteristics for students who would have worked anyway. In contrast, FWS participants who would not have worked otherwise see some negative effects on first-year GPA—though these do not lead to any declines in graduation, and are balanced by later positive impacts on employment. If we use equation (4) to back out the implied effect of *non-*

FWS employment (β_1), we find uniformly negative effects on academic outcomes, and small positive effects on employment outcomes, consistent with prior research.

We do not view these results as conflicting with the mixed-to-negative results on academic outcomes reported in Scott-Clayton (2011) and the large negative first-year GPA results reported in Stinebrickner and Stinebrickner (2003). The more positive results found here may simply reflect the importance of effect heterogeneity. Both prior studies were set in largely rural contexts where students may have had limited other employment opportunities; this may have meant more students were induced into working more hours than they would have otherwise. We also examine a broader set of outcomes. The only outcome available for direct comparison with the most causally rigorous prior work (Stinebrickner & Stinebrickner, 2003) is first-year GPA, and indeed, when we limit our sample to rural students we also find significant negative effects on GPA (-0.10 GPA points in Model 1; see Appendix Table B.4).

The substantial and persistent effects on student debt accumulation are surprising and somewhat puzzling. This finding may indicate that institutional aid officers package loans and work-study together, so that those who take up one program are much more likely to also take up the other. To the extent these two types of aid are packaged together, some of the effects on academic and labor market outcomes could be due to the increased loans rather than to FWS participation, although evidence from Table 6 suggests this is not the case. Still, the relationship between FWS and student loan accumulation is one that should be explored more deeply in future research.

Our subgroup results provide additional support for our conceptual model: groups that we would expect to have a higher likelihood of employment in the absence of the subsidy experience smaller employment impacts, and more positive subsequent impacts. These positive impacts appear driven by reductions in weekly hours and improvements in job amenities (such as on-campus location and relationship to major) for students who would have been working anyway had they not received FWS. An implication is that the effectiveness of Federal Work-Study funds might be increased by modifying the allocation formula—which currently provides disproportionate support to students at elite private institutions—to better target lower-income and lower-scoring students.

References

- Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. B. (2003). Principal stratification approach to broken randomized experiments, *Journal of the American Statistical Association*, 98(462), 299–323.
- Becker, S. O., & Caliendo, M. (2007). Sensitivity analysis for average treatment effects. *Stata Journal*, 7(1), 71–83.
- Brand, J. E., Pfeffer, F. T., & Goldrick-Rab, S. (2012). *Interpreting community college effects in the presence of heterogeneity and complex counterfactuals*. (CCPR Working Paper No. PWP-2012-004). Los Angeles, LA: University of California, California Center for Population Research.
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72.
- Dadgar, M. (2012). *The academic consequences of employment for students enrolled in community college* (CCRC Working Paper No. 46). New York, NY: Columbia University, Teachers College, Community College Research Center.
- Darolia, R. (2014). Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review*, 38, 38–50.
- DeSimone, J. (2008). *The impact of employment during school on college student academic performance* (NBER Working Paper No. 14006). Cambridge, MA: National Bureau of Economic Research.
- DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2002). Simulating the longitudinal effects of changes in financial aid on student departure from college. *Journal of Human Resources*, 37(3), 653–679.
- Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58, 20–29.
- Gronau, R. (1998). *A useful interpretation of r-squared in binary choice models* (Industrial Relations Section Working Paper No. 397). Princeton, NJ: Princeton University.
- Hill, J., Waldfogel, J. & Brooks-Gunn, J. (2002), Differential effects of high-quality child care. *Journal of Policy Analysis and Management*, 21, 601–627. doi: 10.1002/pam.10077
- Hossler, D., Ziskin, M., Kim, S., Cekic, O., & Gross, J. P. (2008). Student aid and its role in encouraging persistence. In S. Baum, M. McPherson, & P. Steele (Eds.), *The*

- effectiveness of student aid policies: What the research tells us* (pp. 101–116). New York, NY: The College Board.
- Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3), 706–710.
- Jo, B., & Stuart, E. A. (2009). On the use of propensity scores in principal causal effect estimation. *Statistics in Medicine*, 28, 2857–2875.
- Kalenkoski, C., & Pabilonia, S. (2010). Parental transfers, student achievement, and the labor supply of college students. *Journal of Population Economics*, 23(2), 469–496.
- Katz, L. (1996). *Wage subsidies for the disadvantaged* (NBER Working Paper No. 5679). Cambridge, MA: National Bureau for Economic Research.
- Light, A. (2001). In-school work experience and the returns to schooling. *Journal of Labor Economics*, 19(1), 65–93.
- Long, B., & Soliz A. (2014). *The causal effect of Federal Work-Study participation on student outcomes in the Ohio public university system*. Cambridge, MA: Harvard University, Graduate School of Education. Unpublished manuscript.
- Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25(3), 17–38.
- National Center for Education Statistics [NCES]. (2003). *Beginning postsecondary students BPS:96/01 restricted CD*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- National Center for Education Statistics [NCES]. (2011). *2004/2009 Beginning postsecondary students longitudinal study restricted use data file*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Oster, E. (2014). *Unobservable selection and coefficient stability: Theory and evidence*. Chicago, IL: University of Chicago Booth School of Business.
- Pascarella, E., & Terenzini, P. (2005). *How college affects students: A third decade of research* (vol. 2). San Francisco, CA: Jossey-Bass.
- Perna, L. W. (2010). Introduction. In L. W. Perna (Ed.), *Understanding the working college student* (pp. xiii–xxvi). Sterling, VA: Stylus Publishing.
- Perna, L. W., Cooper, M. A., & Li, C. (2007). Improving educational opportunities for college students who work. In E. P. St. John (Ed.), *Readings on equal education*, 22, 109–160.

- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York, NY: Springer.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29(1), 185–203.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66(5), 688–701.
- Ruhm, C. J. (1997). Is high school employment consumption or investment? *Journal of Labor Economics*, 15(4), 735–776.
- Scott-Clayton, J. (2011). The causal effect of Federal Work-Study participation: Quasi-experimental evidence from West Virginia. *Educational Evaluation and Policy Analysis*, 33(4), 506–527.
- Scott-Clayton, J. (2012). What explains trends in labor supply among U.S. undergraduates? *National Tax Journal*, 65(1), 181–210.
- Smole, D. P. (2005). *The campus-based financial aid programs: A review and analysis of the allocation of funds to institutions and the distribution of aid to students*. Washington, DC: Congressional Research Service.
- Stinebrickner, T., & Stinebrickner, R. (2003). Working during school and academic performance. *Journal of Labor Economics*, 21(2), 473–491.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.
- The College Board. (2012). *Trends in student aid 2012*. New York, NY: The College Board.
- Tyler, J. H. (2003). Using state child labor laws to identify the effect of school-year work on high school achievement. *Journal of Labor Economics*, 21(2), 381–408.
- U.S. Department of Education. (2000a). *The national study of the operation of the Federal Work-Study program: Summary findings from the student and institutional surveys*. Washington, DC: U.S. Department of Education, Planning and Evaluation Service.
- U.S. Department of Education. (2000b). *National study of the operation of the Federal Work-Study program: Technical appendices for the institutional and student surveys*. Washington, DC: Author.

U.S. Department of Education. (2009). *Federal student aid handbook 2009–2010*. Washington, D.C.: Author.

U.S. Department of Education. (2010a, April 1). Federal campus-based programs data book 2010. Retrieved from U.S. Department of Education website:
<http://www2.ed.gov/finaid/prof/resources/data/databook2010/databook2010.html>

U.S. Department of Education. (2010b). *Federal work-study allocations by institution, 1995–96 to 2009–10*. Washington, DC: U.S. Department of Education, Office of Postsecondary Education.