KU LEUVEN

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES
DPS16.28
NOVEMBER 2016

Does work harm academic performance of students? Evidence using propensity score matching

Tjaša BARTOLJ and Sašo POLANEC

Department of Economics

Faculty of Economics
and Business

Does Work Harm Academic Performance of Students? Evidence Using Propensity Score Matching

Tjaša Bartolj* Sašo Polanec†

November 25, 2016

Abstract

In this article we analyze the effects of student work on academic performance for college students. In order to reduce the endogeneity bias due to selection into treatment, we use propensity score matching technique. This approach allows us to estimate the average treatment effects on the treated separately for different years of study, which is not possible when inside instruments are used to deal with endogeneity of student work. We find predominantly negative treatment effects for all measures of academic performance (GPA, exam attempts, exams passed, and likelihood of passing a year), although many of these are economically and statistically insignificant. We supplement existing studies that do not estimate separate treatment effects for different years of study by showing that work while in college harms study outcomes mostly in the first year of study—by passing smaller number of exams and thereby increasing the likelihood of failing a year. Our results are consistent with evidence on difficulty with adjusting to college studies of first-year students, who face many uncertainties that affect finding the optimal allocation of time between studies, work and leisure.

Keywords: Student work, Academic performance, Treatment effects

JEL Codes: I2, J0

*Institute for Economic Research, Kardeljeva pl. 17, 1000 Ljubljana, Slovenia, email:bartoljt@ier.si.

†University of Leuven, Faculty of Economics and Business, Naamsestraat 69, 3000 Leuven, Belgium, email: saso.polanec@kuleuven.

1 Introduction

The human capital theory predicts that student work can either increase or decrease the stock of accumulated knowledge and consequently improve or worsen individual productivity. Student work may increase human capital through acquisition of new skills, abilities, and knowledge, especially when work is related to studies, which may in turn contribute to academic success and more importantly to the post-college labor market outcomes. At the same time student work might crowd out time for studying and therefore impair academic performance, resulting in a lower accumulation of human capital. In Bartolj and Polanec (2016) we study the potential benefits of student work on post-college outcomes and find evidence of small positive effects of student work experience on wages and likelihood of first regular employment, especially when (i) jobs can be deemed as high skilled and (ii) related to the field of studies. In contrast, the focus of this paper are the effects that student work has on academic performance.

A large body of existing empirical literature concentrated on the impact of student work on high-school average grades. The conclusions of this line of research are, however, mixed. They report (i) *negative effect* (Rothstein 2007; Tyler 2003; Singh 1998; Eckstein and Wolpin 1999; Dustmann and Soest 2007—only for females; Lillydahl 1990—only for sizable levels of student work), (ii) *curvilinear effect* (DeSimone 2006; Oettinger 1999; Post and Pong 2009; Quirk, Keith, and Quirk 2001), and (iii) *no effect* (Lee and Orazem 2010). Papers that analyzed student work during high school also found that it decreases time dedicated to education (Kalenkoski and Pabilonia 2012 and DeSimone 2006), but positively affects graduation rates (Ruhm 1997 and Lee and Orazem 2010).

These results may not be applicable to the post-secondary studies due to important differences between high-school and college studies. The latter are usually less structured and have fewer weekly hours in class, thereby permitting more hours of work even for students enrolled in full-time programs. But at the same time, college students are supposed to take full responsibility for their decisions and are not guided by their teachers and/or parents. Hence they are more likely to worsen their academic performance by engaging in too much paid work. Nevertheless, the empirical evidence on the effects of student work on academic performance for college students is similarly inconclusive. Using GPA as a measure of academic performance, Darolia (2014) and Ehrenberg and Sherman (1987) found no evidence that student work affects GPA, while others found a negative effect (Beerkens, Mägi, and Lill 2011; DeSimone 2008; Callender 2008; Kalenkoski and Pabilonia 2010; Auers, Rostoks, and Smith 2007; Stinebrickner and Stinebrickner 2003). Besides GPA, authors observed also 'graduation-on-time' (Ehrenberg and Sherman 1987; Beerkens et al. 2011), number of credits per term (Darolia 2014), and drop-out probabilities (Ehrenberg and Sherman 1987). All these measures of academic performance were adversely affected by student work.[1]

In this paper we also study the effects of student work on academic performance for a set of business and economics students who were first enrolled in four-year programs at the Faculty of Economics, University of Ljubljana during the period 1997–2004. We use a rich data set on study outcomes of students provided by this faculty, which contains all grades, number of exams passed and information on students' progressing to the next study year. Using unique student identifiers we link these data with administrative data provided by the Slovenian Statistical Office, which contain information on total student pay—our measure of student work—and a large set of variables that are likely affecting student choices regarding labor supply. Using merged data set allows us to make two important contributions to the existing literature. First, we are able to analyze the impact of student work on five distinct measures of academic performance for the same set of students and within one institutional context, which allows us to compare the treatment effects on different outcomes in relative terms. In line with the majority of existing studies listed above, we measure the outcome of intensive margin of study effort—the average GPA. However, in contrast to these studies, we are able to calculate both average GPA for all attempts to pass and for passed exams only. Furthermore, we are also able to estimate the effects on extensive margin of study effort, which is reflected in two measures: number of attempts to pass exams and number of passed exams. Among the studies listed above, only Darolia (2014) considered the effect of student work on the number of credits, which is (in our institutional context) equivalent to the number of passed exams.[2] Finally, we are able to measure the effect of student work on the likelihood of passing/failing the study year, which may be interpreted as an overall measure of academic success, reflecting both intensive and extensive margins of study effort. To the best of our knowledge, this is the first study to use this specific measure of study outcome as other studies report the effects on the probability of dropping out. Our measure is preferred to the probability of dropping out when many students decide to drop out for reasons

[1] Ehrenberg and Sherman (1987) find that only off-campus work negatively affected graduation-on-time and drop-out probabilities in the third and fourth year of study.

[2] In our institutional context students had to pass a minimum number of exams rather than achieve a minimum number of credits per study year.

not related to study success.

Second contribution of this paper is its application of propensity score matching technique to estimation of the average treatment effects on the treated (ATETs) of student work on academic performance. While researchers used this method in other fields of labor economics, this is the first attempt of using it to address this specific question. Our main motivation for using this approach is to deal with endogeneity of treatment variable. Some researchers were able to exploit natural experiments which exhibit exogenous variation in student work (e.g. Stinebrickner and Stinebrickner 2003). Our dataset is, however, constructed from administrative sources, which implies that students included in the sample endogenously chose different amounts of work. Several recent studies used either instrumental variables (IV) estimators (e.g. DeSimone 2008) or generalized method of moments (GMM) estimators (e.g. Darolia 2014) to deal with this issue. These methods yield different treatment effects—local average treatment effects on compliers—which may not be the effects of main interest. More importantly, when using lagged variables as instruments (i.e. inside instruments), it is not possible to estimate the treatment effects separately for all years of study. Our results reveal important differences between different years of study, which is an important advantage of using propensity score matching. Nevertheless, we recognize the fact that propensity score matching can only reduce the part of endogeneity bias that may be captured by observable determinants of student work. While these may be correlated to unobservable characteristics, such as motivation, ability or preferences, we cannot be certain that conditioning on observables has fully eliminated the effect of selection of work effort based on unobservables. In order to minimize the selection bias, we used a large set of personal, economic, family characteristics, and past academic performance, yielding quite high measures of fit for our propensity score estimations.

Our treatment variable is measured with total nominal income that students earned in a given study year, which in principle allows us to use propensity score matching with continuous treatment. Unfortunately our sample did not satisfy the balancing property between all levels of student work. Hence we use the standard propensity score matching for dichotomous variables by discretizing the real annual pay. In particular, we use annual nominal pay of each student and divide it by the average hourly pay for all students and the average number of hours per month to obtain a proxy for the number of months worked. Based on this treatment variable, we calculate ATETs for three levels of student work during a school year: 0–2 months (equivalent to 0–6.7

hours per week), 2–7 months of work (6.7–23.3 hours per week) and more than 7 months.[3] The advantage of this approach are valid estimates of ATETs, but, admittedly, at the expense of estimating treatment effects for heterogeneous groups of students, which cannot be directly compared.

Our estimates of treatment effects suggest that student work indeed harms academic performance, a finding that resonates with many previous studies listed above, but with important differences in terms of size of estimated treatment effects. We find mostly negative ATETs for all measures of academic performance and all study years, although many of these are both small and statistically insignificant. The effects are most harmful in the first year of study, which is consistent with psychological literature providing evidence on difficulties with adjusting to college studies (Baker and Siryk 1984). In particular, for our overall measure of study outcomes—the probability to pass a year—we find significant negative effects for the first and third year of study, while the effects are close to zero for the fourth year of study. For example, first-year students working more than 7 months per study year (2–7 months) compared to those working less than 2 months are 6.8 (4.7) percentage points less likely to pass a year. Adjusted for units of measurement, these effects are twice as large than those reported by Ehrenberg and Sherman (1987) for the probability of dropping out. The effects for the subsequent years of study tend to be much smaller compared to Ehrenberg and Sherman (1987), who find the largest effects for the fourth year of study. The treatment effects for the number of attempts to pass exams and the number of exams actually passed also exhibit similar variation with years of study, although significant negative effects are found for all years of study. The absolute size of these effects is, however, small even for the first year of study. Namely, first-year students working more than 7 months per study year (2–7 months) compared to those working less than 2 months passed 0.5 (0.26) exams less, which is, adjusted for units of measurement, roughly one-tenth of the effect obtained by Darolia (2014). Finally, our estimates of treatment effects of student work on the GPA are (relatively) smallest in size and mostly statistically insignificant. In comparison to existing studies, our results are close to Darolia (2014) and Ehrenberg and Sherman (1987), who found no harmful effect, and significantly lower (in absolute terms) than those obtained by Kalenkoski and Pabilonia (2010), DeSimone (2008) and Stinebrickner and Stinebrickner (2003).

The remainder of this article is organized as follows. Section 2 describes the relevant insti-

[3] These treatment levels are comparable to those used by other studies. For example, Beerkens et al. (2011) use three treatment intervals with boundaries set at 9 hours per week and 25 hours per week.

tutional framework. Section 3 presents data sources and summary statistics. Section 4 specifies the estimation method and discusses findings. Section 5 concludes.

2 Institutional Framework

Our empirical estimation of treatment effects of student work on study outcomes relies on data for Slovenian students who were first enrolled in any four-year undergraduate program offered by the Faculty of Economics, University of Ljubljana (henceforth FELU) during the period 1997–2004. As Slovenian institutional system may not be familiar to the reader, we provide a brief account of its key features relevant to this study.

The FELU is in one of the largest tertiary education organizations in Slovenia, which (in peak years) enrolled as many as 8,000 students in various full- and part-time undergraduate and graduate programs. It is a part of the University of Ljubljana, which is located in country's capital. The university is public and does not charge tuition fees to students with Slovene residence. Students can enrol in the programs offered by the FELU after completing any four-year high school program. The applicants are ranked nationally according to a weighted average grade, calculated from the grade percentage averages achieved in the third and fourth year of high school study and the national exam—matura—a Slovene equivalent of the SAT in the US.[4]

During the period of interest, the FELU offered five business majors (Accounting and Auditing, Business Informatics, Finance, Marketing, and Management and Organization) and three economics majors (Banking and Finance, International Economics, and National Economics). The majority of students chose majors in business studies such as Finance, Management and Organization, and Marketing. The expected time to complete any four-year program at the Faculty of Economics was five years, which includes the additional year for completion of the final thesis. The actual study time typically varied between 4 and 6 years, and could extend beyond 10 years. The grading scheme for undergraduate studies operates on a ten point scale with 1 as the lowest and 10 as the highest grade. The lowest passing grade was 6.[5] Students who failed to pass an exam were allowed to retake it with no limit on the total number of attempts, although the

[4]The high-school grades range between 1 (insufficient) and 5 (excellent); 2 is the lowest passing grade. The matura exam consists of three compulsory (Slovene language, Mathematics, and one foreign language—usually English) and two elective subjects, such as Biology, History, Physics, etc.).

[5]These grades were often a simple mapping from achieved percentages like $Grade = int(Score/10) + 1$, where *int* denotes the integer part of the ratio.

number of exam dates for each course was limited to three per academic year, two exam dates were typically set after the semester of instruction (January and February after winter semester, June and early July after summer semester) and one in September. Due to large number of students, each lecture and class session was generally given more than once, especially in the first two years of study, and students could usually freely choose when they would attend lectures in a given course, which made their time schedule quite flexible.

All full-time students in Slovenia were entitled to generous subsidies, such as free-health care, subsidized meals, and travelling expenses, and could work under different regulations than regular employees. While regular-employment contracts were subject to high social contributions (38.2 percent of gross wage), student-employment contracts—referrals—were not subject to any such tax during the period of interest. Employers were also obliged to pay a bonus for working the night shifts, on Sundays, on holidays, for overtime work, seniority bonus, and bonus for job performance to regular employees, none of which applied to student work.[6] In addition, employers had to cover regular employees' costs for meals during working hours and daily commuting costs (SSC Act 2001). During the period of analysis gross wages were also subject to a progressive payroll tax. All these factors contributed to rather high demand for student work with total value reaching around 1.5% of GDP in the peak years, compared to aggregate wage bill around 45% of GDP.

Student work could be performed by full-time students between 15 and 26 years of age, who were enrolled in any state-approved primary, vocational, high school, or undergraduate programs. Despite preferential tax treatment, student work was not completely tax free. It was subject to a concession fee, value-added tax on concession fee, and personal-income tax. The concession fee was rising over time, starting at 10% of students' gross earnings from 1997 to 2003. From 2003 until 2006 it equalled 12% and afterwards 14% of students' gross earnings. The concession fees were paid by employers on top of students' gross earnings. In addition, employers had to pay value added tax on the concession fee. Therefore the total costs of student work for the employer in 2008 were 116.8 percent of student's gross earnings. Gross earnings of students were also subject to a progressive personal-income tax. While the tax rates were the same for all recipients of different types of personal income, income-tax deduction for students was typically double that applicable to regular employees. As a consequence, net earnings were the same as gross earnings for almost all students, even those who worked full-time entire year

[6] A useful summary of Slovenian labor markets during the period of analysis is given in OECD (2009).

and received average (student) hourly wage. Since we are using data on gross earnings, we do not describe details of personal-income taxation.

3 Data

3.1 Data Sources

As already mentioned above, our analysis uses information on a sample of Slovenian students who were first enrolled in any of the 4-year undergraduate programs offered by the FELU between 1997 and 2004. We used data from several distinct sources, which were merged using person-specific identifiers in a secure room at the Slovenian Statistical Office (henceforth SORS).

The first source of data is the FELU, which provided data from application sheets containing personal information and data from exam records. From this source we extracted information on age, gender, location of permanent residence, chosen major, and study year of students. Based on enrolment history of each student, we also constructed variables that indicate if a student passed a year and repeated a year. Exam records were used to construct the other four measures of study performance—number of attempts to pass, number of exams passed, average grade of all exams and passing exams only.

Next source of data is the Slovenian Tax Authority (henceforth TARS), which collects information on all personal incomes earned. The data on student earnings were reported to TARS by the student employment agencies. While students with sufficiently low earnings were not obliged to fill the personal-income-tax reports, these agencies had a legal obligation to report labor incomes earned by each high-school or college student. Unfortunately data do not contain information on month during which student work was performed, which prevents us from knowing how much work was performed during the semesters and how much during the breaks. TARS is also the source of data on incomes of students' families. Tax filings for personal income tax include both labor and capital incomes, which were used to calculate per capita family incomes. Moreover, labor incomes of families include not only wages and salaries, but also bonuses, perks, wages earned on the basis of short-term labor contracts, and royalties. Capital incomes include interest, dividends, rents, and incomes of sole proprietors.

The third source of data is the National Examination Center, which collects the data on students' high-school performance. We extracted information on the third- and forth-year average grades and the grades from final (external) examination matura. We used these grades to construct a measure of high school GPA.

The last source of data is the SORS. It provided us with the data from the Central Registry of Population, which allowed us to establish parent-child links through unique identifiers of both parents for each student and thus to calculate family incomes and transfers for each student. Having an identity of parents allowed us also to determine their educational attainment, which was collected from the Statistical Registry of Employment. SORS also provided information on all scholarships received by students, ranging from social scholarships targeted to students with low-income families, scholarships given to talented individuals (Zois scholarships), and scholarships granted by prospective employers.

3.2 Construction of Variables and Summary Statistics

Our sample of FELU students enrolled in four-year programs is restricted to (i) those aged between 18 and 20 years when enrolled in the first year of study and (ii) those who finished high school with general matura. The first restriction is due to lack of information on past academic performance for older students, while the second one drops students who passed vocational matura, which is not comparable to general matura and passing it did not suffice for enrolment to university programs.[7] Our final sample contains 3,707, 3,293, 3,201, and 3,103 students in the first, second, third, and fourth year, respectively. The sample size and its structure by gender are presented in Table 1.

Table 1: Sample Size by Gender

	1st Year	2nd Year	3rd Year	4th Year
Number of observations	3,707	3,293	3,201	3,103
Males	1,619	1,402	1,337	1,302
Females	2,088	1,891	1,864	1,801

Let us first describe the outcome variables measuring different aspects of academic performance. As mentioned above we construct five distinct measures, for which we show the means and standard deviations in the top panel of Table 2. The average grade is a variable that re-

[7] Student finishing vocational matura had to pass additional subject of general matura in order to be able to enrol to university courses.

flects the differences in the intensive margin of students' study efforts. It is calculated as an unweighted average of grades achieved in all exam attempts in a given year of study. All negative grades are set to 5, as the differences in negative grades do not exhibit the true variation in knowledge.[8] In the first year the average grade is 6.2, just above the minimum passing grade, but it increases with study years to 7.6 in the last year of study. This is expected, as the less able students drop out of program after failing a year and the remaining students become more apt in finding optimal balance between study, work and leisure. We are also interested in the impact of student work on grades conditional on passing; the averages for all passed exam range between 6.9 and 8.0 in the first and the last year of study, respectively.

The next two measures of academic performance reflect the extensive margins of study effort. These are the number of all attempts to pass exams and the number of exams passed. The numbers of all courses/exams were 10 in the first and second year of study, between 8 and 9 in the third year of study, and between 8 and 10 in the last year of study, depending on a chosen major. In order to pass a year students were obliged to pass all but one course. Students might, however, retake an exam in order to get a passing grade at previously failed exam, or to improve a grade in a course, which they already passed. In the first years of study, the average number of exam attempts exceeds the number of required exams by almost 40 percent, while in the subsequent years this difference is significantly lower. The average numbers of passed exams in the first two years—8.0 and 8.3, respectively—were short of the minimum required number of passed exams to progress to the next year of study, which suggests that some students were not able to pass the study year. The discrepancy between the number of exams passed and minimum required number of exams passed to progress declines with years of study.

Our final measure of academic performance, which captures an overall effect of study efforts, is a binary variable which equals one for students who passed a study year and zero otherwise. Summary statistics reveal that a high percentage of students failed to proceed from the first to the second year of study (33.8 percent), but the vast majority passed the last two years of study.

Next we turn to the measures of treatment—the prevelance of student work. Summary statistics reported in Table 2 show that 81.5% of students were working in the first year of study. This share only increased over the years of study and reached 95.1% by the final study year. Similarly, the average real annual gross pay reported in constant (2004) prices was almost 1,500 EUR in

[8] Although the negative grades range from 1 to 5, grades 1 to 4 were rarely used by some examiners and instead 5 was given to all students who did not pass the required threshold.

the first year of study and increased to 2,614 EUR by the last year of study. As it is customary in this strand of literature to report the effects of student work on study outcomes in terms of time devoted to working, we also express our nominal variables in time equivalents. As we do not observe actual wages for each student and year, we calculate the number of months worked by dividing the nominal earnings by the average gross wage rate in that year, as reported by the largest student employment agency e-Studentski servis, and the average number of working hours in a month.[9] Based on this indicator, we can see that an average student worked around 2.2 months in the first year of study and increased labor supply by roughly 0.5 months each year, reaching 3.8 months by the last year of study.

Table 2: Summary Statistics

	1st Year		2nd Year		3rd Year		4th Year	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Avg. grade	6.230	0.783	6.671	0.813	6.811	0.918	7.618	0.983
Avg. passing grade	6.897	0.588	7.379	0.603	7.343	0.714	7.977	0.777
No. of exam attempts	13.888	3.658	12.920	3.190	10.015	2.828	9.091	2.595
No. of exams passed	7.961	2.482	8.277	2.045	7.085	1.958	7.689	2.326
Passed a year	0.662	0.473	0.750	0.433	0.925	0.263	0.993	0.084
Working during study	0.815	0.389	0.881	0.324	0.915	0.279	0.951	0.217
Gross student work income	1,473	1,663	1,791	1,788	2,161	1,867	2,614	1,968
Months of student work	2.245	2.535	2.730	2.725	3.264	2.831	3.841	2.920
Female	0.563	0.496	0.574	0.495	0.582	0.493	0.580	0.494
Age	18.910	0.418	20.186	0.656	21.405	0.852	22.436	0.881
High school GPA	0.492	0.157	0.506	0.155	0.507	0.155	0.508	0.155
University or higher—mum	0.185	0.388	0.193	0.395	0.197	0.398	0.207	0.405
University or higher—dad	0.219	0.414	0.227	0.419	0.227	0.419	0.230	0.421
Student parent	0.000	0.000	0.001	0.039	0.002	0.053	0.003	0.062
Step parent	0.237	0.425	0.231	0.422	0.235	0.424	0.235	0.424
No. of siblings	1.026	0.719	0.981	0.726	0.928	0.749	0.865	0.747
Non-labor income	6,369	3,774	6,776	4,204	7,228	4,647	7,663	5,161
Conditional-income share	0.127	0.216	0.138	0.228	0.154	0.240	0.160	0.242
Expected net wage	16,244	2,646	16,053	2,590	15,870	2,512	15,851	2,490
Repeated previous year			0.271	0.444	0.200	0.400	0.065	0.246
School year	2,000.3	2.2	2,001.6	2.2	2,002.8	2.2	2,003.9	2.2

Notes: All income-related variables are in constant (2004) Euros. The exchange rate in 2004 was 1 EUR = 1.24 USD. The number of months worked is calculated by dividing the nominal income earned by the average wage rate (for all students) and the average number of hours per month.

The last set of statistics is describing the observable characteristics that are used in our propensity score matching estimations. Personal characteristics are described by gender, age and high school GPA. The latter is our measure of general ability, calculated as a normalized unweighted average of (i) the average grade achieved at matura examination, and (ii) the average grade in the last two years of high school. From Table 2 is evident that the share of

[9]We assume the average number of working hours per month is 175, which is equivalent to 2,100 working hours per year.

females is slightly higher than 50% in all years of study. As females were slightly more likely to progress, we observe slightly increasing shares with study years. Similarly, we also observe that the average high-school GPA increased between the first and second year, suggesting that less able students were more likely to fail the first year of study. The observed students' family characteristics include four binary variables indicating whether each of the two parents obtained university degree or higher, a variable indicating if student has a child and a variable marking students with step parents, and a variable measuring the number of siblings below the age of 27. Table 2 reveals that the share of parents with university degree was around 20% and slightly increasing with study years, share of student parents was below 1%, whereas the share of students with step parents was around 24%. Number of siblings was around 1 in the first year, but decreased in subsequent years of study, mainly due to greater likelihood of siblings of older students passing the age limit for dependent family members.

In order to capture economic background of students, we construct a measure of non-labor income. It is calculated as a sum of (i) net family income per family member, which equals to the sum of parental net income divided by the number of family members,[10] (ii) scholarships, and (iii) pensions received after deceased parents. The average non-labor income was about 6,369 EUR in the first year and increased to 7,663 EUR by the last year of study. As student labor supply decision and academic success might be influenced differently by a part of non-labor income that depends on academic performance, we define a variable—conditional-income share—that measures the proportion of scholarships and pensions in student's non-working income. Table 2 shows that on average around 13% of non-labor income was contingent on study success in the first year of study, and increased to 16% by the last year of study, suggesting that students with greater share of such incomes were more likely to progress to the next study year. We also include a measure of post-graduation expected annual net incomes. This measure captures the incentives that influence allocation of time between work and study. In constructing this measure we assume that students base their expectations of expected income on the most recent wage of persons of the same gender who graduated in their major. Surprisingly, the average expected income declined, which suggests that students coming from majors with lower expected incomes were more likely to pass the year.

Lastly, in our propensity score estimations we also control for grade retention in the previous

[10] We count as family members parents and children under the age of 27, following the personal-income-tax act that defines as a dependent family member a person up to the age of 26 (in addition to other requirements).

year of study, school year, chosen major, and region of permanent address. The grade retention reflects the time available for study and work as students have the opportunity to pass exams for the subsequent study year during the repetition. In Table 2 is presented the share of students that repeated the previous year of study. Not surprisingly, these shares were quite high—27% and 20% in the second and third year of study, respectively. The last three variables attempt to capture the differences in labor market as well as study conditions. For example, chosen major affects the labor demand for students as well as the academic requirements. Similarly, different regions offer diverse job opportunities, but at the same time affect the financial resources and time available for study and work, as those who live in regions located further away from the faculty have to travel daily or rent a room. We present the structure of sample by region in the Appendix (Table 5). From the table it is evident that roughly 45% of all students originate in the Osrednjeslovenska region, where the FELU is located.

4 Empirical Analysis

4.1 Estimation Method

In order to estimate the treatment effects of student work on different measures of academic performance, we match students with different employment choices, but similar predicted probabilities or propensity scores of student employment level. The advantages of using this type of matching approach for our problem are two-fold. First, propensity score matching (PSM) avoids the dimensionality problem of finding matched subjects when there are many control variables. Depending on the year of study, our set of control variables is 39 in the first year, 42 in the second year and 49 in the last-two years, which requires sufficiently large bins for many variables in order to satisfy balancing property. Second advantage is that it imposes minimal structure on estimation. Another feature of matching approach, which we consider as an advantage, is putting emphasis on observations with similar values of regressors. This means that observations at the margin may get little or no weight and thus bear little influence on results. In contrast, OLS tries to minimize squared errors, which may give observations at the margin large weights.

We estimate propensity scores using a logit regression for probability of working k hours during study year (SW_k), using personal characteristics (x) and academic performance in previous study year (A) as explanatory variables:

$$Pr[\mathrm{SW}_{ki} = 1] = \Lambda(\alpha_0 + \alpha_1 x_i + \alpha_2 A_i + u_i), \quad (1)$$

where i indexes individuals and Λ denotes the cumulative logistic distribution function. This conditional probability of receiving treatment (k hours of student work) given x and A is used to match treated observations to controls with similar values of propensity scores. The calculation of the average treatment effect on the treated (ATET) is then based on two assumptions: (i) conditional independence (also called selection on observables, unconfoundedness, or ignorability)[11] and (ii) overlap or matching assumption.[12]

The matching algorithm used in our analysis is radius matching with replacement and imposed common support. This type of matching is a variant of caliper matching that uses all control units within the caliper (or radius) and not only the nearest neighbor as it is done with caliper matching (Dehejia and Wahba, 2002). This feature of radius matching reduces the bias of estimates. Bias is further reduced by matching with replacement, since it allows a treatment unit to be matched to control unit even if control unit was already matched. As suggested by Austin (2011), we use caliper equal to 0.2 of the standard deviation of the logit of the propensity score.[13]

As we expect different levels of student work to have different impact on academic performance, we do not differentiate only between working and non-working students, but instead create three different binary treatment variables, which lead to estimation of three different ATETs. As shown in Table 3, we use students who have less than 2 months of work experience as a control group for two groups: (i) students with 2–7 months of work experience (with ATET_{11} as corresponding treatment effect) and (ii) students with more than 7 months of work experience in a given study year (ATET_{12}). Similarly, students who have 2–7 months of work experience are used as a control group for the groups of students with more than 7 months of experience (ATET_{22}). The rationale for the first boundary set at 2 months is the length of summer holidays,

[11]Conditional on x, outcomes of treatment (y_1) and control group (y_0) are independent of treatment (D). Rosenbaum and Rubin (1983) showed that if the former holds, y_1 and y_0 are also independent of D for given value of propensity score.

[12]For every value of propensity score, there are observations in both control and treatment groups.

[13]We also considered other matching algorithms and other caliper values but obtained qualitatively similar ATETs. We chose this method because it is in line with the recommendation to make a control group as locally comparable as possible to the treated, and baseline differences as little as possible in order to estimate the treatment effects using comparable subjects (Lee, 2005).

while the other two boundaries aim to split the remaining 10 months into two intervals for the winter and summer semesters. In order to be able to compare our results to those in the literature, we can express these intervals in terms of weekly hours of work. Assuming 52 weeks per year and 40 hour work week, 2 months of work is equivalent to 6.7 hours of work per week, whereas 7 months is equivalent to 23.3 hours per week. These values are similar to those used by Beerkens et al. (2011) with boundaries set at 9 and 25 hours per week.

An alternative approach to estimation of treatment effects would be to apply continuous matching as proposed by Hirano and Imbens (2004). As the main advantage, this approach would allow us to estimate local treatment effects for small intervals of student work. However, the response function requires general propensity score to balance pre-treatment variables over all defined intervals, which is in general hard to achieve. The balancing property for continuous matching could not be satisfied for our data, which led us to estimate different propensity scores for different intervals of student work.[14] Namely, we use the same set of personal characteristics in the estimation equation, but allow for different values of regression coefficients.[15] However, a downside of this approach is that ATETs cannot be directly compared. Namely, ATET_{22} is not equal to the difference between ATET_{12} and ATET_{11}, since control groups are, in general, not the same.

Table 3: Construction of Treatment and Control Groups Based on Amount of Student Work

	TREATMENT	
Student work experience	2–7 months	more than 7 months
less than 2 months	ATET_{11}	ATET_{12}
2–7 months		ATET_{22}

The sample size by treatment and control groups is presented in Table 6 in Appendix.

In this manner we estimate the direct effect of student work on academic success, which is measured with average grade, average passing grade, number of exam attempts, number of passed exams, and probability to pass the study year. The indirect effect of student work on academic success in a subsequent period, through academic success in current period, is not accounted for. See also Figure 1 for representation of causal chain and the estimated ATET.

Observed covariates x_i in Equation (1) include indicator variables for different levels of non-

[14]We also considered alternative intervals for calculation of ATETs (e.g. 2-month intervals), but also faced violation of the balancing property in some cases. When balancing property was satisfied, the results we present and those for smaller intervals were nevertheless comparable.

[15]The set of controls increases with years of study as we include past academic performance measures from the second year onwards and indicators for selected majors from the third year of study.

labor income, being female, having step parent, having children, university degrees of mother and father, region of residence, school years and selected majors during college studies in the last two years of study. In addition, set of covariates also comprises share of study-success-contingent income in non-labor income (of students), expected net wage, age, high school GPA, and number of siblings under the age of 27. Furthermore, we add three variables measuring past academic success (A_i) in propensity score equations for the second to fourth year of study. These are number of passed exams and average grade in the previous year, and an indicator variable for repeating previous year of study. We do not control for past student work, as it does not induce imbalance across treatment and control groups once we control for past academic success.

Our use of propensity score matching does not imply that the estimated ATETs may not suffer from any selection bias. When balancing property is achieved the conditional independence assumption holds only for observable characteristics. Moreover, the estimated logit regressions yield McFadden's Pseudo R^2 between 0.044 and 0.14 for different amounts of work and different years of study, which suggests that observed characteristics capture some, but not all, variation in chosen amounts of labor supply. The unobservable heterogeneity between students may still lead to biased ATETs. For our study are particularly worrying the implications of incompletely capturing heterogeneities of students in terms of preferences, motivation and ability with observable characteristics. A priori the direction of omitted variables bias is unclear, as these variables could affect student work and study effort in the same or in the opposite direction. Namely, we would expect an upward (downward) bias in ATETs and less (more) likely negative effect of student work on academic outcomes if omitted variable changes student work and study effort in the same (opposite) direction. It is possible that greater unobserved motivation (or ability, preference for leisure) increases both student work and study effort and the estimated ATETs are upward biased, but also that greater unobserved motivation for studies leads to lower labor supply in which case the ATETs would be downward biased. It is important to keep this caveat in mind when interpreting the results.

4.2 Unconditional Effects of Student Work

Prior to the presentation of the estimated treatment effects, we provide some descriptive evidence on the unconditional relationships between student work experience and the five measures of academic performance. Figures 2 to 5 in Appendix show scatter plots with frequency-weighted markers. These plots reveal rather strong negative relationships between the extent of student work based on total pay and our measures of academic performance, which are suggesting that student work harms study outcomes. Negative relationship is observed for the intensive margin of study effort reflected in the two measures of average grade, the extensive margin of effort reflected in the number of passed exams and number of all attempts to pass exams, and in the overall measure of study success—the probability to pass a year. Based on this descriptive evidence, we are led to conclude that students who work more hours are likely putting less effort in each exam (lower average grade) and prepare for smaller number of exams, which culminates in lower likelihood to pass a year.[16]

Comparison of these unconditional relationships suggests that student work harms academic success more in the early years of study. For example, the overall measure of study success—the probability to pass a year—declines by almost 20 percentage points between 1 month and 7 months of work for students who were enrolled in the first year of study, while in the fourth year of study this difference in probability is less than 5 percentage points. These differences between study years suggest that student work is riskier in the early years of study as students face greater uncertainty about the expected effort required to pass exams, which may be attributed to the well-known effect of adjustment to college (Baker and Siryk, 1984). Moreover, the group of students enrolled in the higher years of studies consists of only those who were able to pass, which makes them less heterogeneous in terms of ability.

Before turning to the discussion of estimated treatment effects, we make two additional remarks regarding unconditional relationships between student work and academic performance. First is about the shape of these relationships and potential biases if one were to ignore the endogenous selection. On one hand the shape appears inverse U-shaped for several measures of academic performance in several years of study with peak at around 3 months.[17] While this may suggest that the trade-off between student work and academic performance only kicks in for those who work more than 3 months, it may also be a consequence of selection into treatment. Namely, if working students are also more motivated (or more able) than those not working, they can achieve better academic results. Similarly, rather large negative effects of student work may

[16]Note that the total number of attempts to exams does not have an a priori negative relationship with student work. Students who work more are less likely to pass and consequently may exhibit more attempts to pass exams.

[17]Note that this pattern is not unique to our data. Darolia (2014) finds a similar inverse U-shaped pattern using US National Longitudinal Youth Survey 1997 with peak around 5 hours of work per week for full-time students, which is equivalent to 1.5 months of work per year.

also be partly driven by selection. In fact, pre-matching comparison of high-school GPAs and parental education between students with different levels of work shows that students who decide to work more are those with worse grades and lower incidence of university degree by parents. These differences are significantly reduced when comparing the pre- and post-matching characteristics of controls and treated on our sample, which suggests that propensity score matching significantly reduces selection bias.

Second remark is about a peculiar feature of our scatter plots, which exhibit very dispersed and even improved academic performance for students who worked around 10–11 months per year. This is observed for four measures of study (exception being the number of all attempts to pass exam) primarily in the first year of study. We believe this is partly due to smaller samples of students used to calculate the averages, but also due to possibility that some of the students participated in tax evading activities. Since taxation of student work was significantly lower than that of regular employment contracts, students could earn a fee by allowing firms to extract cash from their businesses through alleged student work.[18] Students who engaged in such activities often aimed to transfer an amount that was just below the sum of general deduction and student-specific deduction as the marginal tax rate was 0% and the total cost for employer was between 110% and 114% of the gross value of student work. Given the observed patterns it is possible that our estimates suffer from attenuation bias for larger amounts of student work. Hence, the true harmful effects of actual student work on academic performance could be underestimated for larger values of our measures of student work.

4.3 ATETs of Student Work on Academic Performance

In this section we present the treatment effects of student work on academic performance using the propensity score methodology. The propensity score balances all variables in all estimations at p-value of 0.05. This confirms validity of conditional independence assumption for observable characteristics and implies that the estimated ATETs are based on students with very similar characteristics who chose different amounts of student work. The ATETs for average grades, attempts to pass, number of passed exams and likelihood to pass are shown in Table 4. Overall, we find that the estimated ATETs are in line with the descriptive statistics, which suggests that student work indeed harms study outcomes. The treatment effects are, however, significantly

[18] For example, a fee could amount to 5% of transferred cash.

lower in size when compared to the effects implied by unconditional relationships in Figures 2 to 5, which confirms our concern about the selection into treatment.

Let us start with description of the estimated effects on the average grades. The majority of estimated ATETs are negative, a finding that is consistent with observed negative unconditional relationships. In terms of sign and size, the effects for the two GPA measures, calculated from all grades and passing grades only, are fairly similar. Moreover, at least one statistically significant effect is found between second and fourth year of study, while no ATET is significant in the first and third year of study. The most harmful negative effects (in absolute terms) appear to be in the fourth year, although even these are rather small and economically insignificant. For example, the largest effect is -0.13 (1.7% of the average grade in fourth year) when student work is increased from 2-7 months to more than 7 months. In order to compare our results with previous studies, note that a decrease of Slovenian GPA by 0.13 is equivalent to 1.3 percent grade and 0.03 points on US 4.0 scale. Converting this to an hour equivalent using for simplicity uniform distribution of student work within treatment intervals, additional hour of work per week is at most reducing GPA by 0.0025 points. Our estimated effects are closer to those obtained by Darolia (2014) and Ehrenberg and Sherman (1987), who found no evidence of student work affecting GPA, and significantly smaller than those estimated by Kalenkoski and Pabilonia (2010), DeSimone (2008) and Stinebrickner and Stinebrickner 2003), who found that additional hour of work per week reduces GPA by 0.017, 0.11 and 0.16 points, respectively.

Next we discuss the ATETs of student work on the numbers of exam attempts and exams passed (Table 4). Note first that the sign of the effect of student work on the number of attempts is a priori ambiguous. Higher number of exam attempts is due to either students putting more effort into studies and trying to pass as many exams as possible or putting too little effort and failing to pass exams and thus having to retake them. The estimated ATETs are nevertheless mostly negative not only for the total number of exams passed, but also for the number of exam attempts. Yet the negative effects for the number of attempts are significant mainly in the first year, whereas for the number of exams passed significant negative effects are observed in all four years. In contrast to ATETs for average grades, for which the most harmful effects are found in the last two years of study, ATETs for the number of passed exams imply that student work is most harmful in the first and fourth year of study. For example, first-year students working 2–7 months and students working more than 7 months attempted to pass 0.418 (3% of average number of exam attempts) and 0.644 less exams (4.6%) than their peers working less than 2 months. Similarly, first-year students working 2–7 months and students working more than 7 months pass 0.263 (3.3% of average number of exams) and 0.492 (6.2%) less exams than their peers working less than 2 months. The same difference in the amount of work has significantly smaller negative impact in the second and third year of study (less than or around 3%), whereas in the fourth year even bigger effect is observed when comparing more than 7 months of work to 2–7 months of work. In order to compare our result with previous studies, we recalculate the estimated effects per hour of work. Our largest ATET of 0.02 less passed exams per additional hour is significantly lower than the estimated effect by Darolia (2014), who found 0.2 less completed courses per hour worked.

The effects of student work on the two margins of study effort are ultimately reflected in the probability to pass the study year, an encompassing measure of study effort. The ATETs reported in Table 4 are mostly negative, which suggests that student work indeed reduces study effort. The economically and statistically most significant effects are mainly observed for the first year of study, which is consistent with above result that students who work more attempt to pass and actually pass less exams. Namely, students pass the first year of study with 4.7 and 6.8 percentage points lower probability if they work 2–7 months and more than 7 months instead of less than 2 months, respectively. In comparison to the average probability of passing a year (67 percent in the first year), the two ATETs imply 5 and 10% lower probability of passing in the first year. In subsequent study years student work can reduce the probability of pass by 3 percantage points in the second year and 5 percentage points in the third year, although these negative effects only kick in when students engaged in more than 7 months of work compared to less those with less than 2 months of work.

Our results are not directly comparable to other studies in the literature as these estimated effects of student work on different variables of interest. The closest outcome variable to ours is the probability of dropping out, which is one of the outcome variables of interest in the analysis by Ehrenberg and Sherman (1987). This variable reflects both voluntary decisions to drop out by students not interested in studies or pursuing other goals and involuntary drop outs due to inability to pass a year. Thus the effects of student work on this indicator should be higher than those found in our study as students not interested in studies might explore labor market prospects even before dropping out. Ehrenberg and Sherman (1987) found that 20 hours of work per week increased the probability of dropping out by 3-3.4 percentage points in the first three years of study and 4.5 percentage points in the last year of study. Expressing these marginal effects and ATETs per hour of work per week, we find that the probability of not passing a year in the first year increases by 0.0024 per hour of work per week in our data (for both estimates), whereas the marginal effect estimated by Ehrenberg and Sherman (1987) is 0.0016 per hour of work per week. However, our estimates for the subsequent years are significantly smaller than those by Ehrenberg and Sherman (1987).

In a nutshell, the results based on ATETs show that work during studies indeed negatively affects study success. The largest negative effects on the probability of passing a year and the number of passed exams are observed in the first year of study when students change environment and face many uncertainties regarding required work effort to pass a year. This finding is consistent with a large psychological literature, which has extensive evidence on the determinants of first-year college persistence (see, for example, Pascarella and Terenzini 1983 and Kahn and Nauta 2001) and difficulties with adjusting to college (see Baker and Siryk 1984). Harmful effects of work are also observed in subsequent years of study, although these effects are smaller and more nuanced. Sizeable and significant negative effects on the probability to pass a year are observed also in the third year of study, while negative effects in the second year are not significant and in the final year they are negligible. The negative effects of student work on the number of passed exams is observed in all years of study, whereas some negative effects on average grades are observed in the second and fourth year of study.

As already mentioned above we believe our estimates may suffer from measurement error in student work related to tax evading activities of students. This bias likely attenuates the estimated treatment effects particularly for larger values of student work. Nevertheless, we believe that this phenomenon was relatively modest. One strong reason against the importance of tax evasion stems from the observed shape of the distribution of earnings. As it was rational for all students to engage in tax evasion, all that had a possibility to do so, should have exploited this possibility up to the amount of student tax deduction. This would lead to a distribution of earnings with a peak at the level of tax deduction, which is not observed in our data. Moreover, we should not observe any relationship between student work and academic performance if student work was mainly used as a mode for tax evasion. Thus, while we are aware that our measure of student work is not ideal, we believe, that student work had rather modest negative effects on academic performance for the FELU students.

There is, however, also concern that the estimated ATETs are downward biased due to measurement error that arises from using real annual pay as a measure of student work rather than actual hours worked. As we estimate working hours using a fixed average wage in each given year, variation in hours worked does not reflect only actual hours worked, but also differences in hourly wages. If students who were earning more also had better paying jobs, the estimated ATETs would suffer from this additional source of attenuation bias. While this may indeed be a problem, we put forward several reasons why we believe this is not the case. First, as opposed to regular employment, student work does not exhibit large differences in pay based on educational attainment and work experience. Students predominantly perform simple tasks (e.g. basic administrative work, cashiers, furniture movements, etc.) that do not require high human capital and pay comparable rates. Second, our matching approach relies on many personal characteristics (e.g. gender, GPA, college major) that might be correlated with performance of certain highly-paying jobs (e.g. hostess). Third, in order to investigate whether earned student income (or supplied number of working hours) is correlated with wages, we used the data on hours worked and wages from the largest Slovenian employment agency e-Študentski servis and the largest platform for matching students of the FELU with potential employers. Unfortunately, these data are not available for the entire period and all students used in our analysis as it covers the period between 2005 and 2007. The overlap with our main data set is 1,282 observations for students who were working and who were enrolled between second and fourth year of study. Using this restricted set of observations we estimated a regression model for the log of hourly wage with log of annual earnings and the same set of controls we used in the propensity score estimations. The estimate of regression coefficient or partial elasticity for working income is 0.052 (s.e. = 0.17), which confirms the implied positive relationship. However, this estimate yields small differences in absolute wages when we predict wages for different levels of income while keeping other variables fixed. Namely, even if we increase annual work income from 656 EUR (1 month) to 7,872 EUR (12 months), the predicted hourly wage would increase only by 12.7%. This difference is not sufficiently large to generate quantitatively large attenuation bias in ATETs.

Table 4: Estimates of ATETs of Student Work on Academic Performance Measures

	Avg. grade		Avg. passing grade		No. of exam attempts	
	2–7 months	over 7 months	2–7 months	over 7 months	2–7 months	over 7 months
			1st Year			
less than 2 months	−0.001	0.007	0.004	0.021	−0.418**	−0.644*
	(0.028)	(0.056)	(0.021)	(0.042)	(0.134)	(0.261)
2–7 months		0.018		0.028		−0.239
		(0.056)		(0.041)		(0.266)
			2nd Year			
less than 2 months	−0.072*	−0.067	−0.051*	−0.058	−0.022	0.008
	(0.030)	(0.053)	(0.024)	(0.040)	(0.123)	(0.212)
2–7 months		0.040		0.028		−0.046
		(0.053)		(0.041)		(0.209)
			3rd Year			
less than 2 months	−0.053	−0.056	−0.033	−0.032	−0.163	−0.223
	(0.035)	(0.051)	(0.028)	(0.041)	(0.114)	(0.184)
2–7 months		−0.040		−0.022		0.133
		(0.052)		(0.041)		(0.171)
			4th Year			
less than 2 months	−0.051	−0.106	−0.055	−0.113*	0.010	−0.299
	(0.042)	(0.063)	(0.032)	(0.047)	(0.116)	(0.177)
2–7 months		−0.130**		−0.094*		−0.378**
		(0.050)		(0.041)		(0.144)

	No. of exams passed		Passed a year	
	2–7 months	over 7 months	2–7 months	over 7 months
		1st Year		
less than 2 months	−0.263**	−0.492**	−0.047**	−0.068*
	(0.094)	(0.184)	(0.018)	(0.033)
2–7 months		−0.198		−0.013
		(0.198)		(0.035)
		2nd Year		
less than 2 months	−0.202**	−0.167	−0.022	−0.032
	(0.077)	(0.139)	(0.016)	(0.030)
2–7 months		0.069		−0.000
		(0.137)		(0.030)
		3rd Year		
less than 2 months	−0.229**	−0.216	−0.012	−0.050*
	(0.080)	(0.126)	(0.010)	(0.020)
2–7 months		0.036		−0.038
		(0.121)		(0.020)
		4th Year		
less than 2 months	0.009	−0.287	−0.004	−0.005
	(0.106)	(0.163)	(0.003)	(0.005)
2–7 months		−0.469**		0.001
		(0.126)		(0.005)

Notes: Standard errors are reported in parentheses. * and ** denote statistical significance at $p < 0.05$ and $p < 0.01$, respectively. The number of months worked is calculated by dividing the nominal income earned by the average wage rate (for all students) and the average number of hours per month.

5 Conclusion

In this paper we analyze the impact of student work on academic performance using five distinct measures of academic performance. Unlike Stinebrickner and Stinebrickner (2003), who exploit quazi-natural experiment with random assignment of students into different amounts of work, our data are administrative and unconditional relationships based on these may suffer from potential problem of endogenous selection into treatment. We deal with this issue by estimating the average treatment effects on the treated (ATETs) using the propensity score matching with a large set of observable personal characteristics and measures of past academic success. This approach allows us to estimate the treatment effects separately for all four years of undergraduate studies and for different levels of treatment.

Our results support previous studies, which have found harmful effects of student work, as we obtain predominantly negative ATETs. These effects nevertheless vary between different measures of academic performance and years of study, both in terms of economic and statistical significance. We find that student work has the worst effects in the first year of study, which is in line with evidence on difficulties with adjusting to college studies (Baker and Siryk, 1984). For our overall measure of study effort—the probability of passing a year—the largest effect is as large as 6.8 percentage points when students work more than 7 months (per study year) compared to those who work less than 2 months. When adjusted for units of measurement, our ATETs are even higher than those by Ehrenberg and Sherman (1987), who use drop out as an outcome measure. For the subsequent years of study, however, we find quantitatively smaller results than those reported by Ehrenberg and Sherman (1987). As students are required to pass a minimum number of exams, the estimated ATETs for the numbers of exams passed (and number of attempts to pass exams) are similar, although smaller than those reported by Darolia (2014). Namely, the most harmful effects are observed for the first year of study as more than 7 months of work compared to less than 2 months reduces the number of passed exams by 0.5, which represents 5 percent of all required exams in the first year. For subsequent study years the ATETs are typically lower, with exception of fourth year—when students were obliged to pass less exams to progress a year. Finally, we find that student work had the least harmful effects on average grades. Even taking the largest values found for the fourth year of study, our effects are small in size when compared to those reported by Kalenkoski and Pabilonia (2010), DeSimone (2008) and Stinebrickner and Stinebrickner (2003).

In summary, we find that student work has small negative effect on different measures of academic performance for work that exceeds 2 months per academic year. Although our results are based on data for a single institution in a small nation, the results can be generalized to similar institutional contexts, which allow students to adjust study and work schedules quite easily due to non-obligatory attendance rules and repetition of classes. Students in faculties with less flexible timetable may find it harder to balance work and study, and thus the negative effects of work might be larger.

References

Auers, D., Rostoks, T., & Smith, K. (2007). Flipping burgers or flipping pages? student employment and academic attainment in post-soviet latvia. *Communist and Post-Communist Studies*, 40(4), 477–491.

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, 10(2), 150–161.

Baker, R. W., & Siryk, B. (1984). Measuring adjustment to college. *Journal of Counseling Psychology*, 31(), 1251–1271.

Bartolj, T., & Polanec, S. (2016). The Game of Skills: A Comparison of Impacts of Student Work and Academic Performance on Post-College Labour Market Outcomes. Mimeo

Beerkens, M., Mägi, E., & Lill, L. (2011). University studies as a side job: causes and consequences of massive student employment in estonia. *Higher Education*, 61(6), 679–692.

Callender, C. (2008). The impact of term-time employment on higher education students' academic attainment and achievement. *Journal of Education Policy*, 23(4), 359–377.

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.

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

DeSimone, J. (2006). Academic performance and part-time employment among high school seniors. *Topics in Economic Analysis & Policy*, 6(1)

DeSimone, J. S. (2008). The impact of employment during school on college student academic performance. Tech. rep., National Bureau of Economic Research

Dustmann, C., & Soest, A. (2007). Part-time work, school success and school leaving. *Empirical Economics*, 32(2-3), 2–3.

Eckstein, Z., & Wolpin, K. I. (1999). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica*, 67(6), 1295–1339.

Ehrenberg, R. G., & Sherman, D. R. (1987). Employment while in college, academic achievement, and postcollege outcomes: A summary of results. *Journal of Human Resources*, 22(1), 1–23.

Hirano, K., & Imbens, G. W. (2004). The propensity score with continuous treatments. In: Gelman, A., & Meng, X.-L. (Ed.) Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, John Wiley & Sons, Ltd, Chichester, UK, pp. 73–84.

Kahn, J. H., & Nauta, M. M. (2001). Socialcognitive predictors of first-year college persistence: The importance of proximal assessment. *Research in Higher Education*, 42(6), 633–652.

Kalenkoski, C. M., & Pabilonia, S. W. (2010). Parental transfers, student achievement, and the labor supply of college students. *Journal of Population Economics*, 23(2), 469–496.

Kalenkoski, C. M., & Pabilonia, S. W. (2012). Time to work or time to play: The effect of student employment on homework, sleep, and screen time. *Labour Economics*, 19(2), 211–221.

Lee, C., & Orazem, P. F. (2010). High school employment, school performance, and college entry. *Economics of Education Review*, 29(1), 29–39.

Lee, M.-J. (2005). *Micro-econometrics for policy, program, and treatment effects*. Oxford University Press Oxford

Lillydahl, J. H. (1990). Academic achievement and part-time employment of high school students. *Journal of Economic Education*, 21(3), 307–316.

OECD (2009). Reviews of labour market and social policies for slovenia. Tech. rep., OECD Publisihing, Paris

Oettinger, G. S. (1999). Does high school employment affect high school academic performance? *Industrial and Labor Relations Review*, 53, 136–151.

Pascarella, E. T., & Terenzini, P. (1983). Predicting voluntary freshman year persistence/withdrawal behavior in a residential university: A path analytic validation of tinto's model. *Journal of Educational Psychology*, 75(2), 215–226.

Post, D., & Pong, S.-L. (2009). Student labour and academic proficiency in international perspective. *International Labour Review*, 148(1-2), 93–122.

Quirk, K. J., Keith, T. Z., & Quirk, J. T. (2001). Employment during high school and student achievement: Longitudinal analysis of national data. *The Journal of Educational Research*, 95(1), 4–10.

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.

Rothstein, D. S. (2007). High school employment and youths' academic achievement. *Journal of Human Resources*, 42(1), 194–213.

Ruhm, C. J. (1997). Is high school employment consumption or investment? *Journal of Labor Economics*, 15(4), 735–776.

Singh, K. (1998). Part-time employment in high school and its effect on academic achievement. *The Journal of Educational Research*, 91(3), 131–139.

SSC Act (2001). Social Security Contributions Act, *Official Gazette of the RS* [Zakon o prispevkih za socialno varnost, *Uradni list RS*] no. 5/1996, 34/1996, 97/2001

Stinebrickner, R., & Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of Labor Economics*, 21(2), 473–491.

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.

Appendix

Table 5: Structure of Sample by Region

Region	1st Year	2nd Year	3rd Year	4th Year
Pomurska	1.67	1.49	1.56	1.55
Podravska	1.32	1.31	1.28	1.32
Koroška	1.81	1.67	1.69	1.74
Savinjska	7.72	7.44	7.37	7.35
Zasavska	1.92	2.06	2.12	2.19
Spodnjeposavska	2.32	2.25	2.31	2.06
Jugovzhodna	8.96	9.20	9.53	9.38
Osrednjeslovenska	45.51	45.19	44.80	45.25
Gorenjska	13.33	13.54	13.59	13.44
Notranjsko - kraška	2.48	2.46	2.56	2.51
Goriška	7.09	7.29	7.31	7.32
Obalno - kraška	5.88	6.10	5.87	5.90

Table presents shares in percent of respective column total.

Table 6: Sample Size by Treatment and Control Groups

	1st Year	2nd Year	3rd Year	4th Year
Number of observations	3,707	3,293	3,201	3,103
Student work experience				
less than 2 months	2,206	1,717	1,321	979
2–7 months	1,249	1,267	1,501	1,645
more than 7 months	252	309	379	479

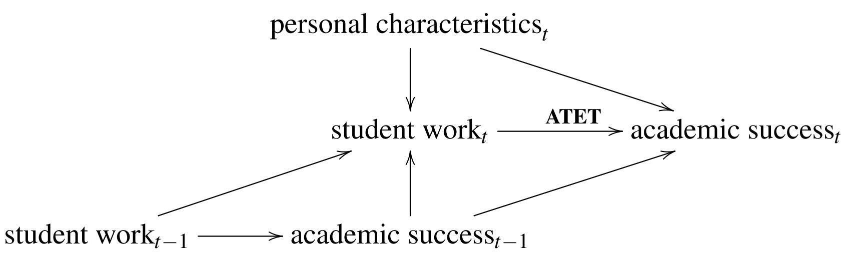

Figure 1: Representation of Causal Chain

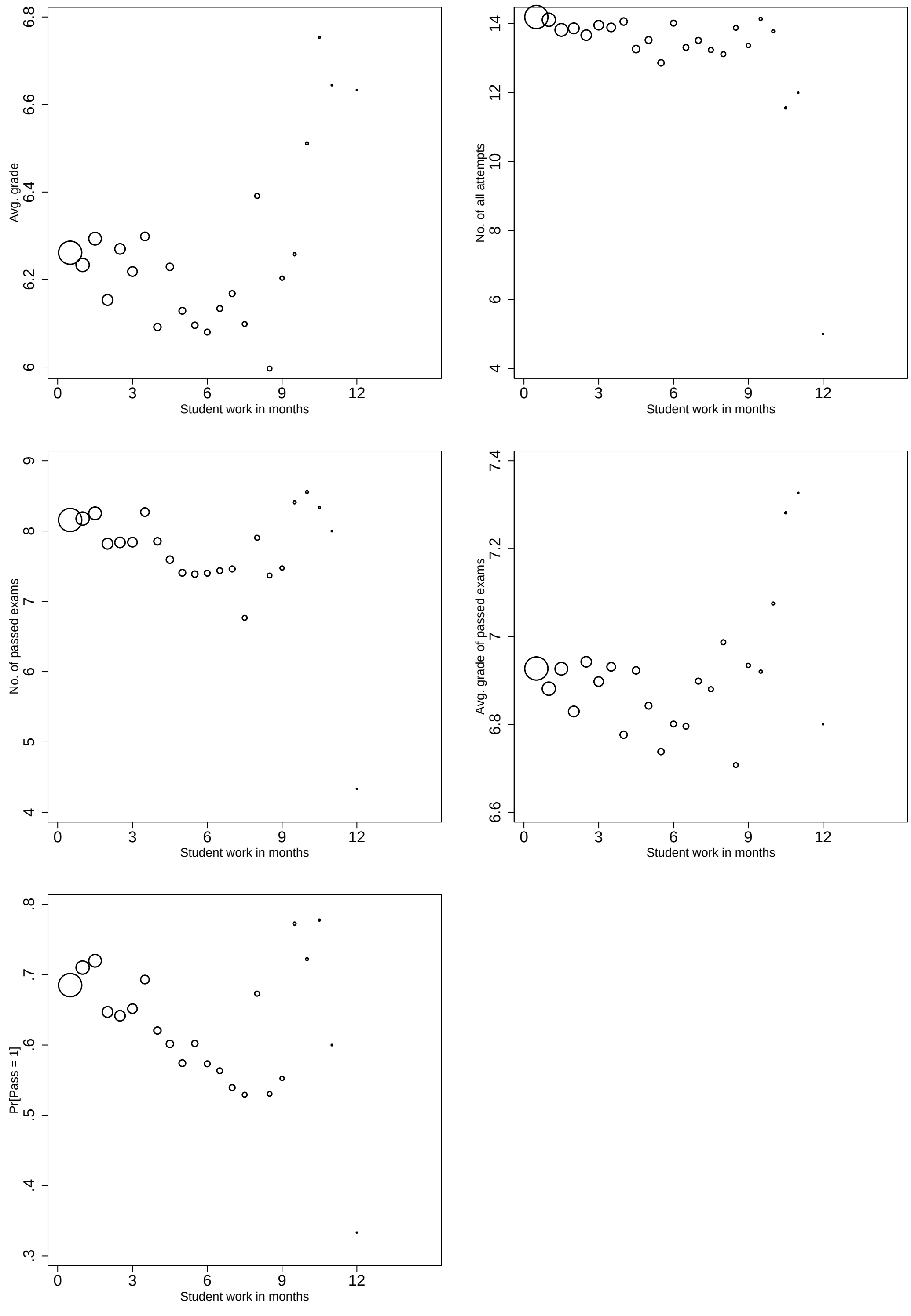

Figure 2: *Academic Performance by Student Work in the First Year of Study.* The size of markers is proportional to the frequency of students with a specific value of student work. Markers with frequency lower than 10 are omitted. The number of months worked is calculated by dividing the nominal income earned by the average wage rate (for all students) and the average number of hours per month.

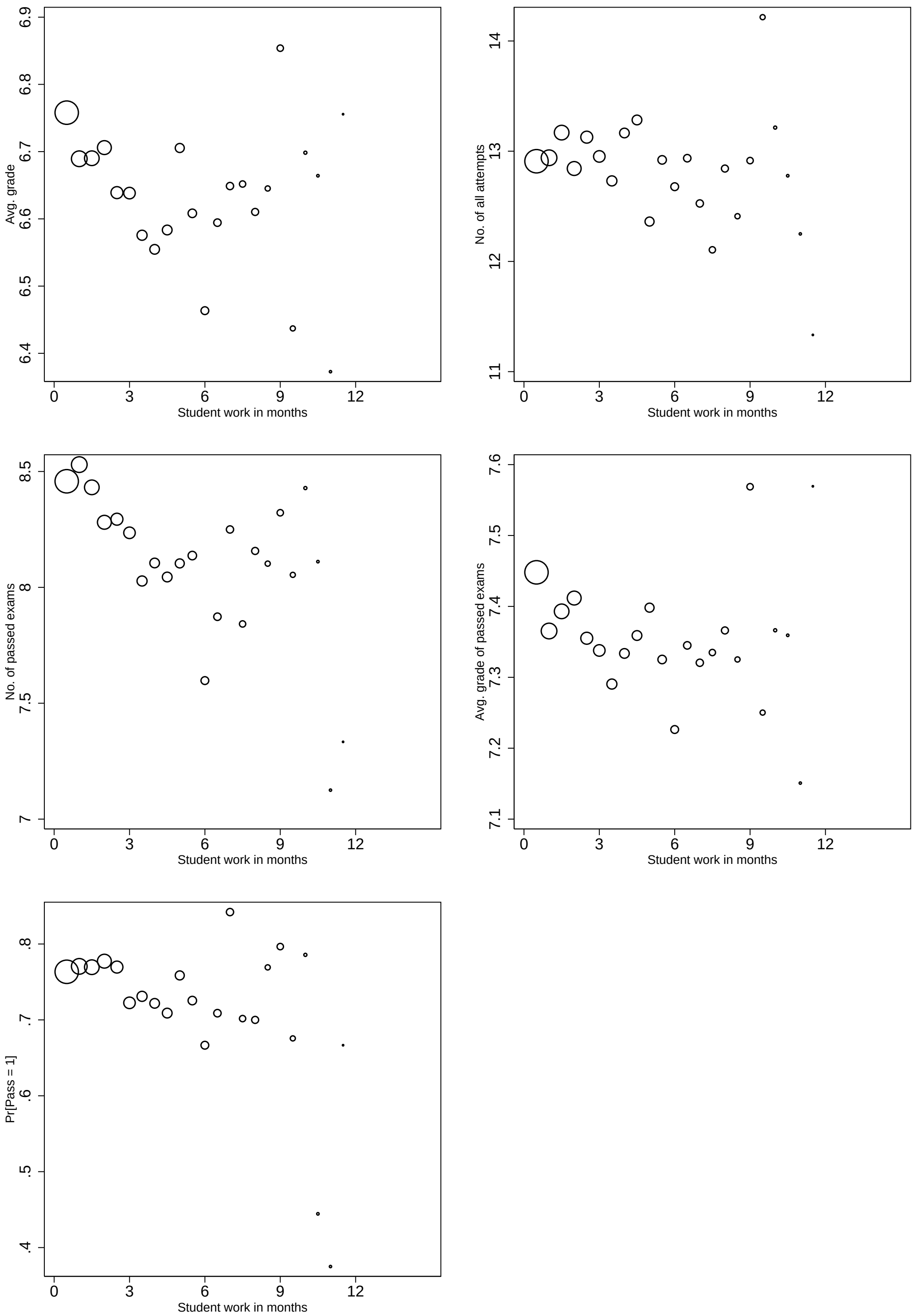

Figure 3: *Academic Performance by Student Work in the Second Year of Study.* The size of markers is proportional to the frequency of students with a specific value of student work. Markers with frequency lower than 10 are omitted. The number of months worked is calculated by dividing the nominal income earned by the average wage rate (for all students) and the average number of hours per month.

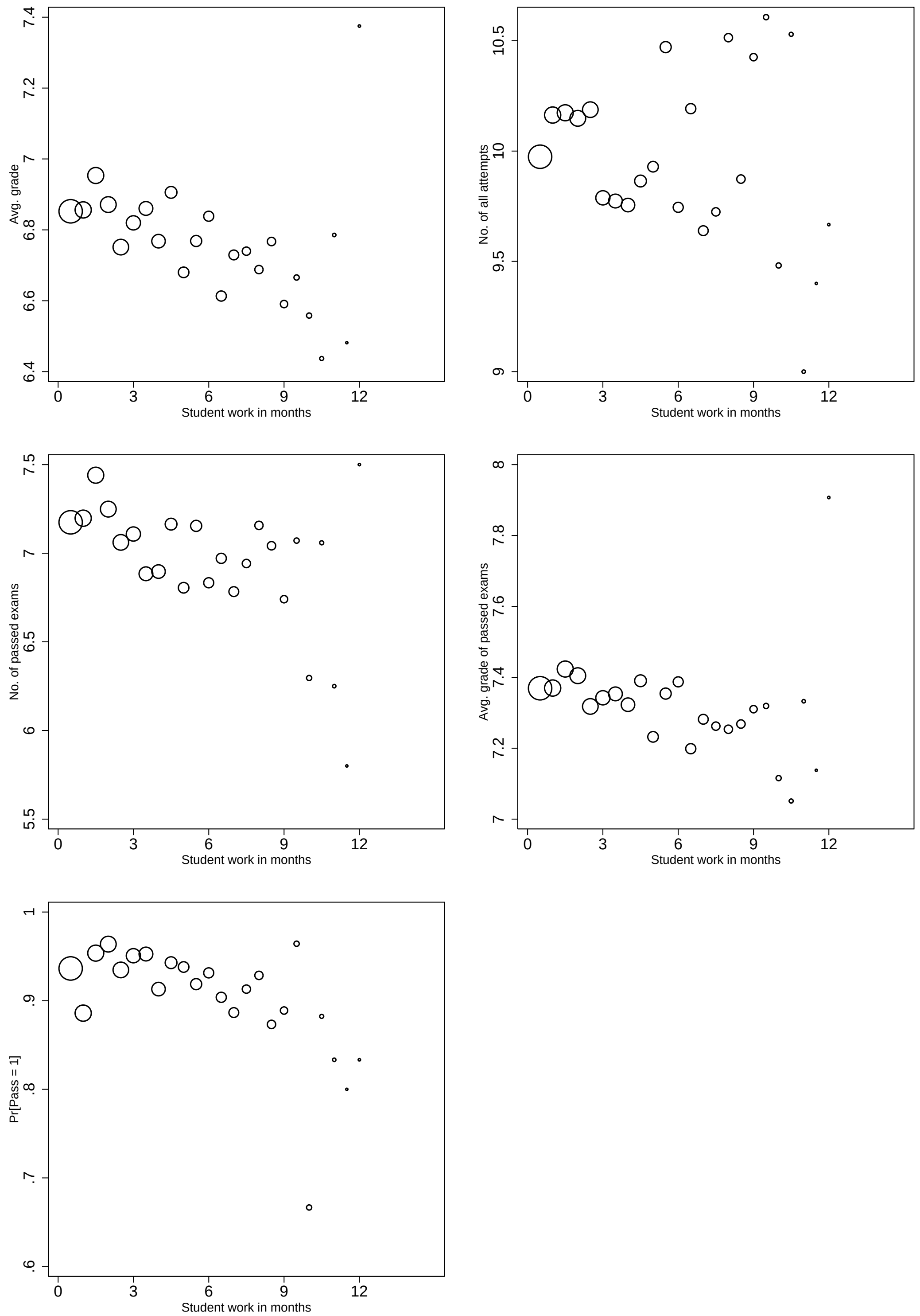

Figure 4: *Academic Performance by Student Work in the Third Year of Study.* The size of markers is proportional to the frequency of students with a specific value of student work. Markers with frequency lower than 10 are omitted. The number of months worked is calculated by dividing the nominal income earned by the average wage rate (for all students) and the average number of hours per month.

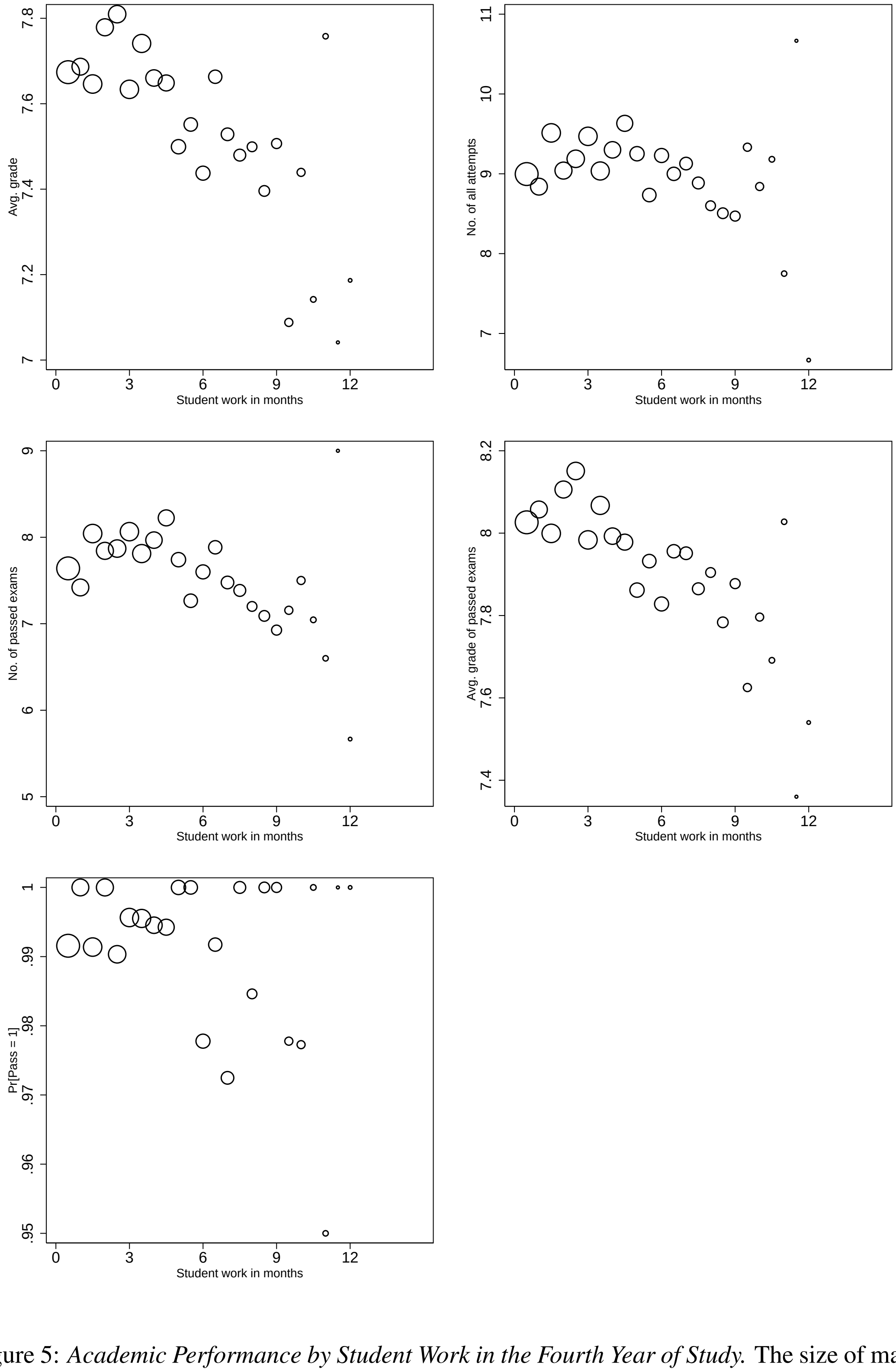

Figure 5: *Academic Performance by Student Work in the Fourth Year of Study.* The size of markers is proportional to the frequency of students with a specific value of student work. Markers with frequency lower than 10 are omitted. The number of months worked is calculated by dividing the nominal income earned by the average wage rate (for all students) and the average number of hours per month.