CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

The Impact of Student Employment on Educational Outcomes: A Meta-Analysis

Master's thesis

Author: Bc. Kateřina Kroupová
Study program: Economics and Finance
Supervisor: doc. PhDr. Tomáš Havránek, Ph.D.

Year of defense: 2021

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, January 4, 2021

Kateřina Kroupová

Abstract

Despite the extensive body of empirical research, the discussion on whether student employment impedes or improves educational outcomes has not been resolved. Using meta-analytic methods, we conduct a quantitative review of 861 effect estimates collected from 69 studies describing the relationship between student work experience and academic performance. After outlining the theoretical mechanisms and methodological challenges of estimating the effect, we test whether publication bias permeates the literature concerning educational implications of student employment. We find that researchers report negative estimates more often than they should. However, this negative publication bias is not present in a subset of studies controlling for the endogeneity of student decision to take up employment. Furthermore, after correcting for the negative publication bias, we find that the student employment-education relationship is close to zero. Additionally, we examine heterogeneity of the estimates using Bayesian Model Averaging. Our analysis suggests that employment intensity and controlling for student permanent characteristics are the most important factors in explaining the heterogeneity. In particular, working long hours results in systematically more negative effect estimates than not working at all or working only a few hours per week. In contrast, studies accounting for student pre-existing characteristics such as ability yield consistently positive estimates.

JEL Classification	I21, I20, I23, C11
Keywords	student employment and educational outcomes, meta-analysis, publication bias, Bayesian Model Averaging
Title	The Impact of Student Employment on Educational Outcomes: A Meta-Analysis

Abstrakt

Dosavadní vědecké poznatky nenabízejí jednoznačnou odpověď na otázku, zdali zaměstnání studentů má pozitivní či negativní dopad na jejich studijní výsledky. Cílem této diplomové práce je za použití metody meta-analýzy kvantitativně přezkoumat 861 odhadů, které pocházejí z 69 empirických studií popisujících vztah mezi prací studentů a jejich akdemickými výsledky. Po představení teorie a metodologických problemů při odhadování tohoto efektu, zkoumáme přítomnost publikační selektivity. Z naší analýzy vyplývá, že negativní odhady jsou reportovány častěji než pozitivní odhady. Pokud se však zaměříme na odhady studií, které uvažují možnou endogeneitu rozhodnutí studentů zapojit se do pracovního procesu, toto nadměrné publikování negativních odhadů (negativní publikační selektivita) zmizí. Kromě negativní publikační selektivity výsledky naší analýzy ukazují, že když očistíme efekt od této odchylky, výsledný půměrný efekt je téměř nulový. Za pomoci Bayesovského průměrování modelů dále zkoumáme, které aspekty primárních studií, ze kterých jsme odhady extrahovali, způsobují rozdílnost těchto odhadů. Na základě výsledků naší analýzy víme, že hlavními faktory způsobující tuto rozdílnost je intenzita zaměstnání, tedy kolik hodin týdně student stráví v práci a vlastnosti studentů. Pokud student věnuje pracovním povinnostem hodně času, jeho studijní výsledky jsou systematicky horší, než když nepracuje anebo pracuje méně. Naopak studie, které ve svých analýzách berou v potaz charakteristiky studentů jako je například jejich studijní způsobilost, vykazují systematicky pozitivní efekt zaměstnání studentů na jejich akademické úspěchy.

Klasifikace JEL	I21, I20, I23, C11
Klíčová slova	zaměstnání studenta a akademické výsledky, meta-analýza, publikační selektivita, Bayesovské průměrování modelů
Název práce	Vliv zaměstnání studenta na akademické výsledky: Meta-analýza

Acknowledgments

The author would like to express her great gratitude to doc. PhDr. Tomáš Havránek, Ph.D. for his guidance, expertise and useful comments during conducting and writing this thesis. The author would also like to thank her fellow students for providing new perspectives on the research topic. Finally, the author is endlessly grateful to her parents, siblings, and partner for their encouragement and support.

This thesis is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 681228.

Typeset in LaTeX using the IES Thesis Template.

Bibliographic Record

Kroupová, Kateřina: *The Impact of Student Employment on Educational Outcomes: A Meta-Analysis.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2021, pages 114. Advisor: doc. PhDr. Tomáš Havránek, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

BMA Bayesian Model Averaging

FAT Funnel Asymmetry Test

FE Fixed-effects Model

FMA Frequentist Model Averaging

IV Instrumental Variable

JCR Journal Citations Report

OLS Ordinary Least Squares

PCC Partial Correlation Coefficient

PET Precision Asymmetry Test

PIP Posterior Inclusion Probability

PMP Posterior Model Probability

RE Random-effects Model

SE Standard Error

SEM Simultaneous Equations Modelling

TSLS Two Square Least Squares

UIP Unit Information Prior

WAAP Weighted Average of Adequately Powered

Master's Thesis Proposal

Author	Bc. Kateřina Kroupová
Supervisor	doc. PhDr. Tomáš Havránek, Ph.D.
Proposed topic	The Impact of Student Employment on Educational Outcomes: A Meta-Analysis

Motivation Student employment during secondary as well as tertiary education has become a prevalent trend in many countries of the Western world (Neyt et al., 2019). More specifically, based on the results of 2012 Survey of Adult Skills, the share of students combining work and studies in OECD countries amounts approximately to 39% (PIAAC, 2012). As this share has been steadily increasing over the past years, the research concerning the relationship between student employment and educational attainment has grown accordingly. However, currently there is no consensus on to what extent student work improves or impairs academic performance.

The extant literature suggests two opposing views on the intersection of students' work behavior and their educational achievement. The first perspective classifies student employment as complementary to education, providing students with additional experience, knowledge and soft-skills (Darolia, 2014). The second perspective views student work as a substitute to education, crowding out time which should be devoted to academic activities, and thus having a detrimental impact on educational outcomes (ibid).

In spite of numerous attempts, the prior research failed to substantiate a conclusive evidence for the negative impact of student work on academic performance. Although most studies investigating the scrutinized relationship report a detrimental effect (Montmarquette et al., 2007; Oettinger, 1999; Darolia, 2014; Moulin et al., 2013), studies employing a more advanced methodological practices often show a negligibly small negative effect (Eckstein & Wolpin, 1999) or no effect at all (Buscha et al., 2011; Rothstein, 2007). Surprisingly, some research studies even demonstrate a positive effect of student part-time employment on their GPA if students work only a certain amount of hours during a week (Salamonson & Andrew, 2005; Quirk et al., 2001).

There are numerous reasons causing variation in the estimated results across different studies. First, a significant amount of empirical studies investigating the scrutinized relationship conducted their research in the absence of any endogeneity checks, which can potentially conceal the true effect (Darolia, 2014). In fact, Neyt et al. (2019) identify the endogeneity problem that pre-existing differences between working and non-working students might account for variation in educational outcomes as well as work behavior as the biggest methodological challenge of this research subject (p. 897). Second, the heterogeneity in the results can be also embedded in the varying features of employed methodological approaches as well as nature of the dataset, country of analysis, educational level or student job characteristics. Third, one cannot eliminate the presence of publication bias in the existing literature as researchers, editors and reviewers might have a strong preference for conventionally expected results (Stanley, 2005). In fact, Buscha et al. (2011) admit that "the view that part-time work has a detrimental effect on educational attainment, relying on academic research, is increasingly widespread in the last 10 years" (p. 383).

Given these ambiguities, the aim of this paper is to quantitatively review the extensive body of existing literature exploring the relationship between student employment and educational performance. Employing the method of meta-analysis will allow me to verify the presence of publication bias, to determine the factors systematically affecting reported findings, and finally to establish "the primary effect [sterilized] from background variation and contaminating influences" (Stanley, Doucouliagos & Jarrell, 2008, p. 2). Meta-analysis is a well-established method of quantitative literature review, vastly used in the field education economics (See for example Havranek, Irsova & Zeynalova, 2018; Groot & Van Den Brink, 2000), benefiting from an objective assessment of a voluminous amount of existing data (Stanley, Doucouliagos & Jarrell, 2008).

To this date, only two literature reviews focusing on youth work experience during their studies have been conducted (Neyt et al., 2019; Riggert et al. 2006). These reviews provide a comparison of the existing empirical research and suggest potential sources of heterogeneity and inconsistencies among available studies. However, none of the aforementioned summaries consider publication bias, nor they rely on objectively compiled sample of studies. Thus, from a theoretical perspective, a quantitative analysis of the empirical research results is vital in order to provide a systematic evaluation on this topic.

Furthermore, examining student work behavior while studying is also relevant on practical level. First, understanding the costs and benefits of working while studying is essential for prominent stakeholders of this issue (Darolia, 2014); for students in order to make a well-informed decision regarding their work experience alongside their studies and for career counselors to deliver their clients an educated advice.

Second, education policy makers should develop their recommendations concerning youth employment in the presence of clear empirical evidence (Ruhm, 1997). Hence, the effect between student work and academic achievement corrected for publication bias can serve the policy makers as a benchmark for designing future policy proposals.

Hypotheses

> Hypothesis #1: The literature estimating the impact of student employment on academic achievement is affected by publication bias.
>
> Hypothesis #2: The publication bias does not plague studies carefully controlling for endogeneity.
>
> Hypothesis #3: Accounting for the inherent endogeneity has a significant impact on the effect estimate.
>
> Hypothesis #4: The effect estimate significantly depends on method of estimation, educational level, country of analysis, student job characteristics, type of educational outcome and student characteristics.

Methodology First, I will construct my own dataset consisting of primary studies. I will use Google Scholar to search for primary studies. Later, when I assemble several primary studies, I will apply the snow-balling method. The chosen primary studies must fulfil these criteria: (1) the study investigates the impact of student employment on academic attainment, not on academic activities such as study time (e.g. Manthei & Gilmore, 2005; Kalenkoski & Pabilonia, 2009), (2) the study must report standard errors, and (3) the independent variable cannot be a dummy, e.g. working vs. non-working student. I will transform the collected effect estimates into partial correlation coefficients as each study uses different proxies for the dependent and independent variable. By standardizing the effect sizes the primary studies become comparable (Cazachevici, Havranek, & Horvath, 2019).

To test for publication bias, I will follow the example of Stanley & Doucouliagos (2010) and examine the asymmetry of the so-called funnel plot. To test for the presence of publication bias in a quantitative manner, I will examine the correlation between partial correlation coefficients and their standard errors. I will follow Gechert et al. (2019) and use the same specifications, if they will be feasible for my data. The same procedure will be conducted on a subset of studies carefully controlling for endogeneity.

To test how various study characteristics influence the reported outcomes, I will rely on Bayesian Model Averaging (BMA), a method running many regressions with different subsets of the additional explanatory variables. To make the estimation feasible, I will apply the Markov chain Monte Carlo algorithm estimating only the most important regression models. This approach is widely used in meta-analyses to account for model uncertainty (See Havranek, Irsova & Zeynalova, 2018). Each variable will be assigned a posterior inclusion probability (PIP) denoting the likelihood of including a certain variable in the true model. To decide which variables drive the heterogeneity, I will follow Jeffreys' categorization (1961) for interpreting the values of PIP.

Expected Contribution So far the academic literature lacked any meta-analysis quantitatively summarizing the existing research on the effect of the student employment on academic achievement. Therefore, my main contribution on the academic level lies in (1) testing whether publication bias plagues the studies estimating the scrutinized relationship, (2) examining how the employed estimation methods differently controlling for the endogeneity problem influence the findings, and (3) evaluating how the environment in which the effect have been studied as well as methodological aspects can capture variation in the reported estimates. On the practical level, results of my thesis can serve as guidelines for education policy makers working on recommendations concerning youth employment or school-to-work programmes integrating professional experience in study programs.

Outline

1. Introduction: I will highlight the lack of consensus in the existing literature and the possible sources of result variation to demonstrate the demand for systematic evaluation of empirical findings. Moreover, I will argue that publication bias is likely to permeate the existing literature, thus it is desirable to test for it.

2. Literature review

3. Estimating the relationship: I will review the methods which have been employed to measure the scrutinized effect, how the main variables (student employment and educational outcomes) were operationalised, and explain how various estimation methods dealt with the endogeneity problem.

4. The data set: I will describe the search strategy, define my selection criteria and explain the transformation of collected estimates into partial correlation coefficients. Then, I will provide the summary statistics of my dataset and list additional explanatory variables collected to examine the heterogeneity.

5. Methods

(a) Publication Bias: I will create a funnel plot and estimate the correlation between the effect estimates and the standard error to assess the presence of publication bias.

(b) Heterogeneity: I will explain BMA and the choices I will make when applying this method.

6. Results: I will discuss results of BMA for each additional explanatory variable and attempt to provide a theoretical explanations for the results.

7. Robustness Checks: I will replicate the analysis on a more homogeneous sample of estimates (same DV, IV measured in the same units, etc.).

8. Concluding remarks â€“ I will summarize my findings, state the implications for public policy purposes and mention the limitations of my study in case any emerge.

Core bibliography

Buscha, F., Maurel, A., Page, L., & Speckesser, S. (2012). The effect of employment while in high school on educational attainment: A conditional difference-in-differences approach. Oxford Bulletin of Economics and Statistics, 74(3), 380396.

Cazachevici, A., Havranek, T., & Horvath, R. (2019). Remittances and Economic Growth: A Meta-Analysis (No. 2019/35). Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

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.

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.

Gechert, S., Havránek, T., Havránková, Z., & Kolcunova, D. (2019). Death to the Cobb-Douglas production function (No. 201). IMK Working Paper.

Groot, W., & Van Den Brink, H. M. (2000). Overeducation in the labour market: a meta-analysis. Economics of education review, 19(2), 149-158.

Havránek, T., Iršová, Z., & Zeynalova, O. (2018). Tuition fees and university enrolment A meta-regression analysis. Oxford Bulletin of Economics and Statistics, 80(6), 1145-1184.

Jeffreys, H. (1961). Theory of Probability, 3rd Edn Oxford: Oxford University Press.

Kalenkoski, C. M., & Pabilonia, S. W. (2009). Does working while in high school reduce US study time?. Social Indicators Research, 93(1), 117-121.

Manthei, R. J., & Gilmore, A. (2005). The effect of paid employment on university students' lives. Education+ Training, 47(3), 202-215.

Montmarquette, C., Viennot-Briot, N., & Dagenais, M. (2007). Dropout, school performance, and working while in school. The Review of Economics and Statistics, 89(4), 752-760.

Moulin, S., Doray, P., Laplante, B., & Street, M. C. (2013). Work intensity and non-completion of university: longitudinal approach and causal inference. Journal of Education and Work, 26(3), 333-356.

Neyt, B., Omey, E., Verhaest, D., & Baert, S. (2019). Does student work really affect educational outcomes? A review of the literature. Journal of Economic Surveys, 33(3), 896-921.

Oettinger, G. S. (1999). Does high school employment affect high school academic performance?. ILR Review, 53(1), 136-151.

PIAAC. (2012). 2012 Survey of Adult Skills [Data Set]. OECD. Retrieved from https://www.oecd.org/statistics/students-at-work.htm.

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.

Riggert, S. C., Boyle, M., Petrosko, J. M., Ash, D., & Rude-Parkins, C. (2006). Student employment and higher education: Empiricism and contradiction. Review of educational research, 76(1), 63-92.

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.

Salamonson, Y., Everett, B., Koch, J., Andrew, S., & Davidson, P. M. (2012). The impact of term-time paid work on academic performance in nursing students: A longitudinal study. International journal of nursing studies, 49(5), 579-585.

Stanley, T. D., & Doucouliagos, H. (2010). Picture this: a simple graph that reveals much ado about research. Journal of Economic Surveys, 24(1), 170-191.

Stanley, T. D., Doucouliagos, C., & Jarrell, S. B. (2008). Meta-regression analysis as the socio-economics of economics research. The Journal of Socio-Economics, 37(1), 276-292.

Chapter 1

Introduction

Being employed while studying at a secondary or tertiary education institution is an indispensable part of students' lives in most countries of the Western world (Singh 1998; Tyler 2003; Marsh & Kleitman 2005). Based on the results of 2012 Survey of Adult Skills, the share of students aged 16-29 combining paid work and studies amounts to 39% in OECD countries (Quintini 2015). As the practice of taking up employment during studies has become a prevalent trend, we have witnessed a substantial growth in researchers' attempts to investigate the relationship between students' employment and their educational outcomes. Although the first attempts to examine the relationship date back to 1960s, the existing research offers little consensus on whether working while studying yields educational benefits or costs.

Theoretical mechanisms underpinning the student work-education relationship provide ambiguous predictions on the direction and origin of the effect. Student employment can be classified as a complement to education, providing students with additional experience, knowledge and soft skills. Oppositely, it can be classified as a substitute, reducing time devoted to academic activities and impeding educational outcomes (Darolia 2014). Alternatively, recently developed theoretical mechanisms stipulate that the relationship is conditional on students' pre-existing characteristics (Lee & Staff 2007) or primary orientation towards education and employment (Baert *et al.* 2018).

Hence, from the theoretical perspective, the impact of student employment on educational outcomes is ambiguous, leaving the search for definite answer to empirical research. Yet, the existing empirical evidence imitates the ambiguity embedded in the theoretical foundations. On one hand, number of studies find a negative impact of student work experience on educational outcomes

(Tyler 2003; Marsh & Kleitman 2005; Stinebrickner & Stinebrickner 2003). On the other hand, multiple authors report negligibly small effect (Eckstein & Wolpin 1999; Darolia 2014; Singh 1998) or no effect at all (Buscha *et al.* 2012; Rothstein 2007; Lillydahl 1990). Some studies even find a positive effect, e.g. when students work only a certain amount of hours during a week (Salamonson & Andrew 2006; Quirk *et al.* 2001). These differences might arise due to number of reasons: researchers are guided by various theories, the endogeneity of the decision to work is ignored or treated differently (Ruhm 1997; Rothstein 2007; Stinebrickner & Stinebrickner 2003), researchers include different sets of covariates, or the employed samples substantially vary in size (Darolia 2014; Kalenkoski & Pabilonia 2010; McKechnie *et al.* 2005).

As Riggert *et al.* (2006) point out, "critical reading of the empirical literature on student employment could legitimately lead different readers to different conclusions" (pg. 85). Given the contradictions on the theoretical and empirical level, the goal of this thesis is to synthesize results from the existing studies and deliver a meta-analytic review of the literature describing the relationship between student employment and educational achievement. Applying the modern methods of meta-analysis, we aim to estimate the true effect and explore to what extent the true effect is influenced by the possible publication bias induced by researchers' shared preference for a certain outcome. Furthermore, we aim to identify the methodological, publication and data-related factors explaining heterogeneity among the empirical results.

Admittedly, the existing literature contains articles attempting to explain variation in the student employment-education relationship. For instance, Riggert *et al.* (2006) provide a qualitative review of aspects causing inconsistencies in studies focusing on higher education. Similarly, Neyt *et al.* (2019) provide a systematic comparison of the empirical findings using descriptive statistics. Finally, combining five datasets from the US, Warren & Cataldi (2006) employ a simple meta-analysis to examine time patterns in the relationship between employment and high school dropout. We advance the meta-analysis by Warren & Cataldi (2006) in two ways. First, in addition to time factor, we inspect how other aspects systematically affect the relationship using Bayesian Model Averaging, an estimation method accounting for model uncertainty. Second, we extend our analysis to estimates capturing all types of educational outcomes, not only dropout decision and document students from different countries and educational levels. Moreover, to our best knowledge, this thesis constitutes the first meta-analysis estimating the true effect of student employment on academic outcomes while correcting for publication bias.

Our results show that a negative publication bias permeates the literature on student employment-education relationship. After filtering out this bias, we obtain a negligible mean effect estimate. This finding is further confirmed by our 'best-practice estimate', an estimate resulting from a synthetic study with predefined ideal conditions, which is close to 0. Furthermore, we observe that the negative publication bias persists even when we control for 26 additional study characteristics. In particular, our analysis of the systematic heterogeneity among reported estimates suggests that negative effect estimates are associated with specifying educational outcome as dropout decision, measuring employment as a continuous variable, number of citations primary study receives, and conducting the analysis in Europe. In contrast, positive estimates are linked with studies using longitudinal data and controlling for students' ability.

Studying the impact of student employment on educational outcomes is important both from the individual and societal perspective (Porter 1997). From the perspective of an individual student, knowing whether working improves or impairs educational performance is crucial as higher educational levels are associated with more favourable, future socioeconomic situation characterized by higher annual income and lower unemployment rate (Riggert *et al.* 2006). From the societal perspective, following neoclassical growth theory, education increases population's human capital, resulting in higher productivity and overall output (Mankiw *et al.* 1992). Alternatively, in line with theories on endogenous growth, higher levels of education contribute to producing innovations in the economy, promoting economic growth. Given its wide individual and societal implications, we feel encouraged to further investigate the student work-education relationship.

This thesis is organized as follows. Chapter 2 provides an overview of theoretical perspectives, explaining the mechanisms underlying the student employment-education relationship. Chapter 3 addresses various estimation strategies and the most challenging estimation obstacles including the endogeneity problem. In Chapter 4, we elaborate on the data collection procedure, outline the data adjustment process, and provide descriptive statistics of our sample. In Chapter 5, we introduce tests of publication bias and summarize its results. Chapter 6 explains Bayesian Model Averaging, addresses its implementation, and finally discusses the factors responsible for heterogeneity between the existing estimates. Finally, Chapter 7 encapsulates the main research results, its contributions and limitations.

Chapter 2

Theoretical Mechanisms

Theoretical mechanisms explaining the intersection between students' work behavior and their educational achievement fail to provide a clear prediction for the effect between these two variables. The traditional course of research relies on well-established models, including Developmental Model utilizing Human Capital Theory and Zero-Sum Model depending on Zero-Sum Theory. On the contrary, recent articles exploit modern theoretical mechanisms explaining the relationship through students' pre-existing differences and their primary orientation towards education and employment. In this section, we provide an overview of the traditional and modern perspectives explaining the theoretical mechanisms behind the investigated relationship.

2.1 Traditional Theoretical Mechanisms

2.1.1 Developmental Model

The first traditional model, building upon the Human Capital Theory (Becker 1965), views student employment as complementary to education enhancing students' human capital. As a result of skill and knowledge transmission from the work environment, the developmental model predicts that term-time employment positively affects educational outcomes (Marsh 1991). In practice, students' work activities can contribute to the development of their soft skills including problem-solving, organizational skills, time-management, communication, working under pressure, and presentation skills (Darolia 2014). If we assume that such skills are transferable, individuals working while studying might benefit from the acquisition of such skills in the academic setting (Buscha *et al.* 2012). Likewise, engagement in occupational activities allows students to apply their academic knowledge in real-life context augmenting their learning experience (Geel & Backes-Gellner 2012). Finally, Rothstein (2007) argues that an early-age work experience might aid students to ascertain their career goals and motivate them to work harder during their studies.

Students' testimonies concerning their motivation to work corroborate plausibility of these arguments. Stern & Briggs (2001), in their qualitative study, demonstrate that high school students perceive school and work as mutually reinforcing since both employment and school demand responsibility, time management skills, and communication with other people (pg. 370). Similarly, Wang *et al.* (2010) report that students enrolled in post-secondary institutions work primarily in order to gain work experience and practical skills.

2.1.2 Zero-Sum Model

The second model, utilizing Zero-Sum Theory (Becker 1965), classifies student work as a substitute to education, crowding out time which should be devoted to academic activities. Therefore, Zero-Sum Model posits that student employment exerts a detrimental impact on educational outcomes (Marsh 1991). As students possess only limited time resources, they face a time-allocation problem: they must distribute their time between leisure activities, studying and working. Devoting more time to work-related activities inevitably leads to less available time for academic pursuits. Hence, time spent working is likely to have a deleterious effect on students' academic productivity as it reduces time available for homework and independent study (Choi 2018; D'Amico 1984). Additionally, besides reduced time for studying, Darolia (2014) argues that working also impairs involvement in academic community undermining students' educational commitment and aspirations. Furthermore, Oettinger (1999) suggests that employment engagement produces excessive fatigue, decreasing students' attentiveness.

Despite the aforementioned arguments, spending time at student work does not have to necessarily detract from time devoted to study-related activities. For instance, Body *et al.* (2014) suggest that students are more likely to decrease their leisure time instead of study time. Furthermore, Kalenkoski & Pabilonia (2012) show that although working high school students from the US tend to lower time devoted to homework, they decrease time spent on free-time activities substantially more.

2.1.3 Threshold Model

The so called 'Threshold model' reconciles the theoretical mechanisms behind Developmental and Zero-Sum Models and assumes a non-linear relationship between working and academic success (Marsh & Kleitman 2005). Using the Zero-Sum Theory and Human Capital Theory, Neyt *et al.* (2019) explain that the first hours of employment contribute the most to gaining valuable transferable skills because during these hours the marginal benefits of working are the highest. With increasing working hours, the marginal benefits of student employment decrease and begin to replace time crucial for successful academic growth. Therefore, depending on the intensity of student employment, some studies show that working may be simultaneously a complement and a substitute to academic performance (Choi 2018). This intensity-dependent perspective holds that working has positive consequences on study engagement only up to a certain threshold of hours worked. After exceeding this threshold, the effect of student employment on educational outcomes reverses as working hours begin to interfere with academic pursuits (Buscha *et al.* 2012). Admittedly, the literature diverges in stating the actual hours threshold, at which the effect reverses (Marsh & Kleitman 2005). While Montmarquette *et al.* (2007) report an inflection point of 15 hours worked per week, Tessema *et al.* (2014) find this point at 10 hours worked per week.

2.2 Modern Theoretical Mechanisms

2.2.1 Primary Orientation Perspective

In contrast to the traditional theoretical mechanisms, modern theoretical perspectives do not assume that student employment is the most decisive factor in determining the relationship between student employment and academic consequences (Warren 2002). Instead, they posit that the relationship may be largely driven by pre-existing differences between working and non-working students (Rothstein 2007).

The Primary Orientation Perspective, also called selection-to-work perspective (Choi 2018) or self-selection perspective (Lee & Staff 2007), holds that various socio-psychological factors including family attitudes towards education, motivation, and educational aspirations form altogether an individual commitment towards education or work experience driving the investigated relationship (Warren 2002). More specifically, the perspective suggests that student employment has more deleterious educational effect for students oriented primarily towards work unlike to students oriented primarily towards education. This difference reflects students' educational engagement or disengagement, which developed before their decision to participate in the labour market (Warren 2002). Hence, the relationship between employment and school performance becomes non-significant or less pronounced as long as researchers account for the primary orientation of students (Choi 2018; Lee & Staff 2007; Warren 2002). In that sense, the Primary Orientation Perspective does not attempt to provide a causal explanation for the scrutinized effect. Rather, it emphasizes students' self-selection process to take up employment, depending on their observable and unobservable pre-existing characteristics forming their prior orientation towards work and education (Neyt *et al.* 2019). Baert *et al.* (2018) investigate the Primary Orientation Perspective empirically and find that estimation using students' primary orientation as a mediator yields a less negative effect between employment and GPA compared to estimation omitting the mediator.

2.2.2 Heterogenous Effect Perspective

The Heterogenous Effect Perspective proposes that the relationship between student work and educational outcomes is directly conditional on individual and student job characteristics (Lee & Staff 2007). "This perspective emphasizes that assuming homogeneous effects of student employment on academic outcomes without considering sources of effect heterogeneity limits an accurate understanding of the impacts of student employment" (Choi 2018, pg. 92). Existing research provides vast evidence for the conditional nature of the relationship between student employment and educational performance; the effect and its magnitude varies greatly by ethnic group (D'Amico 1984), gender (Buscha *et al.* 2012; Holford 2020), job type (McNeal 1997; Sabia 2009), motivation to work (Wenz & Yu 2010), job industry (Dadgar 2012), and educational level (Neyt *et al.* 2019). Failing to take into account the personal and work-environment characteristics results in an incomplete understanding of the examined relationship.

To illustrate this better, consider the distinction between students attending secondary and post-secondary education institution. Compared to high school pupils, university students represent a specific group of youth with a relatively high interest in education as the most academically disengaged students moved to the labour market immediately after high school graduation (Bozick 2007). Furthermore, the organizational structures of secondary and post-secondary education are substantially different; university students have less contact hours and more flexibility in terms of workload and choice of courses (Bozick 2007). Hence, given these varying conditions and characteristics, it is reasonable to assume that the effect of student work on scholastic performance differs between these two groups of students.

Chapter 3

Estimating the Relationship

Credibility of empirical research concerning the effect of student employment on educational outcomes depends on proper controlling for pre-existing heterogeneity among students (Choi 2018). Unless studies account for observable and unobservable characteristics which simultaneously determine students' decision to work and academic performance, they report erroneous results due to the endogeneity problem. In this chapter, we discuss the endogeneity problem and its sources in detail. Furthermore, we discuss the most popular estimation methods and address the varying measures used as proxies for student employment and educational outcomes.

3.1 Endogeneity Bias

Researchers examining the effect of student employment on educational outcomes usually estimate variation of the following model:

$$AP_{it} = \beta_0 + \beta_1 Emp_{it} + \beta_2 X_{it} + \epsilon_{it}, \tag{3.1}$$

where AP_{it} denotes academic performance of student i in time period t, Emp_{it} represents student employment, X_{it} corresponds to the set of included control variables and ϵ_{it} is the error term.

The existing literature implies that researchers encounter one major problem when estimating Equation 3.1: the potential endogeneity of student employment and academic performance. Empirically, the potential endogeneity of students' decision to work represents the main challenge in estimating the true effect as it undermines any attempt to interpret the link between these two variables as causal (Beffy *et al.* 2013). The existing literature suggest two sources of potential endogeneity.

The first source holds that students' labour supply decision is determined by both observable (e.g. family background, gender, ethnicity, etc.) and unobservable characteristics (e.g. ability, motivation, work ethic, time preference, social and peer networks, etc.) that simultaneously influence students' academic performance (Beffy *et al.* 2013). In particular, these characteristics may systematically differ between students who participate in the labour market and students who do not (Rothstein 2007). Thus, if a certain estimation method fails to appropriately control for the pre-existing heterogeneity between students, we cannot conclude that the estimated effect of student employment on educational outcomes is directly attributable to students' employment.

Econometrically, the first source of endogeneity encompasses both the omitted variable bias and the selection bias. The omitted variable bias occurs when researchers fail to include relevant personal observable characteristics under X_i, while the selection bias arises when researchers do not account for students' unobserved characteristics in their methodological approach, e.g. when estimating Equation 3.1 with Ordinary Least Squares (OLS). Estimating Equation 3.1 with OLS yields inconsistent estimates of β_1, unless the exogeneity assumption that student employment is uncorrelated with the error term $Cov(Emp, \epsilon_{it}/X)$ $= 0$ is satisfied. As X_{it} can only include observable controls affecting academic performance, the error term ϵ_{it} consists both of a random component ζ_{it} and a component related to unobserved characteristics μ_{it}, taking the form of $\epsilon_{it} = \mu_{it} + \zeta_{it}$. If these unobserved characteristics μ_{it} are also predictors of Emp_{it}, the disturbance term of Emp_{it} takes the form of $\nu_{it} = \mu_{it} + \psi_{it}$, consisting of the unobserved personal characteristics and a random term. In this case, the correlation between Emp_{it} being a predictor variable of AP_{it} and ϵ_{it} is nonzero, resulting in an inconsistent OLS estimator of β_1.

The OLS estimator of β_1 can be biased both positively (upward) or negatively (downward). To illustrate the opposite directions of the bias, consider students attributed with high levels of ability, a characteristic that is difficult to control for. On one hand, higher levels of ability make these students more likely to participate in other human capital enhancing activities such as work (Stinebrickner & Stinebrickner 2003). In this case, simple econometric methods such as OLS understate the negative impact of student employment on academic performance because thanks to their superior pre-existing characteristics such as ability, these students can successfully combine work and studies.

Hence, OLS generates an upward-biased estimate. On the other hand, if such students are less likely to participate in the labour market, e.g. because they prefer devoting their energy to academic pursuits, the resulting OLS estimator overstates the negative effect of student work. The OLS estimate is overstated because it mainly reflects the effect of students with low ability, who have worse academic performance due to their lower ability, and decide to pursue work experience not to increase their human capital, but to gain more relevant skills for their future career. In this scenario, the OLS estimator is downward-biased.

The second source of endogeneity concerns reverse causation, a situation when both variables are jointly determined resulting in the simultaneous equations bias (DeSimone 2006). This occurs when the estimated effect of student work on academic performance partly reflects a causal impact of academic performance on student work.

The second type of endogeneity bias, the simultaneous equations bias, usually permeates cross-sectional studies. In these studies, researchers do not distinguish between time periods, at which the student employment and academic achievement are measured and observe these variables simultaneously at time t. In consequence, the causal link between these two variables is unclear as we cannot say which variable causes the other (Warren *et al.* 2000). In the endeavour to circumvent this problem, researchers hinge on the availability of longitudinal data (See, for inatance, Apel *et al.* 2008; Lee & Orazem 2010; Kalenkoski & Pabilonia 2010). Longitudinal data overcome the issue of mismatched time periods for used variables: student employment observed at time t is used as a regressor for educational outcomes measured at time $t + x$, where x represents several months or years (Warren *et al.* 2000). Considering the indisputable time order between measuring student employment and educational outcomes, longitudinal data allow researchers to draw causal inferences between these two variables.

Similarly to correcting for reverse causation by using longitudinal datasets, authors of later studies have seeked a solution to the first type of endogeneity. Early studies, utilizing OLS and other elementary estimation methods, treat student employment as exogenous, failing to appropriately account for the endogeneity of the decision to work (Ruhm 1997). Although some of these early studies accommodate multiple control variables to account for observable individual differences, Stinebrickner & Stinebrickner (2003) suggest that "including individual characteristics provides only a relatively low level of protection against this type of endogeneity" (pg. 474). Therefore, early studies generate naive estimates of the examined effect as they cannot control for all observable and unobservable characteristics, demonstrating the need for advanced methodological techniques. In the following subsection, we review methodological solutions researchers use to mitigate endogeneity caused by students' self-selection into employment.

3.2 Estimation Methods

Propensity Score Matching. Some studies address the endogeneity of employment decision using the propensity score matching that accounts for observable heterogeneity between working and non-working students (Choi 2018; Lee & Staff 2007). Propensity score matching technique pairs working and non-working students based on their degree of similarity in various observable socio-psychological and demographic characteristics composing together the propensity score (Lee & Staff 2007). Consequently, the effect of student employment on educational outcomes is compared between the matched students. Despite its attractiveness, propensity score matching method fails to deliver unbiased results due to the violation of the conditional independence assumption. This assumption asserts that all variables affecting the construction of propensity score and the outcome variable must be observable. In light of previous discussion on unobserved individual characteristics, such a condition is hard to fulfill.

Fixed-Effect Model and Random-Effects Model. One solution allowing researchers to control for unobserved differences between working and non-working students entails adding individual unobserved fixed-effects μ_i into the linear regression model:

$$AP_{it} = \beta_0 + \beta_1 Emp_{it} + \beta_2 X_{it} + \mu_i + \epsilon_{it}. \tag{3.2}$$

By subtracting the individual-specific means from the variable values at each time period, the fixed-effects (FE) model allows researchers to control for the time-invariant student-level unobserved characteristics (Darolia 2014). However, as noted by Apel *et al.* (2008), the fixed-effects model yields unbiased and consistent estimates only under the assumption that unobserved student characteristics determining student work habits and academic performance are constant over time. As explained by Oettinger (1999), this assumption is questionable as students' motivation is likely to fluctuate over time. Typically, students pursuing enrollment into tertiary education institution increase their academic effort before their high school leaving exams in order to enhance their chances of being accepted to their top-choice universities.

Apart from the fixed-effects model, a common alternative used by researchers is the random-effects (RE) model (See, for instance, Staff *et al.* 2010; Apel *et al.* 2008). However, unlike to the fixed-effects model, the random-effects model assumes that the unobserved individual characteristics are uncorrelated with the explanatory variables in the model. Considering the discussion on the possible link between students' innate motivation and their work involvement in Subsection 2.2.1, such assumption is rarely satisfied. Hence, RE model also fails to provide unbiased and consistent estimates.

Multi-equation Modelling Approach. An ideal solution to obtain a consistent estimate of β_1 is the Two-Square Least Square (TSLS) instrumental variable procedure, consisting of a multi-equation model:

$$Emp_{it} = \alpha_0 + \alpha_1 Z_{it} + \alpha_2 X_{it} + \nu_{it}, \tag{3.3}$$

$$AP_{it} = \beta_0 + \beta_1 Emp_{it} + \beta_2 X_{it} + \epsilon_{it}. \tag{3.4}$$

In Equation 3.3, which models students' decision to participate in the labour market Emp_{it}, Z_{it} represents a vector of selected instrumental variables, while vector X_{it} refers to the exogenous determinants of a particular educational outcome. The TSLS procedure estimates the effect of student employment on their academic performance in two subsequent stages. In the first step, the student employment is regressed on the chosen instrumental variable(s) and other control variables. In the second step, the estimated values of student employment are used as an explanatory variable for the academic performance.

For TSLS to yield a consistent estimate of β_1, the vector of instrumental variables Z_{it} must fulfill two conditions. First, the instrument relevance condition requires the set of instrumental variables Z_{it} to be correlated with the endogenous variable Emp_{it}, $\text{Cov}(Z_{it},\, Emp_{it}) \neq 0$. Second, the instrument exogeneity condition requires the instrument to be uncorrelated with the error term, $\text{Cov}(Z_{it},\, \epsilon_{it}) = 0$.

Many researchers taking advantage of the instrumental variable (IV) approach hinge on the availability of local labour market conditions, e.g. youth unemployment rate, as the instrumental variable (Rothstein 2007; Beffy *et al.* 2013; Holford 2020; Lee & Orazem 2010). Other articles use child labour laws

(Tyler 2003; Apel *et al.* 2008), the proportion of unearned income (DeSimone 2006), paternal schooling (DeSimone 2008), socio-economic status of students' families (Simón *et al.* 2017), amount of financial aid students obtain (Sprietsma 2015), or the variation in area house prices (Darolia 2014) as their instrumental variables.

Next to the instrumental variable estimation, some researchers rely on the simultaneous equation modelling (SEM) approach (Parent 2006; Kalenkoski & Pabilonia 2010). Identically to to the instrumental variable approach, the SEM approach models the effect of student work on educational outcome by estimating a system of linear equations (Kalenkoski & Pabilonia 2010). However, instead of relying on the TSLS estimator, the simultaneous equations model is usually estimated via maximum likelihood estimator (Kalenkoski & Pabilonia 2010).

Dynamic Discrete Model. The last method overcoming the endogeneity bias is the dynamic discrete approach explicitly modelling students' decision-making process to work (See, for example, Eckstein & Wolpin 1999; Montmarquette *et al.* 2007). In essence, the method estimates the likelihood function of participating in the labour market exploiting the finite number of discrete types of students who differ in the unobservable characteristics (Eckstein & Wolpin 1999). The dynamic discrete model approach advances the other estimation methods in two ways. First, it allows joint modelling of student employment and academic performance. Second, it explicitly controls for students' unobserved individual traits.

3.3 Variable Operationalization

3.3.1 Measuring Student Employment

Besides various estimation methods, reviewed studies also differ in terms of operationalizing the employment variable. The majority of existing studies estimate the effect of *employment intensity* on educational outcomes, while the rest examines the effect of *employment status* on educational outcomes. Researchers using employment status as the independent variable simply distinguish between working and non-working students, defining student employment as a dummy variable (See, for example, McKenzie & Schweitzer 2001; McNeal 1997). In contrast, researchers using the specification of employment intensity define the variable either as a continuous or as a categorical variable.

We find different continuous measures of student employment intensity in the existing studies. For instance, Carr *et al.* (1996) use total hours worked during a semester to estimate the effect, while D'Amico (1984) relies on the percentage of the school year's weeks with work hours being either above or below 20 hours. Nonetheless, researchers usually measure the intensity of student employment as average hours worked during the interview week (Ruhm 1997), a typical non-summer week (Sabia 2009), midterm week (Kalenkoski & Pabilonia 2010), or during two reference weeks in the academic year (Darolia 2014). However, as explained by Oettinger (1999), imputing the typical or survey week's hours worked to the entire school year might contribute to a significant measurement bias. To correct this bias, Oettinger (1999) suggests combining the amount of weeks worked during the year and the average weekly hours worked in the resulting student employment measure. Contrarily to Oettinger (1999), Ruhm (1997) argues that work hours reported for the week preceding the survey might better reflect the reality than work hours reported for periods preceding the survey by several months, given the time proximity and ease of remembering.

Similarly to the continuous variable specification, the categorical specification of student employment intensity may also take various forms (See, for example, Gleason 1993; Torres *et al.* 2010; Staff *et al.* 2010). For instance, Hovdhaugen (2015) divides his sample into 3 bands: 1–19 hours per week, 20–30 hours per week, and more than 30 hours per week. Alternatively, Torres *et al.* (2010) use 5 work intensity categories and Tyler (2003) uses 10 categories, each representing a 5-hour increment.

3.3.2 Measuring Educational Outcomes

Likewise to student employment, the dependent variable *educational outcomes* can take various forms as educational outcomes encompass a wide range of academic goals and results. Neyt *et al.* (2019) distinguish four classes of educational outcomes: educational engagement, educational decisions, exam/test results, and educational attainment.

Educational engagement refers to students' habits associated with their class preparation and discipline they display in school-related activities. This category embraces measures such as class attendance/absence (Schoenhals *et al.* 1998), time spent doing homework or devoted to independent study (Marsh & Kleitman 2005), truancy (Staff *et al.* 2010), or paying attention during class

(Sabia 2009). The second category of educational choices represents the decisions students make during their academic career, including the decision to drop out from a course or a study programme (Warren & Cataldi 2006) and to enroll in further educational level, e.g. the tertiary institution (See, for instance, Steel 1991). The third category, exam and test results, embodies the most employed measure of educational outcomes. Studies, in particular those conducted in the US, employ students' grade point average as the outcome variable (DeSimone 2008; Gleason 1993; Sabia 2009). Other researchers utilize specific course grades (Kouliavtsev 2013), test scores (Tyler 2003), or results of high school final exams (Dustmann & Soest 2007). The very last category, educational attainment, comprises of students' probable and actual achievements, e.g. probability of graduation from the secondary school (Beffy *et al.* 2013) or credits earned during a specific time period (Dadgar 2012).

Taken as a whole, studies examining the relationship of student employment on educational outcomes differ in many aspects, including their approach to account for the endogeneity bias, chosen estimation method and operationalization of the main variables. Moreover, considering the varying estimates, these studies provide an inconsistent lesson about the investigated relationship. Researchers themselves admit that "the range of findings may be an artifact of the different operationalisations" concealing the true effect (McNeal 1997, pg. 208). Hence, it seems viable to conduct a quantitative analysis of the voluminous research on this topic to determine factors systematically affecting reported findings and examine the presence of a potential publication bias. In the next chapter, we turn to explaining the construction of a dataset allowing us to fulfill these objectives.

Chapter 4

Data

To examine the effect of student employment on educational outcomes, we employ the quantitative research design of meta-analysis. To use this method, we construct an original dataset including 861 effect estimates from 69 research papers. In this section, we describe the data collection process, the adjustments we perform to collected data, and finally descriptive statistics of our dataset.

4.1 Data Collection

As a very first step of creating our dataset, we compile a list of primary studies using the Google Scholar search engine. Our search query has been repeatedly modified to cover 10 essential studies in the first 40 hits of the search. We employ the following word combination as our definitive search query[1]:

'employment and academic performance'

We examine abstracts and result sections of the first 500 studies to assess the presence of desirable estimates. After that, applying the snowballing method, we review research papers included in reference lists of the previously inspected studies. Additionally, we check studies cited by Neyt *et al.* (2019) in their multidisciplinary literature review. We terminate the search for primary studies in May 2020.

Primary studies included in our final sample adhere to the following criteria. First, the studies examine the impact of student term-time employment on educational outcomes. Second, selected studies report uncertainty measures

[1]See Appendix A for a list of attempted search queries accompanied by reasons for their omission.

around the estimates such as standard errors, t-statistics or p-values. Moreover, contrarily to Cazachevici *et al.* (2020), we use only studies explicitly reporting uncertainty measures and defer from averaging expected p-values based on the number of asterisks or other graphical means denoting the significance level of an estimate (See Marsh & Kleitman 2005; McCoy & Smyth 2007; Wang *et al.* 2010 for such a type of notation of statistically significant coefficients).

To elaborate on the first criterion, specifically on the response variable, we do not incorporate studies operationalizing educational outcomes as academic engagement such as time spent doing homework or class preparation (See Marsh & Kleitman 2005; Manthei & Gilmore 2005; Schoenhals *et al.* 1998). From our perspective, educational engagement does not represent an outcome, rather a process leading to a comparable educational outcome such as exam score or enrollment into further level of education. Moreover, measures of academic engagement are almost always self-reported being a subject to an individual over- or under-estimation (Applegate & Daly 2006). Therefore, given the subjectivity of such measures, their inclusion in the sample would introduce an additional bias. In a similar spirit, we exclude studies defining the dependent variable as time to obtain a degree (See Theune 2015) because this operationalization might be considerably affected by current trends in study patterns such as the habit of taking gap years and other social factors, e.g. prolonging studies in order to exploit tax benefits of the student status.

Further, we do not use studies focusing on student employment in the primary school setting (See Post & Pong 2000) since in this context student work is illegal and rare. Similarly, we discard studies examining the impact of sandwich work placement, a year-long integrated period of work experience in students' study programme, on student academic performance. We disregard these studies because such programmes are specifically designed to be part of the curriculum with the aim to enhance student academic performance (Jones *et al.* 2017; Scott-Clayton & Minaya 2016). Finally, we exclude studies investigating the relationship between summer employment and educational outcomes (See Leos-Urbel 2014) and strictly adhere to research papers focusing on term-year employment, which is "much more salient because of its infringement on attendance, study habits, and achievement" (McNeal 1997, pg. 210).

To complete the list of conditions for including primary studies in our dataset, there are two practical obstacles preventing us from accommodating all existing research papers on this topic into our sample. These are author's language limitations and study accessibility. The former halts us from analyzing research papers published in other languages than English (See, for instance, Lanzarini *et al.* 2015; Ghavam *et al.* 2005). The latter forces us to disqualify studies inaccessible with the standard university license, although such studies have shown potential for providing desired estimates since they appeared as results of our search query. See Appendix A for the list of unavailable studies.

Following Stanley (2001), no study is disqualified on the basis of publication form. Thus, besides standard research articles from reviewed journals, we employ working papers, conference articles, MA theses, and PhD dissertations in the final sample and control for study quality in the analysis. Table 4.1 provides the final list of primary studies utilized in our meta-analysis.

Consequently, we gather individual effect estimates and their uncertainty measures from the primary studies included in our list. Apart from these coefficients and their standard errors, we also collect other variables to analyze the systematic heterogeneity among the reported coefficients. These variables include estimation characteristics, specific data features, dummy variables for control variables, and publication characteristics. We describe these variables in detail in Section 6.1.

During extracting effect estimates from original studies and coding study characteristics, we have experienced various difficulties related to insufficient reporting practices in primary research papers: presenting effect estimates and uncertainty measures such as t-statistics with erroneous signs (Rochford *et al.* 2009; DeSimone 2008), failing to specify the form of uncertainty measures accompanying the effect coefficients (Simón *et al.* 2017), making mistakes in placing decimal separators in reported values (Tienda & Ahituv 1996), and denoting standard errors associated with estimates as standard deviations (Beerkens *et al.* 2011; Body *et al.* 2014; Callender 2008). Besides imprecise reporting, some research papers lack a thorough description of control variables employed in original regression analyses enhancing the probability of coding some study characteristics erroneously. We completed data collection in August 2020. The final dataset is available in the attached ZIP file to this thesis.

4.2 Data Adjustments

After finalizing data collection, we engage in data adjustment procedures, ensuring that the dataset includes comparable effect estimates. First, unlike to most studies reporting standard regression coefficients, several studies employ the method of a logistic regression and report odds ratios as the effect estimates

Table 4.1: Primary studies used in the meta-analysis

Apel et al. (2008)	Gleason(1993)	Rochford et al. (2009)
Applegate & Daly (2006)	Hawkins et al. (2005)	Rothstein (2007)
Arano & Parker (2008)	Holford (2020)	Sabia (2009)
Auers et al. (2007)	Hovdhaugen (2015)	Salamonson & Andrew (2006)
Baert et al. (2017)	Hwang (2013)	Savoca (2016)
Baert et al. (2018)	Jaquess (1984)	Simon et al. (2017)
Beerkens et al. (2018)	Joensen (2009)	Singh et al. (2007)
Beffy et al. (2013)	Jones & Sloane (2005)	Sprietsma (2015)
Body et al. (2014)	Kalenkoski & Pabilonia (2010)	Staff & Mortimer (2007)
Bozick (2007)	Kohen et al. (1978)	Staff et al. (2010)
Buscha et al. (2012)	Kouliavtsev (2013)	Steel (1991)
Callender (2008)	Lee & Orazem (2010)	Steinberg et al. (1982)
Canabal (1998)	Lee & Staff (2007)	Stinebrickner & Stinebrickner (2003)
Carr et al. (1996)	Maloney & Parau (2004)	Tienda & Ahituv (1996)
Choi (2018)	McKechnie et al. (2005)	Torres et al. (2010)
Dadgar (2012)	McKenzie & Schweitzer (2001)	Trockel et al. (2000)
D'Amico (1984)	McNeal (1997)	Tyler (2003)
Darolia (2014)	McVicar & McKee (2002)	Warren & Cataldi (2006)
DeSimone (2006)	Montmarquette et al. (2007)	Warren & Lee (2003)
DeSimone (2008)	Oettinger (1999)	Warren et al. (2000)
Dustmann & Van Soest (2008)	Parent (2006)	Wenz et al. (2010)
Eckstein & Wolpin (1999)	Paul (1982)	Yanbarisova (2015)
Ehrenberg & Sherman (1987)	Richardson et al. (2013)	Zhang & Johnston (2010)

(See, for example, Bozick 2007 or Warren & Lee 2003). An odds ratio represents an univariate transformation of the estimated beta coefficient quantifying the constant effect of the independent variable on the occurrence likelihood of an outcome (Sribney & Wiggins n.d). We transform the reported odds ratios into the regression coefficients using the inverse form of the following formula:

$$or_{is} = e^{\beta_{is}}, \tag{4.1}$$

where or_{is} denotes the reported odds ratio and β_{is} denotes the desirable effect estimate from i-th specification in study s. To calculate the corresponding standard error from the odds-ratio adjusted standard error, we follow Oehlert (1992) and take advantage of the Taylor series-based delta method. The delta method allows us to approximate the variance of a random variable function by expanding the function with the variable mean (Oehlert 1992). We apply the delta method following this formula:

$$SE(or_{is}) = \sqrt{e^{\beta_{is}} Var(\beta_{is}) e^{\beta_{is}}} = SE(\beta_{is}) e^{\beta_{is}}, \tag{4.2}$$

where $SE(or_{is})$ refers to the odds-ratio adjusted standard error and $SE(\beta_{is})$ to the standard error corresponding to the original coefficient estimate from i-th specification in s-th study. As we want to calculate standard error of the original regression coefficient, we apply Equation 4.2 reversely.

Similarly, some studies examine the non-linear effect between student employment and educational outcomes and report an estimate for a quadratic term. To account for the presence of two estimates of the same effect, we follow Zigraiova & Havranek (2016) and linearize the investigated effect as follows:

$$\beta_{is} = \widehat{\beta}_{lis} + \widehat{\beta}_{qis}\overline{x}_{es}, \tag{4.3}$$

$$SE(\beta_{is}) = \sqrt{SE(\widehat{\beta}_{lis})^2 + SE(\widehat{\beta}_{qis})^2\overline{x}_{es}}, \tag{4.4}$$

where $\widehat{\beta}_{lis}$ refers to the linear estimate of the student employment coefficient, $\widehat{\beta}_{qis}$ corresponds to the estimate of the quadratic term of the student employment coefficient in i-th estimation from study s, and finally $\overline{x}_{es}$ denotes the sample mean of the student employment measured in study s. Further, $SE(\beta_{lis})$ denotes the standard error for the estimate of the linear term and $SE(\beta_{qis})$ denotes the standard error for the estimate of the quadratic term.

Finally, two studies in our dataset estimate an interaction effect between the student employment and other variables, e.g. female students (Steel 1991; Carr *et al.* 1996) and whether student graduated from secondary school (Steel 1991). We follow Cazachevici *et al.* (2020) and calculate the average marginal effect of student employment on academic performance and the corresponding standard errors using the delta method:

$$ME_{is} = \widehat{\beta}_{lis} + \widehat{\beta}_{tis}\overline{x}_{is}, \tag{4.5}$$

$$SE(ME_{is}) = \sqrt{SE(\widehat{\beta}_{lis})^2 + SE(\widehat{\beta}_{tis})^2\overline{x}_{is}}, \tag{4.6}$$

where ME_{is} corresponds to the calculated marginal effects of student employment, $\widehat{\beta}_{lis}$ denotes the estimated linear effect size of student employment, and $\widehat{\beta}_{tis}$ corresponds to the estimate of the included interaction term in i-th estimation from study s. Finally, apart from the mean value of the variable included in the interaction term $\overline{x}_{is}$, the calculation of the corresponding standard errors utilizes the reported standard error associated with the estimate of the linear term $SE(\beta_{lis})$ and standard error of the estimated coefficient for the interaction term $SE(\beta_{tis})$. Admittedly, some research papers (e.g. Carr *et al.* 1996; Ruhm 1997) fail to provide summary statistics or uncertainty measures for quadratic/interaction terms, yielding the transformations infeasible and preventing us from calculating the resulting estimates.

As noted in Section 3.3, researchers use different proxies for measuring educational outcomes. To ensure that the collected estimates reflect the effect of student work on educational outcomes in a unified way, we adjust signs for some of the collected estimates. For instance, McNeal (1997) defines the educational outcome as students' dropout likelihood, whereas Carr *et al.* (1996) measure the educational outcome as the likelihood of completing secondary education. Combining estimates from these studies would be problematic as they yield estimates with opposite signs when essentially reporting the effect of the same direction. In other words, if student employment has a detrimental effect on educational attainment, studies using the dropout likelihood as a proxy report a positive effect, whereas studies using the likelihood of completing a certain education level as a proxy report a negative effect. To further illustrate this issue, consider a German grading scheme employed by Sprietsma (2015), in which 1 is the best grade while 6 is the worst or a class rank used by D'Amico (1984), where lower rank signifies better academic performance. In both cases, we interpret a negative effect estimate as a positive impact of student employment on academic achievement. Hence, to make these estimates comparable with other collected coefficients, we reverse their signs.

Even after performing all aforementioned adjustments, the collected estimates still vary in their econometric specification and measurement units. Following a well-established transformation method employed among others by Havranek *et al.* (2016), Cazachevici *et al.* (2020), and Doucouliagos & Laroche (2003), we standardize the effect sizes of the collected estimates by converting them into partial correlation coefficients (PCC). PCCs make the reported study estimates directly comparable as they represent a unitless measure of strength and direction of the relationship holding other variables constant. Signs of PCCs correspond to the modified signs of the estimates. Using the following formulas, we calculate the partial correlation coefficient and the standard error:

$$PCC_{is} = \frac{t_{is}}{\sqrt{t_{is}^2 + df_{is}}}, \tag{4.7}$$

$$SE(PCC)_{is} = \sqrt{\frac{1 - PCC_{is}^2}{df_{is}}}, \tag{4.8}$$

where PCC_{is} represents the partial correlation coefficient of i-th estimate reported in study s, $SE(PCC)_{is}$ represents the corresponding standard error, t_{is} denotes the t-statistic, and df_{is} represents the number of degrees of freedom.

Our final step in data transformation entails two adjustments. First, Callender (2008) and Sabia (2009) report zero standard errors for some of their estimates. Such values do not allow us to compute PCCs. Hence, we replace standard errors equal to 0 by 0.001, a sufficiently low value. Second, despite the cleaning process, we still observe a few 'extreme' values of the calculated PCCs and corresponding standard errors. To prevent loss of information provided by these observations, we substitute these outliers and their uncertainty measures by values winsorized at the 1% level.

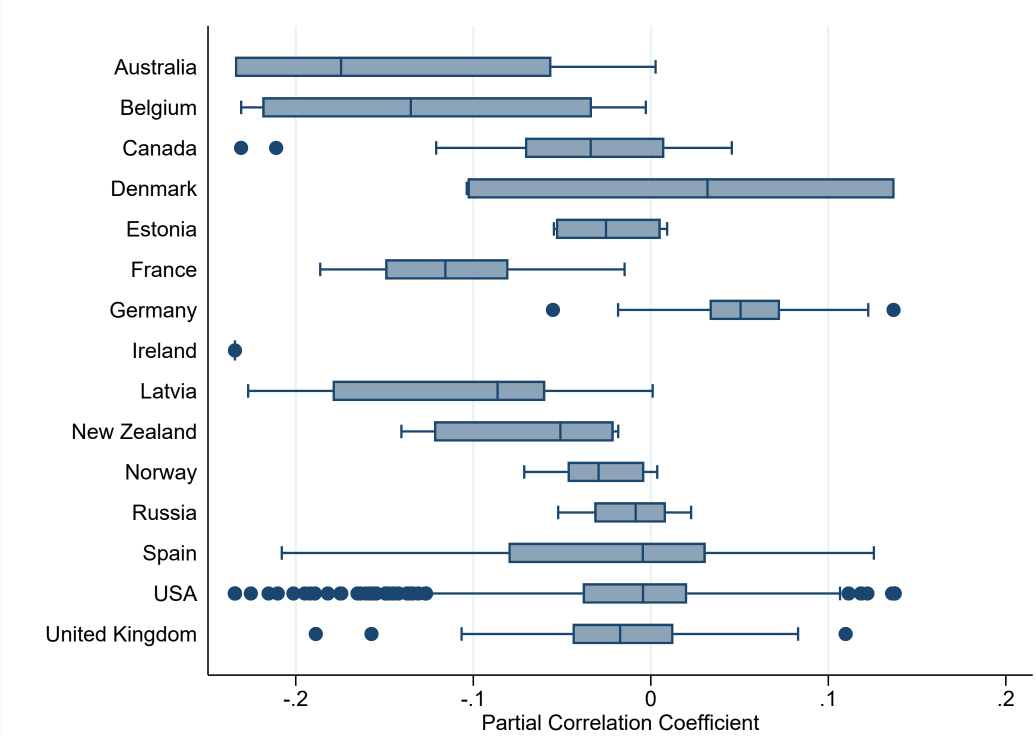
4.3 Descriptive Evidence

Before we dive into the analysis of publication bias and heterogeneity, we shed more light on the created dataset by presenting summary statistics. Following the inclusion criteria and performing data adjustments, our dataset contains 861 partial correlation coefficients from 69 primary studies capturing the effect of student employment on academic achievement. The oldest study in our sample is published in 1978, while the newest one is published in 2020. Within this time-frame, studies included in the sample are relatively equally distributed in terms of their publication year, demonstrating a stable on-going interest in the topic we examine. In Figure 4.1, we show a box plot of PCCs across countries, demonstrating a moderate variability among the effect estimates.

The calculated PCCs describing the effect of student employment on academic achievement can be characterized by a median of -0.006 and a simple mean of -0.017. A simple average falls short of providing us with a valuable insight because it places a higher weight on primary studies reporting a large amount of effect estimates. In fact, our dataset contains seven primary studies reporting only one estimate, whereas Sabia (2009) reports 220 estimates in his study. To counter this disproportion, we weight estimates by the inverse of the number of estimates reported by each study. The weighted mean equals -0.051.

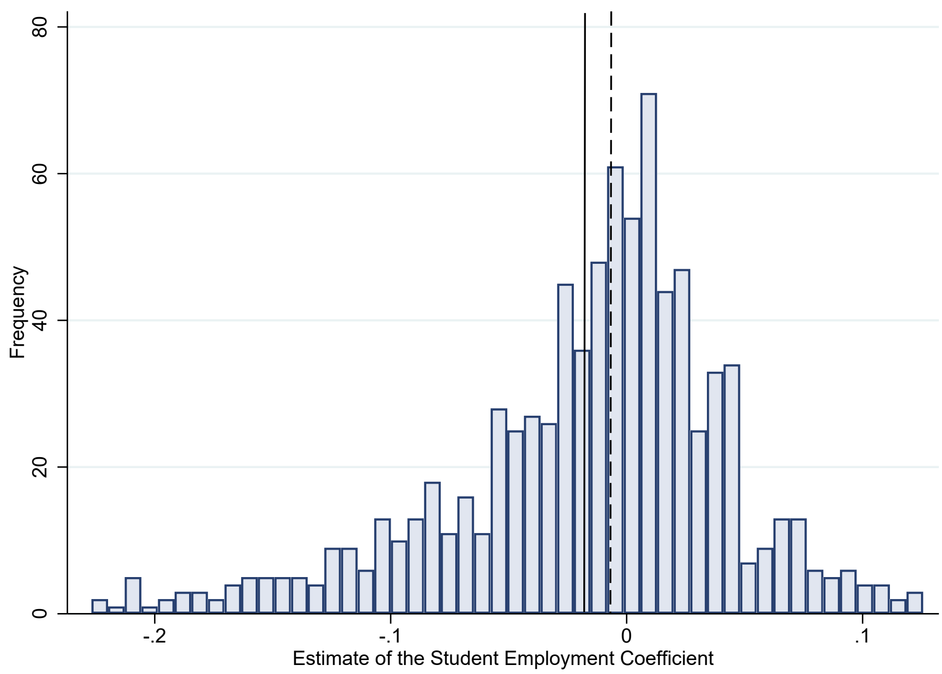
Both means show a modest magnitude of the relationship between employment and students' academic performance. Even the weighted mean suggests a weak and non-significant relation. In light of Doucouliagos' Effect Size Guidelines for Partial Correlations in Empirical Economics (2011), these values do not imply even a small effect. Furthermore, Figure 4.2 depicts the distribution of calculated PCCs. The histogram is pointy documenting a heavy-tailed, uni-

Figure 4.1: Variation of effect estimates within and across countries



Notes: The figure displays a box plot of the partial correlation coefficients capturing the relationship between student employment and academic achievement across different countries.

Figure 4.2: Distribution of calculated PCCs



Notes: The figure shows the frequency distribution of the calculated PCCs. The solid vertical line represents the mean. The dashed lines denotes the median.

modal distribution with a slight negative skewness. Additionally, a box plot of PCCs across studies available in Appendix A suggests a substantial within- and between-study heterogeneity in PCCs.

To obtain some preliminary clues on the heterogeneity, we examine simple and weighted mean values for different categories of our dataset, reported in Table 4.2. Although we provide simple and weighted mean values, the following discussion relies on the weighted mean values that are more reliable. Studies failing to control for the endogeneity of students' decision to work yield substantially higher negative weighted mean compared to studies accounting for endogeneity, which supports the view that the negative relationship is attributable to pre-employment differences (Rothstein 2007). Correspondingly, mean effect estimates for different estimation methods diminish in magnitude as their ability to control for pre-existing unobservable differences enhance. Nevertheless, apart from dynamic discrete estimation approach, the differences among estimation techniques are rather small. Next, we observe that the negative effect of student employment on educational outcomes is more pronounced for university and male students compared to high school and female students, respectively.

Further, Table 4.2 shows that estimates signify little spatial and time variation. On the other hand, we observe clear differences between estimates reported in published and unpublished studies implying a possible presence of publication bias in collected estimates. Finally, the mean effect estimates of different educational outcomes are relatively comparable as opposed to mean estimates of different employment measures. Primary studies using the continuous form of student employment yield substantially higher negative effect compared to studies using either dummy or categorical specification. We notice that for categorical specification, the mean is driven by positive estimates reported for low work intensities. This could bias our results in the forthcoming analyses. For this reason, in the next sections, we create more homogeneous subsamples that we use as robustness checks.

Table 4.2: Partial correlation coefficients for various data categories

		Unweighted			Weighted		
	n	Mean	95% CI		Mean	95% CI	
All estimates	861	-0.017	-0.022	-0.013	-0.051	-0.057	-0.045
Endogeneity							
Endogeneity control	307	-0.022	-0.028	-0.016	-0.036	-0.043	-0.029
No endogeneity control	554	-0.015	-0.021	-0.009	-0.057	-0.065	-0.049
Data type							
Longitudinal data	729	-0.008	-0.012	-0.004	-0.025	-0.030	-0.020
Cross-sectional data	132	-0.069	-0.086	-0.052	-0.087	-0.105	-0.069
Estimation method							
Elementary approach	525	-0.013	-0.021	-0.009	-0.057	-0.065	-0.049
Matching approach	29	-0.041	-0.057	-0.025	-0.060	-0.073	-0.048
Joint modelling	138	-0.041	-0.051	-0.030	-0.045	-0.055	-0.034
Panel methods	148	-0.007	-0.012	-0.002	-0.035	-0.043	-0.028
Dynamic discrete	21	-0.009	-0.053	0.035	-0.009	-0.048	0.030
Data characteristics							
Male students	218	-0.013	-0.020	-0.007	-0.056	-0.067	-0.044
Female students	222	0.002	-0.003	0.007	-0.023	-0.028	-0.017
Secondary education	621	-0.008	-0.012	-0.004	-0.029	-0.035	-0.024
Tertiary education	240	-0.043	-0.055	-0.031	-0.070	-0.082	-0.057
Academic outcome							
Educational choice	147	-0.030	-0.040	-0.020	-0.039	-0.048	-0.030
Educational attainment	116	-0.022	-0.036	-0.008	-0.039	-0.056	-0.022
Test scores	598	-0.014	-0.019	-0.001	-0.057	-0.065	-0.050
Student employment							
Continuous variable	261	-0.039	-0.048	-0.030	-0.080	-0.092	-0.069
Dummy variable	158	-0.007	-0.016	0.003	-0.024	-0.035	-0.014
Categorical variable	442	-0.008	-0.014	-0.003	-0.022	-0.029	-0.015
Low-intensity	185	0.013	0.006	0.021	0.014	0.004	0.024
Medium-intensity	94	-0.011	-0.023	0.000	-0.035	-0.051	-0.020
High-intensity	163	-0.031	-0.041	-0.021	-0.044	-0.054	-0.034
Spatial variation							
European countries	125	-0.029	-0.045	-0.014	-0.061	-0.079	-0.044
USA	694	-0.013	-0.018	-0.009	-0.042	-0.048	-0.036
Other countries	42	-0.052	-0.076	-0.029	-0.075	-0.101	-0.048
Publication status							
Unpublished	75	-0.004	-0.022	0.014	-0.021	-0.038	-0.004
Published	786	-0.019	-0.023	-0.014	-0.057	-0.063	-0.050
Publication date							
Until 1990	40	-0.061	-0.087	-0.034	-0.055	-0.089	-0.022
1991-2000	103	-0.030	-0.040	-0.020	-0.039	-0.053	-0.025
2001-2010	453	-0.010	-0.016	-0.004	-0.053	-0.061	-0.045
Since 2011	265	-0.019	-0.027	-0.011	-0.053	-0.063	-0.042

Notes: The table shows mean values of PCCs for different subsets of data. We report the unweighted and weighted mean where we weight the values by the inverse of number of estimates per study.

Chapter 5

Publication Bias

5.1 Definition and Motivation of Publication Bias

In the previous sections, we reviewed theories and methodological approaches concerning the relationship between student employment and educational outcomes. Further, we summarized the collected evidence in order to identify the average effect and the most pronounced empirical patterns. Although such approach is valuable as it uncovers new insights, it fails to challenge the credibility of collected empirical evidence resulting in possibly biased summary effects (Ioannidis *et al.* 2017).

Credibility of certain findings has been traditionally evaluated by journal editors and referees who ultimately decide which results get published. Despite all efforts to standardize this decision-making process, findings with certain features are more likely to be accepted by journal editors. In consequence, such approach might produce a biased overview of the empirical knowledge on a certain topic. Stanley (2005) identifies three features enhancing the publication probability of effect estimates. First, both researchers and editors favour statistically significant estimates. Second, reviewers and researchers treat more favourably research findings consistent with the dominant theory. Third, researchers themselves use the dominant theory, although unintentionally, as a guiding mechanism when designing their model specification. Generally, we call this tendency of journal reviewers and researchers to "prepare, submit and publish" research findings corresponding to their expectations of statistical significance and direction 'publication selection bias' or simply 'publication bias' (Dickersin 2005, pg. 13).

Publication bias has been excessively described in the past (Dickersin 2005).

Only recently, however, researchers have begun to study the systematic misreporting of effect estimates in various economic topics including minimum-wage research (Card & Krueger 1995; Doucouliagos & Stanley 2009), budgetary implications of numerical fiscal rules (Heinemann *et al.* 2018), or consumption habit formation (Havranek *et al.* 2017). Inevitably, reported results and their distortion have been also scrutinized in other research areas including economics of education. For example, Havranek *et al.* (2018) find a negative publication bias in the evidence on the relationship between tuition fees and demand for tertiary education. Similarly, Minasyan *et al.* (2019) test whether publication selection bias permeates literature concerning the link between gender inequality in education and per capita economic growth.

To enrich the research of systematic misreporting in the field of economics of education, in this section, we investigate whether the existing empirical evidence on educational implications of student employment suffers from publication selection bias. A less immersed observer could reject the presence of publication bias in this empirical evidence for two reasons. For one, as there are multiple theories offering convincing explanations for positive and negative estimates, research findings are less likely to be influenced by a 'commonly accepted theory'. For two, researchers themselves (See, for instance, Oettinger 1999; Sabia 2009; Tyler 2003) acknowledge that there seems to be little consensus on whether student employment hinders or improves academic performance implying minor preference for a certain type of results among researchers.

Nonetheless, a thorough look into the existing literature suggests counterarguments, indicating the presence of negative publication bias. For example, Buscha *et al.* (2012) admit that "the view that part-time work has a detrimental effect on educational attainment [...] is increasingly widespread in the last 10 years" (pg. 383). Similarly, Neyt *et al.* (2019) hint at the possibility of negative publication bias by showing that the most convincing studies in terms of their methodological advancement yield less negative effect estimates compared to less advanced studies. If the collected evidence on the effect between student employment and academic performance is subject to publication bias, the descriptive statistics are biased. Hence, the sample mean reported in the previous section does not represent the true average effect. To assess the credibility of empirical evidence regarding the effect of student employment on educational outcomes, in this section, we test whether the collected estimates are influenced by publication bias.

5.2 Testing for Publication Bias

5.2.1 Graphical Test: Funnel Plot

We begin our investigation of potential publication bias by employing a visual tool for its assessment called funnel plot. In essence, funnel plot represents a scatter plot, where the effect estimates from individual studies depicted on the horizontal axis are plotted against the precision measure on the vertical axis (Sterne *et al.* 2005).

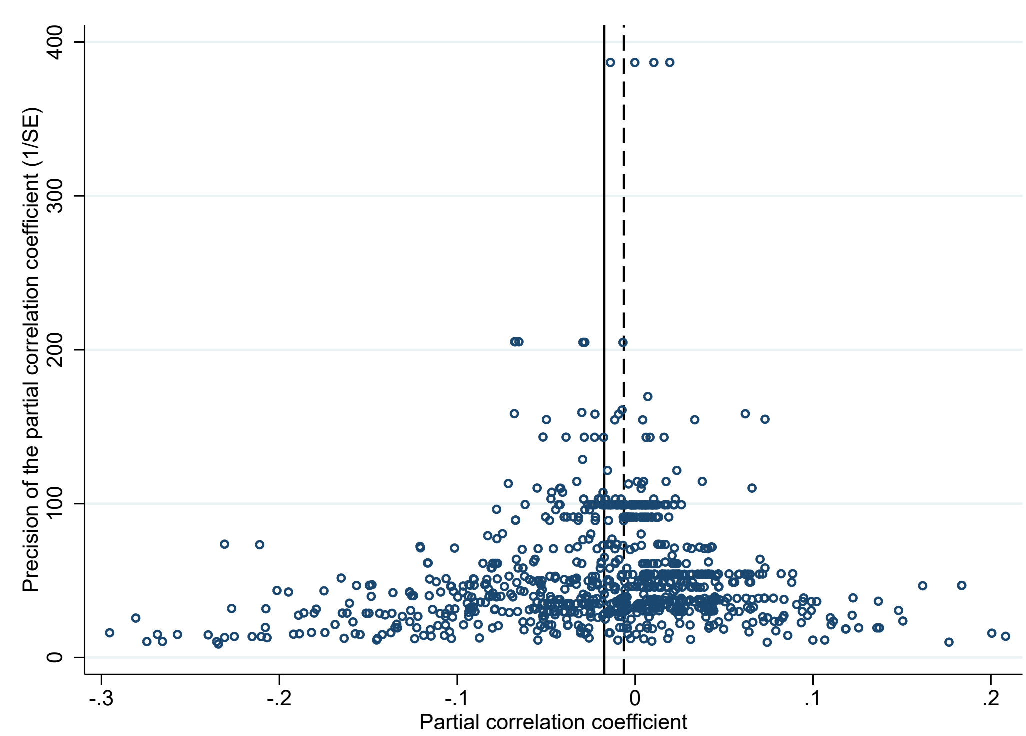
Assumingly, the most precise estimates should lie close to the true mean effect in the top part of the graph, with the spread widening out at the bottom of the graph as the precision of effect estimates decreases (Egger *et al.* 1997). Therefore, in the absence of publication bias, the graph should form the shape of a symmetrical inverted funnel. Conversely, an asymmetry of the funnel plot indicates the existence of publication bias, e.g. preference for positive or negative estimates. Moreover, funnel plot indicating an overrepresentation of statistically significant results over statistically non-significant results, irrespective of their direction, appears wide and hollow (Stanley 2005).

We construct the funnel plot presented in Figure 5.1 in the following way. On the horizontal axis, we plot the partial correlation coefficients. On the vertical axis, we plot the inverse of standard errors. Although using the inverse of standard errors as a measure of precision is a standard practice in a meta-analytic research, this choice slightly modifies the expected shape of funnel plot: in the absence of publication bias the outer lines of the resulting funnel plot should be curved inwards (Sterne *et al.* 2005).

The diagram presented in Figure 5.1 roughly forms the predicted inverted funnel shape with a high level of symmetry. Even very imprecise estimates concentrated at the bottom of the diagram are reported. After a detailed inspection, we notice that the graph depicts more negative estimates of a high magnitude compared to high positive estimates. This results in a left-skewed funnel plot indicating that large positive estimates are slightly underrepresented in the literature.

Nevertheless, this mild underrepresentation of large positive estimates does not have to be a result of a negative publication bias. Variation in estimation methods and any other difference between research designs of primary studies is capable of introducing asymmetry to the funnel plot (Stanley 2005). Unfortunately, the graphical means of funnel plot does not allow us to examine other causes of its skewness apart from publication selectivity. Moreover, the funnel plot technique is a highly subjective assessment of publication bias as it relies on researchers' visual evaluation. Hence, given little asymmetry in the funnel plot and the subjectivity of this tool, at this point of our analysis, we cannot conclude that a negative publication bias is present in the literature on the relationship between student employment and academic achievement. To complement the graphical analysis with a more rigorous assessment, in the next subsections, we utilize quantitative tests of publication bias.

Figure 5.1: Funnel plot of partial correlations coefficients



Notes: The figure displays the funnel plot of partial correlation coefficients capturing the effect of student employment on academic achievement. The solid vertical line represents the mean of PCCs, the dashed vertical line indicates the median. We use unwinsorized data for constructing this graph, however, for quantitative tests we use winsorized data.

5.2.2 Linear tests

Formally, we inspect the publication bias using the linear regression approach suggested by (Egger *et al.* 1997). Following the linear regression approach, we estimate:

$$PCC_{is} = \beta_0 + \beta_1 SE_{PCC_{is}} + \epsilon_{is}, \tag{5.1}$$

where PCC_{is} and $SE_{PCC_{is}}$ are the calculated partial correlation coefficients and their corresponding standards errors, respectively. Finally, ϵ_{is} represents the regression error term. We interpret the constant β_0 (effect beyond bias) as the true effect corrected for publication bias. On the other hand, beta coefficient β_1 (publication bias) and its test for statistical significance convey the information regarding the existence, direction and magnitude of the potential publication bias.

We call Equation 5.1 the funnel asymmetry test (FAT) as it expresses the symmetry of funnel plot in quantitative terms (Egger *et al.* 1997). In essence, the funnel asymmetry test explores the correlation between reported estimates and their standard errors. In the absence of publication bias, the test detects the correlation to be zero as the effect estimates are randomly distributed across studies (Stanley 2005). Conversely, if the test diagnoses publication bias, the correlation between reported estimates and their standard errors is non-zero.

To estimate the regression Equation 5.1, we pursue different model specifications. First, we estimate the equation with ordinary least squares (OLS). Second, we test for publication bias by estimating the model with between-study variance. We do not estimate the model using within-study variance because the nature of our dataset does not allow us to pursue this specification; our dataset is unbalanced and some primary studies report only one effect estimate. In the third specification, we introduce a weighting scheme widely used in meta-analytic studies (For instance in Havranek *et al.* 2018 or Gechert *et al.* 2019). In the third specification, we weight effect estimates by the inverse of the number of observations per study, ensuring that estimates generated from studies of varying sample sizes have an equal impact. Finally, Stanley (2005) points out that estimating Equation 5.1 might be biased due to random sampling error and joint determination of effect estimates and their standard errors by the chosen estimation method in the primary study. To remedy the potential endogeneity between reported estimates and their standard errors, we estimate Equation 5.1 with the instrumental variable approach. We use the square root of the number of observations as an instrument. This instrument should fulfill both the relevance and exogeneity conditions. By definition, the instrument is correlated with standard errors because large-sample studies typically generate small standard errors. On the other hand, it is unlikely, however not impossible, that sample size is correlated with a chosen estimation approach.

The results of these specifications are presented in Panel A of Table 5.1. We observe a mild negative publication bias that is statistically significant at 5% or lower significance level for all four specifications. On the contrary, based on the low and statistically non-significant values for mean corrected for publication bias, we cannot reject the hypothesis that the underlying effect of students' employment on their educational outcomes is non-existent.

Table 5.1: Linear tests of publication bias

Panel A: Various model specifications	OLS	Between effects	Study	Instrument
Standard Error	-0.881**	-1.974***	-1.779***	-0.904**
(Publication Bias)	(0.31)	(0.36)	(0.38)	(0.31)
	[-1.629, -0.273]		[-2.585, -0.874]	[-1.589, -0.292]
Constant	0.006	0.017	0.010	0.007
(Mean Beyond Bias)	(0.01)	(0.01)	(0.01)	(0.01)
	[-0.023, 0.035]		[-0.012, 0.031]	[-0.022, 0.036]
Observations	861	861	861	861
Studies	69	69	69	69
Panel B: Model specifications weighted by precision	**WLS**	**Between effects**	**Study**	**Instrument**
Standard Error	-0.544	-1.336***	-1.305**	-0.561
(Publication Bias)	(0.31)	(0.36)	(0.44)	(0.31)
	[-1.239, 0.095]		[-2.211, -0.2261]	[-1.224, 0.094]
Constant	-0.003	-0.003	-0.004	-0.003
(Mean Beyond Bias)	(0.01)	(0.01)	(0.01)	(0.01)
	[-0.020, 0.011]		[-0.032, 0.013]	[-0.017, 0.011]
Observations	861	861	861	861
Studies	69	69	69	69

Notes: The table reports the results of linear regression testing for the presence of publication bias. The standard errors of the regression parameters are clustered at the study level. The simple uncorrected mean equals -0.017, the weighted uncorrected mean - 0.051. In panel A we present the following specifications: OLS = ordinary least squares, BE = study-level between effects, Study = weighted by the inverse of the number of estimates reported per study, IV = the inverse of the square root of the number of observations acts as an instrument for the standard error. In Panel B, the same specifications are additionally weighted by the inverse of PCC's standard errors. Standard errors in parentheses. 95% confidence intervals from wild bootstrap clustering in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To further address the heteroscedasticity of Equation 5.1, we follow another standard practice among meta-analysts (See, for instance, Stanley 2005 or Zigraiova & Havranek 2016) and apply a weighting scheme, in which we assign more weight to more precise estimates. To achieve higher efficiency, we weight the equation by the inverse of their standard errors (1/SE):

$$PCC_{is}.\frac{1}{SE_{PCC_{is}}} = t_{is} = \beta_0.\frac{1}{SE_{PCC_{is}}} + \beta_1 SE_{PCC_{is}} + \nu_{is}. \tag{5.2}$$

In Equation 5.2, β_0 denotes the extent and direction of publication bias, while β_1 signifies the mean effect corrected for publication bias. Moreover, t_{is} stands for t-statistic of the i-th PCC obtained from the s-th study. We call Equation 5.2 precision asymmetry test (PET) (Stanley 2005). Apart from applying this heteroscedasticity correcting weighting scheme separately, we pursue it simultaneously with the other model specifications described above. Panel B in Table 5.1 reports results for these estimations. Likewise to Panel A, we observe that the underlying corrected mean effect is substantially smaller than the simple uncorrected mean effect (-0.017), close to zero and non-significant. Regarding the publication bias, we still observe a negative publication bias, however its significance at 5% level fades out for OLS weighted by precision and instrumental variable estimation weighted by precision.

Our final advancement accounting for the heteroscedasticity of FAT-PET entails clustering standard errors at the study level. The standard assumption that error term ϵ_{is} is independently and identically distributed as $\epsilon_{is} \sim iid$ $(0,\sigma^2)$ is most likely violated. It is likely that there is some correlation between standard errors of effect estimates reported in the same primary study. To avoid erroneous results and inferences from statistical tests, we cluster standard errors at the study level. By this clustering, we assume correlation between estimates from the same study and independence between estimates generated by different studies. Although we reach the minimum amount of clusters for reasonable inference, our sample suffers from unequal cluster sizes that might introduce another bias to the inference (MacKinnon & Webb 2017). To account for this cluster unbalancedness, we follow Gechert *et al.* (2019) and employ wild bootstrap clustering. We report the 95% confidence interval of the wild bootstrap for all specifications apart from study-level between-effect estimation.

5.2.3 Non-linear tests

Both funnel and precision asymmetry tests (FAT-PET) serve as powerful tests for examining publication bias and the true underlying effect corrected for this bias (Stanley 2008). However, the major limitation of these tests stems from their assumption that there is a linear relation between publication bias and standard errors. Stanley *et al.* (2010) suggest that this assumption is likely to be violated for highly precise effect estimates. Estimates concentrated at the top of the funnel plot are less likely to be contaminated by publication bias bias due to their sufficiently small standard errors. Such small standard errors allow them to achieve statistical significance without particular selection (Stanley *et al.* 2010). Therefore, when applying linear approximation, publication bias might be overstated, shifting the true underlying effect downwards. For this reason, we perform alternative non-linear tests used in recent meta-analyses to identify the authentic empirical effect clear of publication bias (See, for instance, Cazachevici *et al.* 2020; Havranek *et al.* 2020; Matousek *et al.* 2019). Table 5.2 displays values of the true underlying effect using these non-linear techniques.

Table 5.2: Non-linear tests of publication bias showing the underlying true effect

	Stem-based method	Endogenous kink	WAAP	Selection model
Effect Beyond Bias	0.004	-0.010	0.004	-0.019**
	(0.021)	(0.027)	(0.007)	(0.003)
Observations	861	861	861	861
Studies	69	69	69	69

Notes: The table reports the results of non-linear tests, showing the magnitude and significance of the true underlying effect corrected for publication bias. The simple uncorrected mean equals -0.017, the weighted uncorrected mean - 0.051. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

First, we employ the stem-based method introduced by Furukawa (2019). The stem-based method builds upon the 'Top 10' developed by Stanley *et al.* (2010). Both methods operate under the assumption that a subsample of the most precise estimates should exert a minimum publication bias. Hence, the calculated true underlying effect based on this subsample should be more efficient. While 'Top 10' method relies on an arbitrary cut-off point of the 10% most precise estimates, the Stem-based method determines the number of studies for estimating the true empirical effect endogenously. Furukawa (2019) obtains the desired number of studies by minimizing the mean squared error of the estimates. The goal of the estimation is to find an optimal number of studies that is sufficiently high to ensure efficiency of the true effect estimator and sufficiently low, so it does not include many imprecise estimates.

The next method we use, the Endogenous Kink method as proposed by Bom

& Rachinger (2019), also follows logic embedded in 'Top 10' approach. Assuming that more precise estimates are less likely to suffer from publication bias, the objective of Endogenous Kink technique is to isolate them and use them for computing the average effect. Similarly to Furukawa's Stem-based method, Endogenous Kink finds the fraction of most precise estimates endogenously. The Endogenous Kink method obtains the cut-off value by fitting a piecewise linear meta-regression of estimates on their standard errors. This regression consists of two branches; a horizontal branch for the most precise estimates featuring no relation with their standard errors a positively-sloped branch mirroring the positive correlation between standard errors and estimates contaminated by publication bias. The kink, at which the branches meet, signifies the cut-off value. Using stem-based method and the Endogenous Kink technique yield the mean of true underlying effect estimate of 0.004 and -0.010, respectively.

Further, we follow a novel method designed by Ioannidis *et al.* (2017) and compute the weighted average of adequately powered (WAAP) estimates. The WAAP estimator calculates the true effect using only adequately powered estimates with statistical power higher than 80% and weights them by optimal weights ($1/SE^2$). Stanley *et al.* (2017) point out that the main limitation of the WAAP method lies in its unfeasibility when being applied to datasets with no sufficiently powered studies. Our winsorized sample is an example of such a dataset. Therefore, when calculating WAAP, we resort to using unwinsorized data. Estimating WAAP for the unwinsorized dataset, we obtain the mean effect of 0.004.

Our final non-linear test is the Selection Model introduced by Andrews & Kasy (2019). Similarly to Hedges (1992), Andrews & Kasy (2019) assume that the chance of publishing certain effect estimate is dependent on its statistical significance and that this chance changes once a certain level of t-statistic is achieved. The method uses maximum likelihood to identify the publication probability for different windows of data bounded by critical t-statistic thresholds. Consequently, it calculates how many estimates in these windows are underrepresented and assigns them with more weight. The Selection Model gives us the mean effect of student employment on educational outcomes equal to -0.019.

On average, the values of true underlying effect generated by non-linear tests are slightly more negative than the ones generated by linear tests. This confirms that linear tests sometimes overstate publication bias and understate the true underlying effect. Nonetheless, given the zero-close estimates, non-linear tests are consistent with linear tests, showing that the true underlying effect is negligible and mostly statistically non-significant. This finding is reconcilable with modern theoretical mechanisms. These mechanisms state that the effect of student work on academic performance is not driven by the employment itself. However, it is rather conditional on students' primary orientation or pre-existing characteristics, resulting in attenuation or absence of the effect.

5.2.4 Caliper test

Non-linear tests allow us to verify the magnitude and existence of the true underlying effect, while relaxing the linearity assumption between effect estimates and standard errors embedded in funnel and precision asymmetry tests (FAT-PET). Nevertheless, non-linear tests do not circumvent another limitation of FAT-PAT, which is the exogeneity assumption between standard errors and effect estimates. Traditional solution to this issue involves estimating FAT-PAT with instrumental variable that we perform in Subsection 5.2.2. To further corroborate or reject the negative publication bias, we seek an alternative test, relaxing the exogeneity assumption.

With this objective in mind, we employ Caliper test proposed by Gerber & Malhotra (2008) and used in modern meta-analyses (See, for instance, Havranek *et al.* 2020; Matousek *et al.* 2019). Similarly to instrumental variable estimation of FAT-PET, Caliper test drops the assumption of zero correlation between effect estimates and their standard errors in the absence of publication bias. It compares number of estimates below and above certain critical value of t-statistics in a given band. In the absence of publication bias, estimates just above and just below these arbitrary thresholds (in our case -1.96, 0, and 1.96) should be equally likely which seems to be violated in our case by looking at Figure 5.2 showing the frequency distribution of t-statistics of computed PCCs. If the opposite is true, these critical values have a substantial impact on which findings get published, supporting the presence of publication selection bias.

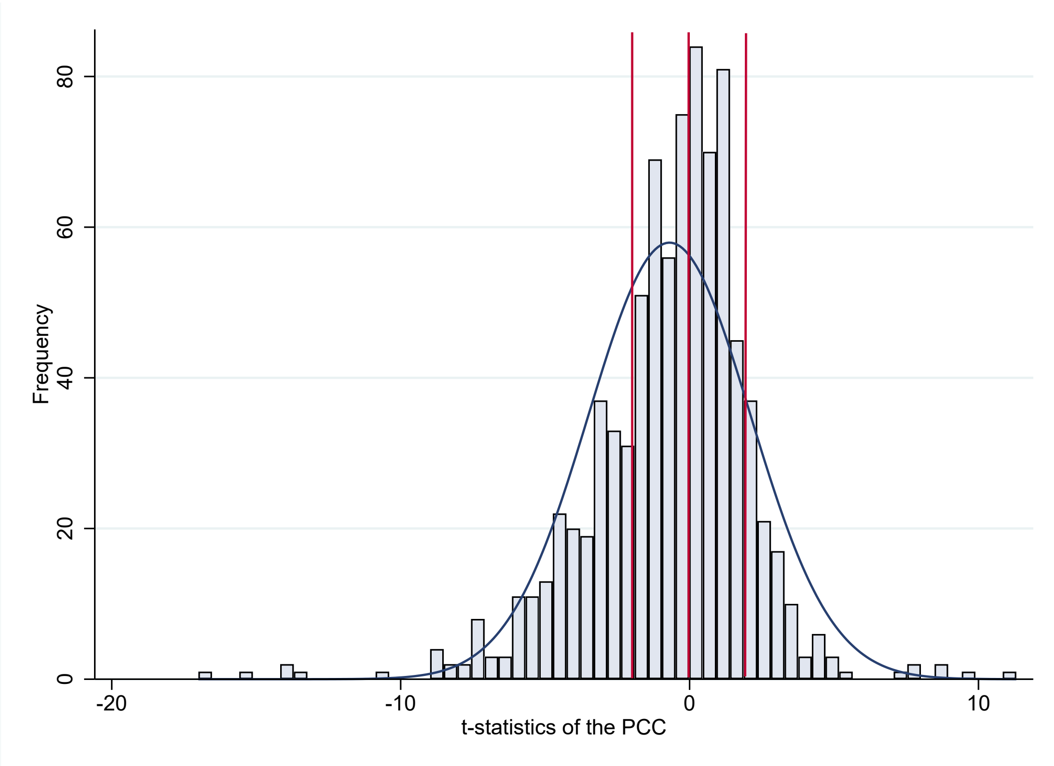
Our results for Caliper test reported in Table 5.3 do not show any publication selection bias between significant and non-significant positive estimates where the frequencies of estimates below and above the 5% significance threshold gradually equalize as we narrow down the caliper. Similarly, Caliper test detects minimum disparities between positive and negative estimates that are equally distributed around the threshold of 0. Although, given the results of

Table 5.3: Caliper test for detecting publication bias

Threshold for t-statistic	-1.96	0	1.96
Caliper size: 10%			
Share of estimates above the threshold	0.472***	0.548***	0.577***
	(0.08)	(0.09)	(0.10)
Caliper size: 5%			
Share of estimates above the threshold	0.750***	0.545***	0.545***
	(0.11)	(0.16)	(0.16)
Observations	861	861	861
Studies	69	69	69

Notes: The table reports the share of estimates being above the critical value of t-statistic (in absolute terms) in a 10% and 5% caliper. To illustrate the interpretation of the coefficients, a coefficient of 0.750 means that the ratio of negative significant estimates to non-significant is 75% to 25%. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5.2: Distribution of t-statistics of partial correlation coefficients



Notes: The figure displays the distribution of t-statistics of partial correlation coefficients. Vertical lines denote the critical thresholds of -1.96, 0, and 1.96. We use unwinsorized data to construct this graph.

the previous tests, we would expect negative estimates to dominate positive ones, the equal distribution of estimates in the caliper around 0 is sensible as the average effect fluctuates around 0. For the last critical value of t-statistic equal to -1.96, we notice that negative significant estimates dominate the negative non-significant estimates 75% to 25% for the tightest band of 5%. This result refines our analysis; it shows that researchers not only prefer negative estimates to positive ones as corroborated by linear tests, but also prefer significant negative estimates to negative non-significant findings.

5.3 Robustness checks

To further validate our results, we conduct the above-described tests on three subsamples of our original dataset. For better flow of this thesis, we present results for these robustness checks in Appendix B.

The first subsample discards estimates from studies that fail to control for the endogeneity of students' decision to participate in the labour market. Demonstrated by descriptive statistics, the weighted mean for estimates ignoring the endogeneity bias is almost twice as negative as the mean for estimates accounting for endogeneity. Moreover, as pointed out by Neyt *et al.* (2019), the most advanced studies accounting for endogeneity yield less negative estimates compared to less advanced studies. Hence, it is reasonable to assume that authors of these methodologically advanced studies do not hold any specific preferences regarding the effect because they realize that the negative effects of previous studies are driven by insufficient controlling for endogeneity. To test this presumption, we test for the presence of publication bias on a subset of estimates generated by studies controlling for endogeneity. This subsample contains 307 observations with a simple mean of -0.022 and median of -0.008.

Linear tests of publication bias for the subsample of estimates controlling for endogeneity do not provide convincing evidence for the presence of negative publication bias as the coefficients capturing the bias are mostly statistically non-significant. Moreover, although Caliper test corroborates the existence of negative publication bias for the wider caliper, the trend reverses for the tighter caliper where we observe that positive estimates dominate negative estimates 67% to 33%. Nevertheless, the result for the tighter caliper is statistically non-significant. Ultimately, these findings show that studies carefully controlling for endogeneity of student employment are not plagued by negative publication bias. This finding implies that the publication bias present in the overall sample might not be driven by a shared preference for negative effect estimates, rather by a common failure to carefully control for students' individual characteristics mutually influencing their educational outcomes and decision to work.

In the second subsample, we include only PCCs of studies that operationalize the independent variable of student employment as a continuous variable. This more homogenous dataset consists of 261 observations, yielding the unweighted mean of -0.039 and median of -0.024. Funnel and precision asymmetry tests performed on this subsample yield results consistent with our previous findings; they demonstrate a weak negative publication bias significant at least at 10% level for most of our specifications. Similarly to our analysis on the whole sample, results of Caliper test for this subsample show that negative significant estimates dominate negative non-significant estimates 71% to 29%.

Consequently, as our final robustness check, we conduct the analysis of publication bias on a subsample of original estimates, not transformed into PCCs. As none of the primary studies reports the effect of student work in the form of elasticities, we specify our own criteria for a comparable economic effect. Observations included in this non-transformed subsample measure student employment as the average amount of hours worked per week during the academic year. Moreover, they use a 4-point grade average as the educational outcome. Admittedly, accommodating only estimates using GPA as a response variable disqualifies research papers conducted in other countries than the US. Nonetheless, including other grading scheme would make the economic effect again incomparable. The resulting non-transformed dataset consisting of 92 estimates from 16 studies can be characterized by a simple unweighted mean of -0.007 and a median of -0.004. Linear tests performed on this untransformed dataset indicate the presence of negative publication bias, yet its magnitude and statistical significance varies. Unfortunately, we do not perform Caliper test for this subsample due to the insufficient amount of observations.

Finally, for all three subsamples, both linear and non-linear analyses fail to reject the null hypothesis that the true underlying effect is different from zero. Even in the untransformed subsample the mean beyond bias is close to zero, indicating that the economic effect is very small. Hence, as this finding permeates all our specifications and robustness checks, we conclude there the effect of students' employment on their academic achievement is negligible.

Chapter 6

Heterogeneity

As noted earlier, the reported estimates of the effect of student employment on educational outcomes differ in terms of their magnitude and direction. The previous section accounted a certain portion of heterogeneity among reported estimates to the negative publication bias. In this section, we explore other sources of heterogeneity associated with the characteristics of the study design. We feel encouraged by previous research that articulated the need for investigating heterogeneity. For instance, Tyler (2003) concludes his review of the past research on educational implications of student employment by this suggestion: "Taken as a whole these studies do not offer consistent lessons about the relationship between school-year work and academic achievement. The reasons for the inconsistencies are likely related to some combination of different data sets, different age students, different dependent variables, and different empirical methods across the studies" (pg. 386).

Hence, in this chapter, we identify design study characteristics that explain the variation among the partial correlation coefficients. First, we describe the variables we collected to capture the systematic differences. Second, we explain the Bayesian Model Averaging (BMA) technique that we use for investigating heterogeneity and finally, we present our results.

6.1 Explanatory variables

To investigate reasons causing heterogeneity between the reported estimates, we introduce 31 additional variables capturing the differences. We group the variables in 6 blocks: data characteristics, design of the analysis, estimation methods, variable specification, publication characteristics, and student characteristics.

Admittedly, the set of included additional variables is not exhaustive as there is a myriad of factors influencing the relationship under consideration. For instance, DeSimone (2006) acknowledges the potential link between population size and grade responsiveness to employment. However, given his solitude in examining this link, we defer from using population size as another explanatory variable. Similarly, Body *et al.* (2014) and McNeal (1997) investigate how the scrutinized relationship varies depending on the employment sector and student's field of study. For instance, Body *et al.* (2014) show that working in a private sector has a greater detrimental effect on students' success compared to working in a public sector due to time flexibility. Likewise, Richardson *et al.* (2013) argue that unlike to other students, fine-art students face a less-flexible study environment that might hinder their experience of combining work and studies. Although these student job dimensions influence the estimated relationship, they do not represent a common researchers' choice. Therefore, we do not include them in our analysis. On similar grounds, we exclude other variables such as whether students live at home or on campus (Body *et al.* 2014), association with religious communities (McVicar & McKee 2002), region of student's residency (McVicar & McKee 2002), etc.

Block 1 – Data characteristics: Besides the obvious data characteristics including the effect estimate and the corresponding standard error, we collect three additional variables describing the nature of the datasets employed in the primary studies. First, for each study we collect the number of observations used for the estimation. The size of data samples utilized in primary studies differs considerably. For instance, DeSimone (2006) relies on a nation-wide dataset with more than 149,000 individuals. In contrast, Hwang (2013) uses a sample including only 215 students following the course on Principles of Economics. Darolia (2014) argues that restricted datasets are problematic as they are more likely to miss or mask the true effect of student employment on educational outcomes. Ideally, datasets should reflect experiences of nationally representative sample of students (Darolia 2014; Callender 2008). Hence, to account for these differences, we create the variable *Sample size*.

Second, we create the variable *Average data year*, denoting the average year of the employed data in a standardized form. We assume that estimates capturing the effect of student employment on educational outcomes differ between student generations due to the varying working and studying habits. For instance, Babcock & Marks (2011) show that university students substantially decreased time devoted to studying between 1961 and 2003. On the contrary, students' work engagement and working hours have been steadily rising since 1980s (Singh 1998; McVicar & McKee 2002).

Third, instead of using cross-sectional data, some researchers exploit longitudinal datasets. Apart from establishing causality, longitudinal data allow researchers to control for the time-invariant individual unobserved heterogeneity, and thus account for the endogeneity of student employment (Oettinger 1999). Our analysis accounts for this specific type of dataset by introducing a dummy variable *Longitudinal data*.

Block 2 – Estimation methods: We code five methods used for the estimation of the investigated relationship: *Elementary approach* encompassing OLS and linear probability models, *Matching approach* representing the propensity score matching approach, *Panel methods* including fixed-effects and random-effects method, *Joint modelling* including instrumental variable approach and simultaneous modelling approach, and finally *Dynamic discrete* approach. Considering the varying underlying assumptions of these techniques and the degree to which these estimation methods account for students' unobservable differences, we expect estimation approaches to affect the reported estimates. Indeed, using the same dataset, Stinebrickner & Stinebrickner (2003) employ OLS, FE and IV approach to estimate the relationship between student work and academic performance and obtain three fundamentally different estimates.

Block 3 – Design of the analysis: The estimation techniques vary in their ability to control for potential endogeneity of students' decision to work. To address the degree to which different studies correct for the endogeneity bias, we include a dummy variable *Endogeneity*.

Further, we focus on the geographical variation among primary studies. While most primary studies utilize datasets obtained in the US, some studies use datasets including observations on students from Europe and other parts of the world. In particular, our sample contains 64% of primary studies conducted in the US, 25% of studies conducted in Europe, and the remaining share of studies use datasets concerning Australian, Canadian or Russian students. To account for varying educational systems and student attitudes across the world, we distinguish studies based on the country of analysis and create corresponding dummy variables *USA*, *Europe*, and *Other countries*.

Another feature of the primary studies we control for is the *Educational* *level*, at which the effect of student employment on academic performance is measured. The existing literature suggests opposing views on how the effect of student employment on educational outcomes differs between secondary and higher education students. On one hand, we would expect the effect to be less negative for university students. As noted earlier, Bozick (2007) argues that university students compared to high school students, have more flexible study environment and more favourable attitude to education as they decided to continue with their studies. On the other hand, our descriptive statistics presented in Table 4.2 shows more negative effect for tertiary students. Neyt *et al.* (2019) also report more negative effect for university students and explain that tertiary education students might be less successful in combining work and studying due to the more challenging content and less structured setting of their studies.

Finally, we control for the *Number of explanatory variables* included in the original estimations. Researchers usually attempt to employ a high number of variables in order to control for individual differences among students. In our sample, the average number of covariates included in original estimation is 17.

Block 4 – Variable specification: Researchers use various specifications to operationalize the student employment status. As discussed earlier, most of them utilize student employment as a continuous variable, while others create a categorical or a dummy variable. To control for the variation arising from different operationalization of student employment, we code the specification of *Employment* as *Continuous, Categorical* or *Dummy.* Equivalently, researchers examine the effect of student employment on various educational outcomes including educational choices, test/exam scores, and educational attainment (Neyt *et al.* 2019). We control for these differences by creating corresponding dummy variables *Educational choice*, *Educational attainment*, and *Test scores.*

Block 5 – Publication characteristics: Following the traditional practice in meta-analyses (See, among others, Havranek *et al.* 2016; Gechert *et al.* 2019), we further include several publication characteristics. First, to account for potential methodological innovations and time trend in the existing literature, we include the *Publication year* of the study as an explanatory variable. Second, to control for the quality of primary studies included in our sample, for each study we add two variables; the number of Google Scholar *Citations* and a dummy variable reflecting whether a study was *Published* in a reviewed journal. We standardize the number of citations per year in order to avoid penalizing the most recent studies.

Moreover, we also collect Journal Citations Report (JCR) impact factor for estimates reported in published studies. The annual *JCR impact factor* measures the citation frequency of a journal as the ratio of citations divided by the available citable items (Clarivate Analytics 2019). Admittedly, the JCR impact factor is a simple impact factor, failing to attribute each citation with an appropriate weight as opposed to recursive impact factor RePEc that is commonly used for assessing economics journals (Zimmermann 2013). Despite the improved methodology of recursive impact factors, we use the simple JCR impact factor as a journal quality measure since most of the primary studies in our sample have been published in journals concerning educational policy. As the RePEc list encompasses primarily journals on economic research, outlets wherein primary studies from our sample are published are scarcely included in the list.

Block 6 – Student characteristics: Another source of heterogeneity entails the distinct students' characteristics and preferences. As noted before, researchers aim to control for various student characteristics in their original regressions with the objective to account for individual characteristics that could potentially impact academic performance or employment decision.

One of such characteristics is students' intrinsic motivation that seems to be an important factor in determining the employment-education relationship. Empirically, Richardson *et al.* (2013) demonstrate that employment is less likely to hamper academic performance if students work because they want to than because they have to. Another sensible factor researchers control for is students' cognitive ability (Arano & Parker 2008; McNeal 1997; Staff & Mortimer 2007). For example, Oettinger (1999) finds that more able students systematically select different employment schedules than less able students. By incorporating dummy variables *Motivation* and *Ability* we reflect whether original estimation uses these student characteristics as control variables.

Similarly, researchers recognize the impact family background can exert on the association between student employment and educational outcomes by including family background variables into their estimation. Naturally, students' educational outcomes are influenced by the economic situation of their parents. Although, it would be sensible to distinguish between studies considering parental economic situation and those ignoring this aspect, we take a step further and focus on parental educational capital. Carneiro & Heckman (2003) suggest that student educational choices are better explained by family permanent features, such as parents' education levels which directly contribute to family permanent income. Apart from that, students growing up in families with higher education levels are likely to perform better academically as education is more valued in such families (Arano & Parker 2008). To address these mechanisms, we add a dummy variable *Parental education*, indicating whether the original estimation includes this control variable in the original model.

In addition, we include dummy variables for studies controlling for standard demographic characteristics such as students' *Ethnicity* and *Age*. Empirically, these factors have been shown to have a substantial impact on the link between student work and academic performance. For instance, Oettinger (1999) finds a negative effect of student employment on their GPA only for students from ethnic minorities. Correspondingly, Kohen *et al.* (1978) argues that the negative association is less pronounced for older students who tend to be more mature and committed to their educational and occupational goals.

Finally, we also aim to explore whether students' gender can drive heterogeneity in the effect of student employment on educational outcomes. Prior research provides some indication for this proposition. For instance, Montmarquette *et al.* (2007) find a negative association between student employment and educational outcomes only for males. Likewise, Sabia (2009) and Holford (2020) report more negative effect estimates for male students. Unlike to previous student characteristics, we control for the impact of gender directly, by incorporating two dummy variables, reflecting whether a certain estimate applies to *Male* students or to *Female* students.

Finally, we also collect four additional variables indicating whether original estimations control for factors such as marital status (Beffy *et al.* 2013), taking care of a child (Arano & Parker 2008), class attendance (Auers *et al.* 2007), or study time (Holford 2020). However, after inspecting means of these variables, we decide to exclude them from the analysis, given the minimum variation in their means. Similarly, we also drop variables [1] attributable only to university students as these variables have many missing observations and do not provide any additional information about high school students. Table 6.1 reports all collected additional variables, their definition and simple means.

[1]These variables reflect whether the estimate applies to undergraduate students, part-time students, and students studying two-year college and whether the original estimation controls for variables such as working on- or off-campus and receipt of subsidy.

Table 6.1: Description and summary statistics of additional variables

Variable	Description	Mean	SD
PCC	The partial correlation coefficient capturing the effect of student employment on educational outcomes.	-0.017	0.066
Standard error	The estimated standard error of the PCC.	0.027	0.017
Data characteristics			
Sample size	*The logarithm of the number of observations used in the primary study.*	7.619	1.211
Average data year	The logarithm of the mean year of the data used minus the earliest average year in our data (base = 1967).	3.243	0.530
Longitudinal data	= 1 if longitudinal data are used to estimate the effect (0 if cross-sectional survey data).	0.847	0.360
Estimation methods			
Elementary approach	= 1 if elementary approaches (OLS, logit regression, etc.) are used for estimation.	0.610	0.488
Matching approach	= 1 if propensity score matching approach is used for estimation.	0.034	0.181
Panel method	= 1 if panel methods (fixed-effects, random-effects) are used for estimation.	0.172	0.378
Joint modelling	= 1 if instrumental variable approach or simultaneous equation modelling is used for estimation.	0.160	0.367
Dynamic discrete (ref. category)	= 1 if dynamic discrete approach is used for estimation.	0.024	0.154
Design of the analysis			
Endogeneity	*= 1 if the estimation method accounts for potential endogeneity.*	0.357	0.479
Europe	= 1 if country of analysis is in Europe.	0.145	0.352
USA	= 1 if country of analysis is the USA.	0.806	0.396
Other countries (ref. category)	= 1 if country of analysis is different from the USA and outside of Europe.	0.049	0.216
Educational level	= 1 if educational outcomes are measured on secondary education level (0 if applicable to tertiary education level).	0.721	0.449
No. of variables	The logarithm of the number of explanatory variables used in the model in the primary study.	2.566	0.856
Variable specification			
Educational Choice	= 1 if educational outcome is specified as educational choice.	0.171	0.376
Educational Attainment	= 1 if educational outcome is specified as educational attainment.	0.135	0.342
Test scores (ref. category)	= 1 if educational outcome is specified as test and exam results.	0.695	0.461

Table 6.1: Description and summary statistics of additional variables (continued)

Variable	Description	Mean	SD
Employment: Continuous	= 1 if student employment is a continuous variable.	0.303	0.460
Employment: Dummy	= 1 if student employment is a dummy variable.	0.184	0.387
Employment: Categorical (ref. category)	= 1 if student employment is a categorical variable.	0.513	0.500
Publication characteristics			
Publication date	The logarithm of the year in which the study was published minus the year (1978), in which the first study in our sample was published.	3.261	0.917
Citations	The logarithm of the mean number of Google Scholar citations received per year since the study was published (collected in August 2020).	1.695	1.006
Published	= 1 if study was published in a scientific journal.	0.913	0.282
Impact	The JCR impact factor trend of the outlet in which the primary study had been published (collected in August 2020).	1.620	1.269
Student characteristics			
Ability	= 1 if original estimation accounts for students' ability, e.g. SAT scores, prior education, class rank, etc.	0.366	0.482
Motivation	= 1 if original estimation controls for students' academic motivation	0.338	0.473
Parental education	= 1 if original estimation includes variable(s) reflecting parents' educational level	0.545	0.498
Age	= 1 if original estimation controls for students' age	0.462	0.499
Ethnicity	= 1 if original estimation includes control variables reflecting respondents' affiliation to ethnic minority	0.596	0.491
Male	= 1 if the estimates apply only to male students.	0.253	0.435
Female	= 1 if estimates apply only to female students only.	0.258	0.438

Notes: The table reports summary statistics for additional explanatory variables. Variables not included in the final estimation due to high collinearity are in italics. Dummy variables representing omitted category are denoted as 'ref. category'. Mean = simple unweighted mean. SD = Standard Deviation.

6.2 Estimation method

6.2.1 Describing BMA

In total, we collect 31 study characteristics reflecting various aspects of study design. Our goal is to establish which study characteristics systematically influence the coefficients capturing the effect between student employment and their educational outcomes. An intuitive approach to this exercise would entail regressing the computed partial correlation coefficients on the whole set of additional explanatory variables. However, proceeding with this full regression model would yield imprecise results due to the inflated standard errors and the fact that the true specification of the empirical model is unknown, given the high number of regressors we want to include. To account for the model uncertainty in this exercise, we employ Bayesian Model Averaging (BMA) estimation technique to analyse heterogeneity (Moral-Benito 2012; Havranek *et al.* 2015).

Instead of arbitrarily choosing one model or presenting results from all plausible models, BMA solves the model uncertainty by considering all possible models with different choices of covariates (Raftery 1995). In essence, BMA estimates a large amount of regressions using different subsets of explanatory variables. Consequently, it constructs a weighted average of all these combinations (Zeugner & Feldkircher 2009). In this way, BMA provides superior predictive ability compared to other estimation strategies such as sequential regression (Raftery 1995).

In BMA, the K number of regression models M are weighted by Posterior Model Probabilities (PMP) arising from Bayes theorem. Considering data D, the posterior model probability of model M_k where k = 1, 2,..., K is given by:

$$pr(M_k|D) = \frac{pr(D|M_k)pr(M_k)}{pr(D)} = \frac{pr(D|M_k)pr(M_k)}{\sum_{l=1}^{K} pr(D|M_l)pr(M_l)}. \tag{6.1}$$

Based on Equation 6.1, we see that the posterior model probability $pr(M_k|D)$ is proportional to the product of integrated likelihood of the model $pr(D|M_k)$ capturing the probability of utilized data considering model M_k and the prior model probability pr(M_k). This product is then divided by the sum of integrated likelihoods of K regression models. While the posterior model probability indicates goodness-of-fit of model M_k, the prior model probability *pr(*M_k*)* refers to researcher's prior beliefs regarding the probability of model M_k before considering the data (Zeugner 2011).

Consequently, BMA uses the computed posterior model probabilities to calculate the weighted posterior mean and the weighted posterior variance (or weighted posterior standard deviation) for each included explanatory variable. We interpret these two statistics as the estimate of regression coefficient and the standard error of the estimated regression parameter. The weighted posterior mean and the weighted posterior variance are defined as:

$$E(\beta_i|D) = \sum_{k=1}^{K} \widehat{\beta}_{ik} pr(M_k|D), \tag{6.2}$$

$$Var(\beta_i|D) = \sum_{k=1}^{K} (Var(\beta_i|D, M_k) + \widehat{\beta}_{ik}^2) pr(M_k|D) - E(\beta_i|D)^2, \tag{6.3}$$

where $\widehat{\beta}_{ik}$ represents the estimated regression coefficient for i-th variable in k-th model. In Equation 6.2, we see that the posterior mean is a weighted average of β_i from all models, including models where variable i is not considered (Zeugner 2011).

Finally, BMA also generates Posterior Inclusion Probability (PIP). In contrast to computation of posterior mean, the posterior inclusion probability is defined as the sum of the posterior model probabilities only of models which include variable i. Thus, we define PIP as:

$$PIP = \sum_{k=1}^{K} pr(M_k|\beta_i \neq 0, D). \tag{6.4}$$

We interpret PIP as the probability that a given variable is a significant predictor of the dependent variable. To illustrate this better, PIP equal to 1 indicates that all effective models utilize the variable corresponding to this PIP. When interpreting the magnitude of posterior inclusion probability, researchers usually follow Jeffreys (1961). Jeffreys (1961) distinguishes between weak, positive, strong, and decisive effect if the value of the corresponding PIP falls into the interval of 0.5-0.75, 0.75.-0.95, 0.95-0.99, and 0.99-1, respectively.

When applying BMA, researchers face two computational problems, making BMA hard to implement. First, computing integrals included in the integrated likelihood function $pr(D|M_k)$ is demanding (Hoeting *et al.* 1999). Second, the enormous model space makes the estimation infeasible for a personal computer (Hoeting *et al.* 1999). For instance, with 31 additional variables, there would be 2^{31} possible regressions, representing a serious computational challenge.

One way to overcome this computational hardship is to apply the Markov chain Monte Carlo method with the use of Metropolis-Hastings algorithm. This approach became a standard practice in meta-analytic studies (See, for instance, Gechert *et al.* 2019; Havranek *et al.* 2015). Markov chain Monte Carlo method diminishes the computational demands of BMA by estimating only models with the highest PIP. The Metropolis-Hastings algorithm determines these models by comparing a benchmark model M_i with a competing model M_j in terms of their posterior inclusion probabilities (Zeugner 2011). If M_i is accepted in favour of M_j, a new competing model is selected and compared. If the opposite occurs and M_j is accepted, M_j becomes a new benchmark model and the procedure repeats (Zeugner 2011).

Furthermore, besides solving computational difficulties, BMA requires the researchers to specify the distribution priors over the parameter space denoted by g and over the model space denoted by $pr(M_k)$. Eicher *et al.* (2011) encourage researchers to use priors, reflecting all available information on models and parameters. If researchers possess only small amount of prior information, it is reasonable to turn to non-informative or so called default priors. Popular choices of such default g-priors represent the unit information prior ($g = N$, where N denotes number of observations) or 'BRIC' benchmark prior ($g = max(N,X^2)$, where X denotes number of regressors) (Zeugner & Feldkircher 2009). The higher the value of parameter prior, the more weight is placed on data compared to prior beliefs (Hasan *et al.* 2018). When specifying prior distribution over model space, researchers usually opt for uniform model prior (UMP), attributing equal prior probability to every model or a less restrictive beta-binomial model prior (Hasan *et al.* 2018). For a detailed discussion on prior choice, consult Eicher *et al.* (2011) for a theoretical discussion and Hasan *et al.* (2018) for a practical application.

6.2.2 Implementing BMA

After introducing BMA method into our data analysis tool kit, we apply it practically and estimate the following meta-regression equation using BMA:

$$PCC_{is} = \beta_0 + \beta_1 X_i + \beta_2 SE_{PCC_{is}} + \epsilon_{is}. \tag{6.5}$$

In Equation 6.5, PCC_{is} represents the partial correlation coefficient, X_{is} stands for additional explanatory variables introduced in Section 6.1, β_2 measures the direction and magnitude of publication bias, and ϵ_{is} denotes the error term. The constant β_0 has no interpretation as it reflects the mean effect corrected for publication bias conditional on covariates X.

Although we wish to include all explanatory variables reported in Table 6.1, this turns out to be undesirable for the following reasons. First, if we incorporated all dummy variables that we coded for different categories of certain study characteristic, we would end up with a perfect multicollinearity, yielding infinite variance inflation factors. To avoid the dummy variable trap, we include *n-1* variables for each characteristic with n categories. Variables excluded on this basis are denoted as a 'reference category' in Table 6.1. Second, as our sample includes also unpublished studies, variable *JCR impact factor* has many missing values. BMA automatically excludes observations with missing information. To prevent the exclusion, we code missing values for *JCR impact factor* as 0. Third, we inspect the correlation matrix for the remaining explanatory variables to further investigate the multicollinearity. We notice three problematic variables in the correlation matrix reported in Appendix C, namely *Sample size*, *Endogeneity*, and *USA*. The correlation between sample size and estimated standard error amounts to -0.92. This value is not surprising as we used number of observations, more specifically degrees of freedom (number of observations minus number of explanatory variables in the original regression) to calculate the standard error corresponding to PCC. Similarly, in case of *Endogeneity*, relatively high correlation values with various estimation methods, especially with *Elementary approach* (-0.93), are sensible as estimation techniques are characterized by their ability to control for endogeneity. To prevent imprecise results, we drop variables *Endogeneity* and *Sample size* (denoted by italics font in Table 6.1). Lastly, for *USA* and *Europe* we observe a correlation value of -0.84, demonstrating small geographical variability in the collected estimates. Although this correlation poses a threat to the precision of our results, we maintain both of these variables and aim to control for this correlation by calibrating BMA, e.g. choosing an optimal prior on the model space. By and large, we run BMA with 26 explanatory variables including the estimated standard errors.

Before we proceed with the application of BMA, we specify prior distributions on individual regression parameters and model probabilities. Given that the amount of prior information on the parameter space available to us is small, we follow Eicher *et al.* (2011) and opt for a commonly used default prior - the unit information prior (UIP). UIP sets g equal to number of observations, providing approximately the same amount of information as one individual observation in the dataset (Eicher *et al.* 2011). Regarding our prior choice on model space, we do not follow the traditional approach of using uniform model prior assigning the same probability to each model, irrespective of the number of included control variables. Instead, in light of our discussion on correlation among included covariates, we follow George *et al.* (2010) and employ the collinearity adjusted dilution model prior. Unlike to uniform model prior, dilution model prior drops the assumption of zero correlation between explanatory variables. When applying the dilution model prior, the posterior probabilities of models including highly correlated covariates are adequately down-weighted to account for this multicollinearity (Hasan *et al.* 2018).

Given our choices, BMA estimated with the unit information prior and dilution model prior represents our baseline model. We use an R package 'BMS' written by Zeugner (2011) to estimate BMA. Moreover, considering the high unbalancedness of our dataset in terms of number of estimates reported by each study, we follow Havranek *et al.* (2018) and weight our baseline model by the inverse of the number of estimates in each study. On top of the baseline model, we employ the following robustness checks:

(i) BMA with different weighting
First, we check the sensitivity of our results by applying a different weighting scheme to our baseline BMA. We weight our BMA by the precision (1/SE) of collected estimates, attributing more weight to more precise estimates and increasing the efficiency of estimated results. Havranek *et al.* (2017) explain that precision-weighted BMA is a viable robustness check as it tackles the heteroscedasticity of the estimated relationship.

(ii) Frequentist Model Averaging (FMA)
As our alternative specification to BMA, we employ Frequentist Model Averaging (FMA). Similarly to BMA, FMA incorporates the model uncertainty by averaging the desired estimator across different models. However, in contrast to BMA, FMA is entirely data-dependent and does not require prior specification (Wang *et al.* 2009). To implement FMA, we adopt the approach suggested by Havranek *et al.* (2017). Following their approach, we estimate the Mallows' model averaging estimator that determines the weights by minimizing the Mallows' criterion (Amini & Parmeter 2012). The smaller the Mallows' criterion, the smaller the model variance and the better the goodness-of-fit of the model. The approach builds upon Magnus *et al.* (2010) and reduces the model space from 2^{26}

to the number of explanatory variables equal to 26, taking advantage of the orthogonalization of the covariate space (Amini & Parmeter 2012).

(iii) Simple OLS (Frequentist check)
Finally, we proceed with estimating Equation 6.4 with simple OLS. In this specification, we cluster standard errors at the study level and include only variables that yield posterior inclusion probability higher or equal to 0.5 in the baseline BMA.

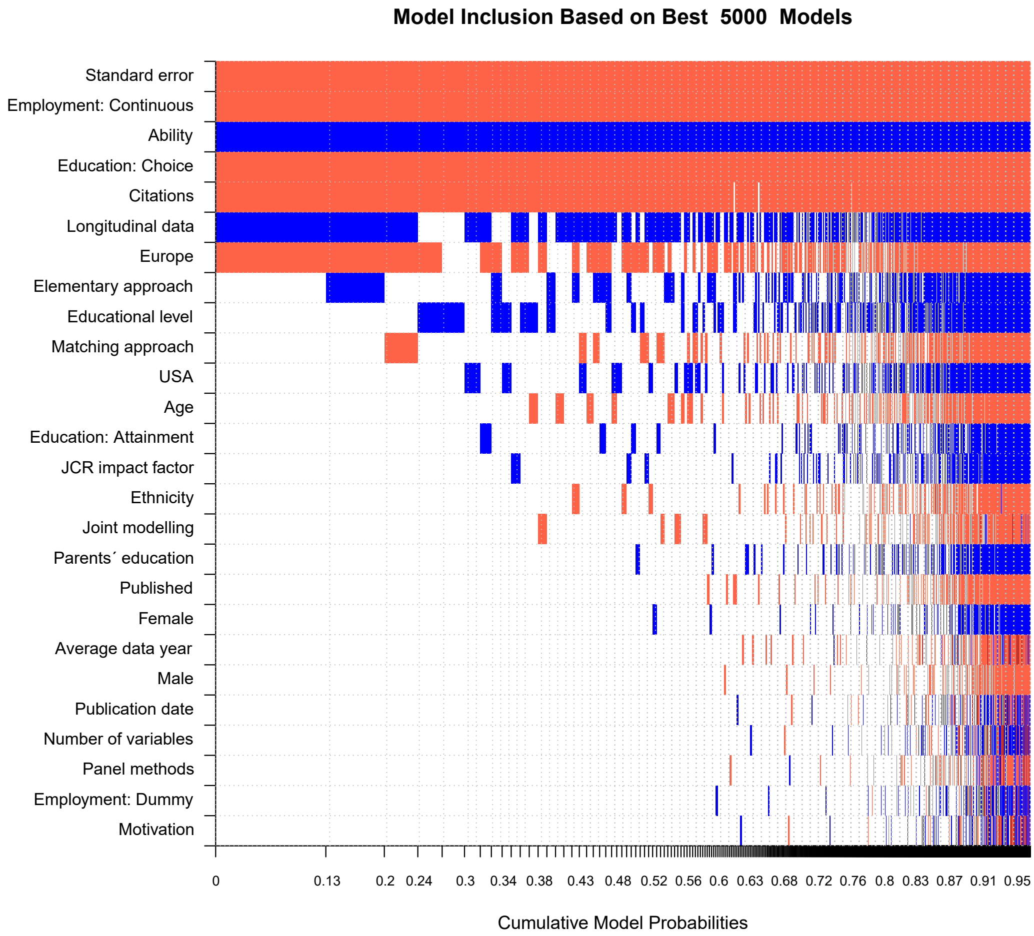
Finally, although it is a common practice to run BMA with different sets of priors, we refrain from running alternative specifications. Our reason for not pursuing BMA with a different set of priors is twofold. First, given the high collinearity between employed regressors, other prior distributions over model space would yield imprecise results. Second, as we know little about the true parameter sign, we find it redundant to randomly apply other non-informative default g-priors.

6.3 Results

Results of our BMA exercise are visualized in Figure 6.1. The vertical axis lists the explanatory variables from the largest to the smallest by their posterior inclusion probability. Hence, the most significant predictors lie on the top of the plot. The horizontal axis depicts the best models. The width of each column corresponds to the posterior model probability. Coloring of the figure has the following meaning: white space signifies exclusion of a particular variable from the model, red color (lighter in a greyscale picture) indicates a negative coefficient for a particular variable, and blue color (darker in a greyscale picture) indicates a positive coefficient. Surpassing the PIP value of 0.5, we identify seven variables explaining heterogeneity among the calculated PCCs: *Standard error, Employment: Continuous, Ability, Education: Choice, Citations, Longitudinal data, Europe*.

We accompany the graphical output of BMA with a quantitative results reported in Table 6.2. The numerical results corroborate conclusions drawn from the plot. In light of Jeffreys's (1961) categorization, variables *Standard Error*, *Education* operationalized as a *Choice*, *Employment: Continuous* and controlling for students' *Ability* have a decisive effect. Further, the results demonstrate strong effect for number of *Citations* and weak effect for *Longitudinal data* and *Europe*.

Figure 6.1: Model inclusion for our baseline BMA estimation



Notes: The figure shows results of the baseline BMA estimation (g-prior = unit information prior, model prior = dilution prior) weighted by the inverse of the number of estimates reported per study. The vertical axis plots the explanatory variables according to their posterior inclusion probabilities in descending order. The horizontal axis depicts the posterior model probability. White cells indicate that variables is not included in the model. Blue (darker in the greyscale) cells imply that the estimated coefficient of variable is positive and red cells (lighter in the greyscale) indicate negative estimated coefficient of variable.

Block 1 - Publication bias and data characteristics: In line with our previous findings, we find support for the negative publication bias permeating the literature on the educational implications of student employment also in our heterogeneity analysis. The direction and significance of publication bias demonstrated by the negative posterior mean and high value of PIP of the variable *Standard error* preserves even after expanding our analysis by explanatory variables. Moreover, this finding remains robust for all our specifications including the precision-weighted BMA estimation, where we interpret the posterior mean of the intercept as the intensity of publication bias.

Taking advantage of longitudinal datasets seems to have a substantial impact on explaining differences in the estimated PCCs. *Longitudinal data* systematically generate more positive effect estimates of the student employment-education relationship compared to cross-sectional studies. This is in line with prior research demonstrating that studies based on longitudinal data yield less negative (Rothstein 2007; Oettinger 1999) or more positive (Stinebrickner & Stinebrickner 2003) effect estimates. Advancement of longitudinal studies over cross-sectional studies is twofold. First, longitudinal data tackle better the endogeneity of the decision to work that is considered the main reason for reporting a negative effect (Neyt *et al.* 2019; Rothstein 2007). Due to the time span, longitudinal data mitigate the self-selection bias by differencing out the fixed unobserved individual heterogeneity (Oettinger 1999). Second, longitudinal data overcome the difficulties of drawing causal inferences as work habits are verifiably measured before educational outcomes (Moulin *et al.* 2013). As a result, cross-sectional studies failing to control for time-invariant individual characteristics generate downward-biased estimates.

In contrast to the predictive importance of *Longitudinal data*, our results indicate that *Average data year* of the original dataset has no impact on heterogeneity of PCCs, showing no structural differences among student populations over the years. This result is consistent with the conclusion of Warren & Cataldi (2006), who find little time variation in the relationship between student work and high school dropout between years 1966-1997.

Block 2 - Estimation methods: Remarkably, our results suggest that estimation methods do not yield systematically different effect estimates. On one hand, the finding is surprising if we consider the varying ability of these methods to control for the endogenous students' decision to work. On the other hand, this finding is consistent with Darolia (2014), who points out that although effect estimates generated by different estimation techniques differ in terms of their magnitude, in general they are statistically non-significant.

An alternative explanation for little impact of estimation methods might be embedded in the difficulty to correctly employ more advanced estimation techniques such as instrumental variable approach. As noted by Oettinger (1999), it is challenging to find an instrument satisfying the exogeneity and relevance assumption. For instance, Baert *et al.* (2017) explain that conditions on the local labour market, often used as an instrument, may affect students' decision to work, e.g. highly saturated market labour decreases students' chance of finding a job, and hence influence students' educational outcomes. Similarly, Buscha *et al.* (2012) argue that state child labour laws do not have to be necessarily exogenous to educational outcomes as they reflect the general importance of academic attainment in the specific region.

Block 3 - Design of the analysis: We observe that analyses conducted in *Europe* deliver more negative estimates of the student work-education relationship compared to analyses conducted in the USA or other countries. This finding is consistent with Neyt *et al.* (2019), who conclude that European countries yield more negative estimates because they predominantly examine the effect for university students, for whom the mean effect is more negative.

This explanation fits our sample relatively well. The weighted mean effect reported in Table 4.2 is more negative for tertiary students compared to secondary students. Moreover, in our sample 84% of European countries investigate the effect for higher education students. Neyt *et al.* (2019) argue that employment experience has more detrimental impact on educational outcomes at the university level due to the more demanding curriculum in tertiary institutions and the changing attitude towards the importance of their studies. As students move closer to their working life, they value more their work experience, neglecting their studies. We find this explanation plausible for higher education students in Europe. Education on tertiary level is free of charge in many European countries. Therefore, compared to their peers in the US, European students need smaller financial resources to sustain their living expenses while studying, resulting in different incentives to work. European students tend to view employment as a signal of their capabilities and competence to their future employers (Beerkens *et al.* 2011). In combination with decreasing legitimacy of university diploma (Yanbarisova 2015), individuals studying in Europe change their attitude towards their studies and perceive education only as an addition to their work experience, resulting in a more negative effect of their work experience on their academic success.

Block 4 - Variable specification: Furthermore, the BMA estimation suggests that operationalizing educational outcome as *Educational Choice* generates more negative PCCs. This finding resonates with Neyt *et al.* (2019) who report that studies operationalizing educational outcome as study progression deliver consistently negative or neutral relationship compared to studies using other educational outcomes. Intuitively, one can explain the negative relationship via the mechanism of zero-sum theory. Crowding out study time translates into poor test performance and exam failures, resulting progressively in a state, in which students prefer to drop-out from a certain course or study programme

(Parent 2006). Hence, our finding provides support for the notion that the effect of student employment "grows in cumulative importance" (Warren *et al.* 2000, pg. 949) and has long-term effects on educational outcomes.

Nevertheless, this explanation overlooks students' diverse backgrounds and expectations, mediating the relationship. Eckstein & Wolpin (1999) develop a structural model of high school attendance and show that although student employment increases the probability of dropout decision, the effect is driven by students' specific characteristics such as their ability, motivation, and preferences concerning free time. Hence, students for whom being employed results in dropping out from their studies are systematically different from students who are successful in combining their work experience with studying, e.g. they are primarily work-oriented or possess characteristics less applicable in academic setting. Thus, contrarily to zero-sum theory, explanation suggested by Eckstein & Wolpin (1999) rather supports the modern theoretical perspectives.

Another factor negatively influencing the effect estimate is whether student *Employment* is specified as a *Continuous* variable. This finding shows that what primarily matters for the effect of student employment on educational outcomes is the intensity of students' work schedule. This finding is consistent with the zero-sum perspective and conclusion cited in multiple studies; working long hours while studying has detrimental impact on educational outcomes (D'Amico 1984; Montmarquette *et al.* 2007; Buscha *et al.* 2012; Lee & Staff 2007). For instance, Montmarquette *et al.* (2007) show that "working less than fifteen hours per week is not necessarily detrimental to success in school" (pg. 759). Beffy *et al.* (2013) confirm this inflection point and show that spending at work more than 16 hours per week has strong negative effect on the graduation probability, whereas working less than 16 hours has much smaller effect. We explore the possibility that low working hours have less negative or even positive effect on educational outcomes in our further analysis.

Block 5 - Publication characteristics: We find little evidence that *Publication date*, whether study is *Published* or not, and journal quality measured by *JCR impact factor* systematically influence the reported effect estimates. On the other hand, what matters for heterogeneity of effect estimates is *Number of citations* a particular study has received, considering the high PIP value for this variable. Frequently cited studies yield systematically more negative effect estimates. We propose the following explanations for this trend: i) researchers cite these studies more often to corroborate their negative findings, ii) researchers refer to studies reporting negative estimates when highlighting the improvements of their studies that yield more positive estimates, iii) research papers yielding negative estimates possess superior methodological quality, and hence are cited more often. Unfortunately, our analysis cannot confirm nor reject any of these explanations, making them equally plausible.

Block 6 - Student characteristics controlled for in original estimation: BMA results further indicate that accounting for students' *Motivation, Age, Ethnicity*, and *Parental education* in the original estimation is not important for explaining variation in the effect of student employment on educational outcomes. Contrarily, controlling for students' *Ability* results in more positive effect estimates. This finding is in line with Arano & Parker (2008) and Carr *et al.* (1996), who find that measures of cognitive ability such as ACT or ASVAB score[2] are positively correlated with academic performance. In practice, the positive impact of controlling for students' *Ability* in the original estimation means that regressions not controlling for this covariate yield downward-biased estimates overstating the negative effect of student employment on academic performance and reflecting mostly experience of students with lower ability.

As noted earlier, as robustness checks we pursue BMA with different weighting, FMA and frequentist check. The results of these specifications are reported in Table 6.2. The key findings described in this section hold for FMA and frequentist check as opposed to BMA estimation weighted by precision, where we notice some differences. In line with our baseline BMA estimation, we observe that *Education: Choice, Employment: Continuous*, and *Longitudinal data* retain their strong effect as suggested by their PIP surpassing value of 0.95. On the other hand, PIP values for variables *Europe, Citations* and *Ability* reduce below 0.5 mitigating the importance of these variables. At the same time, looking at other PIP values, we observe that variables such as *Elementary approach, Motivation* and *Ethnicity* gain on significance. As a result of applying precision weights to BMA, we obtain a modified combination of study characteristics explaining heterogeneity between PCCs. Nonetheless, we look at the results of this robustness check cautiously, given that interpreting precision-weighted BMA estimation is troublesome due to increased correlation (Havranek *et al.* 2018). The graphical results of the precision-weighted BMA are presented in Appendix C. We further investigate robustness of our results in the next section, where we run the BMA exercise on two subsamples of our data.

[2]ACT refers to American College Testing, a standardized test capturing academic preparation. ASVAB stands for Armed Services Vocational Aptitude Battery test used to predict students' predisposition for military career.

Table 6.2: Explaining heterogeneity in PCCs capturing the student employment-education relationship

	BMA (weighted by No. obs.)			*BMA (weighted by SE)*			*Frequentist Model Averaging*			*Frequentist check (OLS)*		
	Post. Mean	*Post. SD*	*PIP*	*Post. Mean*	*Post. SD*	*PIP*	*Coef.*	*SE*	*p-value*	*Coef.*	*SE*	*p-value*
Standard error	-1.445	0.140	1.000	NA	NA	NA	-1.441	0.156	0.000	-1.430	0.351	0.000
Data characteristics												
Average data year	0.000	0.001	0.029	0.001	0.024	0.015	-0.010	0.006	0.077			
Longitudinal data	0.016	0.012	0.706	0.117	0.015	1.000	0.013	0.008	0.123	0.028	0.017	0.107
Estimation methods												
Elementary approach	0.004	0.007	0.287	0.053	0.016	0.970	0.016	0.012	0.205			
Matching approach	-0.007	0.016	0.190	-0.010	0.025	0.167	-0.025	0.020	0.209			
Panel methods	0.000	0.002	0.018	0.000	0.003	0.021	-0.002	0.016	0.889			
Joint modelling	-0.001	0.004	0.067	0.000	0.003	0.019	0.004	0.014	0.748			
Design of the analysis												
Europe	-0.013	0.011	0.654	0.000	0.003	0.017	-0.021	0.010	0.030	-0.024	0.014	0.073
USA	0.003	0.007	0.189	0.000	0.004	0.026	0.009	0.008	0.267			
Educational level	0.005	0.009	0.286	0.001	0.005	0.039	-0.001	0.008	0.915			
Number of variables	0.000	0.000	0.019	-0.007	0.010	0.341	0.002	0.004	0.686			
Variable specification												
Education: Choice	-0.043	0.008	1.000	-0.113	0.014	1.000	-0.045	0.008	0.000	-0.047	0.015	0.001
Education: Attainment	0.001	0.004	0.085	0.000	0.002	0.014	0.014	0.008	0.072			
Employment: Continuous	-0.047	0.005	1.000	-0.072	0.011	1.000	-0.048	0.007	0.000	-0.051	0.013	0.000
Employment: Dummy	0.000	0.001	0.017	0.001	0.004	0.040	0.006	0.008	0.448			

Table 6.2: Explaining heterogeneity in PCCs capturing the student employment-education relationship (continued)

	BMA (weighted by No. obs.)			*BMA (weighted by SE)*			*Frequentist Model Averaging*			*Frequentist check (OLS)*		
	Post. Mean	*Post. SD*	*PIP*	*Post. Mean*	*Post. SD*	*PIP*	*Coef.*	*SE*	*p-value*	*Coef.*	*SE*	*p-value*
Publication characteristics												
Publication date	0.000	0.000	0.019	0.001	0.004	0.040	0.009	0.004	0.038			
Citations	-0.010	0.002	0.988	-0.001	0.004	0.122	-0.011	0.003	0.000	-0.011	0.006	0.068
Published	-0.001	0.004	0.047	0.000	0.003	0.017	-0.015	0.010	0.107			
JCR impact factor	0.000	0.001	0.085	0.000	0.002	0.044	0.007	0.003	0.012			
Student characteristics												
Ability	0.036	0.005	1.000	0.000	0.002	0.017	0.037	0.006	0.000	0.032	0.012	0.008
Motivation	0.000	0.001	0.015	0.035	0.019	0.840	-0.004	0.006	0.502			
Parents' education	0.001	0.003	0.061	0.002	0.007	0.084	0.009	0.006	0.140			
Age	-0.002	0.005	0.146	0.000	0.002	0.024	-0.005	0.007	0.479			
Ethnicity	-0.001	0.003	0.077	-0.042	0.019	0.890	-0.014	0.007	0.034			
Male	0.000	0.002	0.020	0.000	0.002	0.021	-0.011	0.009	0.245			
Female	0.000	0.003	0.031	0.000	0.003	0.032	0.010	0.011	0.327			
Intercept	0.004	NA	1.000	-0.028	NA	1.000	0.005	0.002	0.002	0.020	0.028	0.482
Studies	69			69			69			69		
Observations	861			861			861			861		

Notes: The table displays results for BMA baseline model weighted by the inverse of the number of observations reported per study, the alternative BMA model weighted by the inverse of standard error, FMA estimation and frequentist check that includes only variables with PIP > 0.5 in the BMA baseline model. BMA uses the unit information prior and the dilution prior accounting for potential collinearity. Variables are described in Table 6.1. Post. Mean = Posterior Mean. Post. SD = Posterior standard deviation. PIP = Posterior inclusion probability.

6.4 Robustness check using subsamples

To check sensitivity of our results, we perform BMA on two subsamples of our dataset. For both subsamples, we run two BMA specifications using the unit information prior and dilution model prior, while weighting estimates by the number of reported estimates per study and by precision.

The first subsample contains only estimates specifying student employment variable in a continuous form. Compared to the baseline dataset, this subsample excludes variables *Employment: Continuous* and *Employment: Dummy*. Moreover, this subsample does not include *Matching approach* since the variable contains no data for this subsample. Moreover, we substitute estimation method *Joint Modelling* by *Instrumental variable* as no study specifying student employment as a continuous variable uses *Simultaneous modelling* for estimation. In the second subsample, we use only estimates where the student employment variable is operationalized as a categorical variable, distinguishing between different levels of work intensity. Similarly, to the first subsample, we omit variables describing the form of student employment variable. Furthermore, to reflect the varying work intensity of estimates, we employ two additional dummy variables describing students' work intensity. Variable *Low intensity* applies to estimates capturing the effect for students working up to 15 hours per week and variable *Medium intensity* captures the weekly intensity varying from 16 to 30 hours. Variable *High intensity* is not included to avoid dummy variable trap. We also intended to perform BMA analysis on the untransformed homogenous subsample as we did in Section 5.3, however, BMA is infeasible for the untransformed subsample.

We report quantitative results of these robustness checks in Appendix C. In line with our previous findings, we find that the negative publication bias is present in these subsamples. Although the coefficients of posterior mean for Standard error (for BMA weighted by number of estimates reported per study) and Intercept (for precision-weighted BMA) decrease in magnitude, they retain their negative signs. Moreover, negative publication bias maintains its significance demonstrated by PIP value equal to 1, apart from BMA weighted by number of estimates for the second subsample.

Looking at the first subsample, BMA estimation explains little variation in estimates specifying student employment as a continuous variable. Apart from *Number of citations* and *Ability* in the estimates-weighted specification and *Longitudinal data* in the precision-weighted specification, no variables have an impact on the estimates. The other two variables (*Education: Choice* and *Europe*) with PIP above 0.5 in the baseline model, do not surpass the 0.5 value in this subsample. We suggest two explanations for this. First, once accounting for work intensity, estimates become relatively uniform and show little variation. Second, the factors we include in BMA are less relevant for this subsample of estimates.

In contrast to the first subsample, running BMA for the second subsample yields quite a different set of significant predictors of heterogeneity. This is not surprising as estimates included in this subsample are highly heterogeneous, given the different work intensity levels. Nevertheless, we would like to point out the importance and direction of variable *Low intensity*. Estimates of low work intensities generate systematically more positive estimates of the effect of student employment on educational outcomes. This finding is in line with Buscha *et al.* (2012), who argue that less intense involvement is beneficial to study outcomes. Furthermore, it corroborates the plausibility of threshold perspective, which predicts that student employment has positive effect on educational outcome up to a certain amount of working hours, at which the effect reverses.

6.5 Best practice estimate

The findings of the previous analyses show that the literature concerning educational implications of student employment suffers from negative publication bias and that the heterogeneity in collected estimates can be explained by specific study characteristics. In this section, we explore what would be the mean effect if we controlled for these influences. In particular, using the synthetic study approach, we construct a 'best practice' estimate corrected for the underrepresentation of positive estimates and impact of study characteristics. In practice, we obtain the 'best practice' PCC by estimating a linear regression, where we weight each included variable by a preferred value. The preferred value corresponds to sample minimum if the variable indicates bad practice, sample maximum for the best practice and sample mean if we have no preference.

Admittedly, this exercise is of an exploratory nature. Thus, from the very beginning, we would like to point out shortcomings of this practice. First, defining a best practice estimate is inherently subjective as we decide about the aspects of an ideal study. Second, in our case, the best practice estimate lacks a clear economic meaning due to the use of PCCs.

In our view, an ideal estimate should fulfill the following conditions. First, it should be estimated using *Longitudinal data*, allowing researchers to discern causal effects from correlations. Hence, we use sample maximum for this variable. Second, although BMA yields estimation methods to be non-significant, we want to filter out potential endogeneity bias and prefer estimation techniques, circumventing the endogeneity issue. Therefore, we plug sample maximum for *Panel methods* and *Joint modelling*. Regarding specification of student employment, we prefer when student *Employment* is measured as a *Continuous* variable since it directly reflects the impact of student employment intensity. Also, we plug 1 for all dummies reflecting whether original studies control for specific student characteristics. Additionally, as we are interested in the most recent estimate, we set the *Average data year* at its sample maximum. Finally, we believe that newest, published studies from journals awarded with a high *JCR impact factor* are likely to produce the best estimate. In consequence, we set the publication characteristics at their sample maxima, apart from *Citations* where we plug the sample mean because highly cited studies do not necessarily render superior quality. Contrarily, as we want to avoid publication bias, we set *Standard error* at its sample minimum. As for the operationalization of educational outcome, we do not hold any preferences because from the methodological perspective none of the measurement forms is superior. Finally, we do not hold any preferences for the remaining variables, and thus we leave them at their sample mean.

Following our preferences, Table 6.3 reports multiple best practice estimates and the corresponding 95% confidence intervals. We calculate the confidence intervals using OLS with clustered standard errors at the study level. The best practice estimate of the partial correlation coefficient equals -0.020. We observe twice as high negative effect estimate for studies conducted in Europe compared to studies conducted in USA. Moreover, we see that student employment has the most detrimental effect on educational choice. The effect on educational attainment such as graduation probability and exam scores is similar.

Nonetheless, all the implied best-practice estimates are close to 0. Even if we change weights for variables, the very low negative predicted estimates persist. These results further substantiate our findings from the previous sections, where we found a negligible effect of student employment on educational outcomes.

Table 6.3: Predicted 'best practice' estimates

	Predicted Estimate	95% Confidence Interval	
All	-0.020	-0.094	0.047
Europe	-0.032	-0.116	0.026
USA	-0.016	-0.086	0.056
Education: Choice	-0.057	-0.129	0.004
Education: Attainment	-0.012	-0.085	0.076
Education: Test scores	-0.013	-0.089	0.054

Notes: The table reports the 'best practice' estimate of the PCC capturing the effect of student employment on educational outcomes. The choices reflected in the 'best practice' are described in the text. The confidence intervals are constructed using OLS with clustered standard errors at the study level.

Chapter 7

Conclusion

The objective of this thesis is to conduct a quantitative review of studies investigating the impact of student employment on educational outcomes. Specifically, we test for the presence of publication bias using graphical and quantitative methods. Furthermore, we identify study characteristics explaining heterogeneity among reported estimates using the Bayesian Model Averaging, an estimation method accounting for model uncertainty. To perform these analyses, we construct an original dataset including 861 partial correlation coefficients capturing the student work-education relationship in 69 studies.

From the theoretical perspective, our work contributes to the existing research on the student work-education relationship in three ways. First, to our knowledge, this thesis constitutes the first meta-analysis of primary studies investigating the educational implications of students' work experience. Second, by synthesizing estimates from different studies, we examine the possibility that researchers reckon on an effect of certain direction, resulting in the publication selection bias. Finally, we explore which study characteristics, methodological aspects and employed control variables cause variation in the reported estimates across studies. As this is usually done using a sample of students from one course/school/country, our work advances other studies in the level of external validity as our results combine findings of multiple studies using different student bodies. Therefore, our findings are better generalizable to other student populations.

As a result of these analyses, first, we find that positive estimates are slightly underrepresented in the literature on the student work-education relationship, indicating the presence of a negative publication bias. However, this negative publication bias disappears for a subsample of studies controlling for the endogeneity of students' decision to work. Therefore, we suggest that researchers report negative estimates more frequently not because of shared preference for such estimates, but because of neglecting the fact that work experience is a self-selected activity influenced by students' characteristics that simultaneously affect educational outcomes.

Second, the linear tests of publication bias indicate that the mean effect sterilized from the publication bias is negligible, oscillating around zero. This small mean estimate remains robust across different weighting schemes and specifications. Further, the non-linear tests of publication bias also generate a very small mean estimate. This finding supports the view of modern theoretical perspectives, stating that the effect of student employment on educational outcome is not directly attributable to employment but rather to students' pre-existing characteristics and attitude towards education and employment.

Third, results of the Bayesian Model Averaging estimation provide evidence for systematic variation in the estimated partial correlation coefficients caused by study characteristics. We find that estimates generated by studies employing longitudinal data and controlling for students' ability are systematically more positive. This implies that studies failing to control for students' permanent traits such as ability tend to overstate the negative effect of student employment on students' educational outcomes. Similarly, we observe that operationalizing educational outcome as the dropout rate yields consistently negative estimates. Further, we find that studies operationalizing student employment in a continuous form, e.g. as the number of hours worked per week, produce systematically more negative effect estimates. In practice, it means that working more hours per week yield more detrimental effect of employment on one's academic performance. This finding is not only in line with the zero-sum perspective, stating that employment crowds out time important for studying, but also with the threshold perspective arguing that working has a negative impact on academic performance only when a certain amount of hours worked per week is exceeded. We further corroborate the threshold perspective by conducting a robustness check, where estimates of low work intensity yield systematically positive effect.

In addition to the academic relevance highlighted above, our results also provide practical implications for education policy makers. For instance, our findings might serve as guidelines for policy recommendations concerning youth employment programmes, integrating professional experience in study programs. Alternatively, our findings are relevant for school counsellors, who should primarily consider students' work-education attitudes, learning predispositions and time intensity of potential work experience when providing guidance to students.

Beyond the contributions of this thesis, we consider two limitations of this work. First, the funnel and precision asymmetry tests (FAT-PET) operate under two strong assumptions. The first concerns the linear relationship between effect estimates and their standard errors, while the second one entails the exogeneity condition of zero correlation between estimates and standard errors in the absence of publication bias. Since the results of linear tests of publication bias are conditional on these assumptions, we pursue alternative tests that relax these assumptions. To circumvent the linearity assumption, we employ non-linear techniques including Weighted Average of Adequately Powered introduced by Ioannidis *et al.* (2017) or novel methods such as Selection Model and Stem-based method introduced by Andrews & Kasy (2019) and Furukawa (2019), respectively. To address the exogeneity assumption, we conduct Caliper test and estimate FAT-PET with an instrumental variable. Nevertheless, none of these techniques can perfectly resolve the exogeneity issue. For Caliper test, the exogeneity assumption must still hold within each individual caliper. For estimating FAT-PET with an instrument, we cannot say with certainty that the chosen instrument for standard error, square root of the number of observations, is uncorrelated with estimation technique.

Second, while the use of partial correlation coefficients represents a common practice in meta-analyses to achieve a comparable effect size, this transformation of original effect estimates has been criticized. While Sachar (1980) points out the problematic interpretation of partial correlation coefficients, Reed (2020) shows that using partial correlation coefficients might yield different mean effect estimates and aspects responsible for heterogeneity when compared to the original estimates. Given this criticism, we would prefer to use original effect estimates. However, primary studies examining the educational implications of student work use various measures of the dependent and independent variables, ruling out the option to conduct a meta-analysis with original estimates. Nevertheless, as a robustness check we conduct tests of publication bias on the untransformed subsample of original effect estimates, yielding results consistent with our findings when using partial correlation coefficients. Hence, although it is important to acknowledge the pitfalls of partial correlation coefficients, we find it reasonable to use partial correlation coefficients as they yield reliable results and are regularly embraced by meta-analysts.

Bibliography

Amini, S. M. & C. F. Parmeter (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **109(8)**: pp. 2766–2794.

Apel, R., S. D. Bushway, R. Paternoster, R. Brame, & G. Sweeten (2008): "Using state child labor laws to identify the causal effect of youth employment on deviant behavior and academic achievement." *Journal of Quantitative Criminology* **24(4)**: pp. 337–362.

Applegate, C. & A. Daly (2006): "The impact of paid work on the academic performance of students: A case study from the university of canberra." *Australian Journal of Education* **50(2)**: pp. 155–166.

Arano, K. & C. Parker (2008): "How Does Employment Affect Academic Performance Among College Students?" *Journal of Economic Insight* **34(2)**: pp. 65–82.

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

Babcock, P. & M. Marks (2011): "The Falling Time Cost of College: Evidence from Half a Century of Time Use Data." *The Review of Economics and Statistics* **93(2)**: pp. 468–478.

Baert, S., I. Marx, B. Neyt, E. V. Belle, & J. V. Casteren (2018): "Student employment and academic performance: an empirical exploration of the primary orientation theory." *Applied Economics Letters* **25(8)**: pp. 547–552.

Baert, S., B. Neyt, E. Omey, & D. Verhaest (2017): "Student Work, Educational Achievement, and Later Employment: A Dynamic Approach." *IZA Discussion Papers 11127*, Institute of Labor Economics (IZA), Bonn.

Becker, G. S. (1965): "A theory of the allocation of time." *The Economic Journal* **75(299)**: pp. 493–517.

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

Beffy, M., D. FougĂ¨re, & A. Maurel (2013): "The Effect of College Employment on Graduation: Evidence from France." *CEPR Discussion Papers 9565*, C.E.P.R. Discussion Papers.

Body, K. M.-D., L. Bonnal, & J.-F. Giret (2014): "Does student employment really impact academic achievement? The case of France." *Applied Economics* **46(25)**: pp. 3061–3073.

Bom, P. & H. Rachinger (2019): "A kinked meta-regression model for publication bias correction." *Research Synthesis Methods* **10(4)**: pp. 497–514.

Bozick, R. (2007): "Making it through the first year of college: The role of students' economic resources, employment, and living arrangements." *Sociology of Education* **80(3)**: pp. 261–285.

Buscha, F., A. Maurel, L. Page, & S. Speckesser (2012): "The effect of employment while in high school on educational attainment: A conditional difference-in-differences approach." *Oxford Bulletin of Economics and Statistics* **74(3)**: pp. 380–396.

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

Card, D. & A. B. Krueger (1995): "Time-series minimum-wage studies: A meta-analysis." *The American Economic Review* **85(2)**: pp. 238–243.

Carneiro, P. M. & J. J. Heckman (2003): "Human capital policy." *IZA Discussion Papers 821*, Institute of Labor Economics (IZA), Bonn.

Carr, R. V., J. D. Wright, & C. J. Brody (1996): "Effects of high school work experience a decade later: Evidence from the national longitudinal survey." *Sociology of Education* **69(1)**: pp. 66–81.

Cazachevici, A., T. Havranek, & R. Horvath (2020): "Remittances and economic growth: A meta-analysis." *World Development* **134(C)**.

Choi, Y. (2018): "Student employment and persistence: Evidence of effect heterogeneity of student employment on college dropout." *Research in Higher Education* **59(1)**: pp. 88–107.

Clarivate Analytics (2019): "The clarivate analytics impact factor." Retrieved from https://clarivate.com/webofsciencegroup/essays/impact-factor/. Accessed: 21/10/2020.

Dadgar, M. (2012): "The academic consequences of employment for students enrolled in community college." *CCRC Working Paper 46*, Community College Research Center.

D'Amico, R. (1984): "Does employment during high school impair academic progress?" *Sociology of Education* **57(3)**: pp. 152–164.

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(C)**: pp. 38–50.

DeSimone, J. (2006): "Academic Performance and Part-Time Employment among High School Seniors." *The BE Journal of Economic Analysis & Policy* **6(1)**: pp. 1–36.

DeSimone, J. S. (2008): "The Impact of Employment during School on College Student Academic Performance." *NBER Working Papers 14006*, National Bureau of Economic Research, Inc.

Dickersin, K. (2005): *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, chapter Publication Bias: Recognizing the Problem, Understanding Its Origins and Scope, and Preventing Harm, pp. 11–33. The Atrium, Southern Gate, Chichester,West Sussex PO19 8SQ, England: John Wiley & Sons.

Doucouliagos, C. & P. Laroche (2003): “What do unions do to productivity? a meta-analysis.” *Industrial Relations: A Journal of Economy and Society* **42(4)**: pp. 650–691.

Doucouliagos, H. (2011): “How large is large? preliminary and relative guidelines for interpreting partial correlations in economics.” *Working Papers* $eco_2 011_5, DeakinUniversity, DepartmentofEconomics.$

Doucouliagos, H. & T. D. Stanley (2009): “Publication Selection Bias in Minimumâ€Wage Research? A Metaâ€Regression Analysis.” *British Journal of Industrial Relations* **47(2)**: pp. 406–428.

Dustmann, C. & A. Soest (2007): “Part-time work, school success and school leaving.” *Empirical Economics* **32(2)**: pp. 277–299.

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

Egger, M., G. D. Smith, M. Schneider, & C. Minder (1997): “Bias in meta-analysis detected by a simple, graphical test.” *BMJ* **315(7109)**: pp. 629–634.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): “Default priors and predictive performance in bayesian model averaging, with application to growth determinants.” *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Furukawa, C. (2019): “Publication Bias under Aggregation Frictions: Theory, Evidence, and a New Correction Method.” *EconStor Preprints 194798*, ZBW - Leibniz Information Centre for Economics.

Gechert, S., T. Havranek, Z. Irsova, & D. Kolcunova (2019): “Death to the Cobb-Douglas Production Function.” *Working Papers IES 2019/26*, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Geel, R. & U. Backes-Gellner (2012): “Earning while learning: When and how student employment is beneficial.” *Labour* **26(3)**: pp. 313–340.

George, E. I. *et al.* (2010): “Dilution priors: Compensating for model space redundancy.” In “Borrowing Strength: Theory Powering Applications - A

Festschrift for Lawrence D. Brown," volume 6, pp. 158–165. Institute of Mathematical Statistics.

Gerber, A. & N. Malhotra (2008): "Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals." *Quarterly Journal of Political Science* **3(3)**: pp. 313–326.

Ghavam, M., F. Pourmalek *et al.* (2005): "Effects of dentistry students' employment on their academic success (2003-2004)." *Journal of Islamic Dental Association of Iran (Majallaj-I-Dandanpiyiskhi)* **17(1)**: pp. 104–112.

Gleason, P. M. (1993): "College student employment, academic progress, and postcollege labor market success." *Journal of Student Financial Aid* **23(2)**: pp. 5–14.

Hasan, I., R. Horvath, & J. Mares (2018): "What Type of Finance Matters for Growth? Bayesian Model Averaging Evidence." *World Bank Economic Review* **32(2)**: pp. 383–409.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

Havranek, T., R. Horvath, & A. Zeynalov (2016): "Natural resources and economic growth: A meta-analysis." *Working Papers IES 2016/03*, Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Havranek, T., Z. Irsova, L. Laslopova, & O. Zeynalova (2020): "The elasticity of substitution between skilled and unskilled labor: A meta-analysis." *MPRA Paper 102598*, University Library of Munich, Germany.

Havranek, T., Z. Irsova, & O. Zeynalova (2018): "Tuition fees and university enrolment: A meta-regression analysis." *Oxford Bulletin of Economics and Statistics* **80(6)**: pp. 1145–1184.

Havranek, T., M. Rusnak, & A. Sokolova (2017): "Habit formation in consumption: A meta-analysis." *European Economic Review* **95(C)**: pp. 142–167.

Hedges, L. V. (1992): "Modeling publication selection effects in meta-analysis." *Statistical Science* **7(2)**: pp. 246–255.

Heinemann, F., M.-D. Moessinger, & M. Yeter (2018): "Do fiscal rules constrain fiscal policy? A meta-regression-analysis." *European Journal of Political Economy* **51(C)**: pp. 69–92.

Hoeting, J. A., D. Madigan, A. E. Raftery, & C. T. Volinsky (1999): "Bayesian model averaging: a tutorial." *Statistical Science* **14(4)**: pp. 382–401.

Holford, A. (2020): "Youth employment, academic performance and labour market outcomes: Production functions and policy effects." *Labour Economics* **63(C)**.

Hovdhaugen, E. (2015): "Working while studying: The impact of term-time employment on dropout rates." *Journal of Education and Work* **28(6)**: pp. 631–651.

Hwang, J.-K. (2013): "Employment and student performance in principles of economics." *International Review of Economics Education* **13(C)**: pp. 26–30.

Ioannidis, J. P. A., T. D. Stanley, & H. Doucouliagos (2017): "The Power of Bias in Economics Research." *Economic Journal* **127(605)**: pp. 236–265.

Jeffreys, H. (1961): "Small corrections in the theory of surface waves." *Geophysical Journal International* **6(1)**: pp. 115–117.

Jones, C. M., J. P. Green, & H. E. Higson (2017): "Do work placements improve final year academic performance or do high-calibre students choose to do work placements?" *Studies in Higher Education* **42(6)**: pp. 976–992.

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

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

Kohen, A. I., G. Nestel, & C. Karmas (1978): "Factors affecting individual persistence rates in undergraduate college programs." *American Educational Research Journal* **15(2)**: pp. 233–252.

Kouliavtsev, M. (2013): "The impact of employment and extracurricular involvement on undergraduates' performance in a business statistics course." *Journal of Economics and Economic Education Research* **14(3)**: pp. 53–66.

Lanzarini, L., M. E. Charnelli, & J. Diaz (2015): "Academic performance of university students and its relation with employment." In "2015 Latin American Computing Conference (CLEI)," pp. 1–6. IEEE.

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

Lee, J. C. & J. Staff (2007): "When work matters: The varying impact of work intensity on high school dropout." *Sociology of Education* **80(2)**: pp. 158–178.

Leos-Urbel, J. (2014): "What is a summer job worth? the impact of summer youth employment on academic outcomes." *Journal of Policy Analysis and Management* **33(4)**: pp. 891–911.

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

MacKinnon, J. G. & M. D. Webb (2017): "Wild Bootstrap Inference for Wildly Different Cluster Sizes." *Journal of Applied Econometrics* **32(2)**: pp. 233–254.

Magnus, J. R., O. Powell, & P. Prufer (2010): "A comparison of two model averaging techniques with an application to growth empirics." *Journal of Econometrics* **154(2)**: pp. 139–153.

Mankiw, N. G., D. Romer, & D. N. Weil (1992): "A contribution to the empirics of economic growth." *The Quarterly Journal of Economics* **107(2)**: pp. 407–437.

Manthei, R. J. & A. Gilmore (2005): "The effect of paid employment on university students' lives." *Education + Training* **47(3)**: pp. 202–215.

Marsh, H. W. (1991): "Employment during high school: Character building or a subversion of academic goals?" *Sociology of Education* **64(3)**: pp. 172–189.

MARSH, H. W. & S. KLEITMAN (2005): "Consequences of employment during high school: Character building, subversion of academic goals, or a threshold?" *American Educational Research Journal* **42(2)**: pp. 331–369.

MATOUSEK, J., T. HAVRANEK, & Z. IRSOVA (2019): "Individual Discount Rates: A Meta-Analysis of Experimental Evidence." *EconStor Preprints 194617*, ZBW - Leibniz Information Centre for Economics.

MCCOY, S. & E. SMYTH (2007): "So Much to Do, so Little Time: Part-time Employment among Secondary Students in Ireland." *Work, Employment and Society* **21(2)**: pp. 227–246.

MCKECHNIE, J., K. DUNLEAVY, & S. HOBBS (2005): "Student Employment and its Educational Impact: A Scottish study." *Scottish Educational Review* **37(1)**: pp. 58–67.

MCKENZIE, K. & R. SCHWEITZER (2001): "Who succeeds at university? Factors predicting academic performance in first year Australian university students." *Higher Education Research & Development* **20(1)**: pp. 21–33.

MCNEAL, R. B. (1997): "Are students being pulled out of high school? The effect of adolescent employment on dropping out." *Sociology of Education* **70(3)**: pp. 206–220.

MCVICAR, D. & B. MCKEE (2002): "Part-time work during post-compulsory education and examination performance: Help or hindrance?" *Scottish Journal of Political Economy* **49(4)**: pp. 393–406.

MINASYAN, A., J. ZENKER, S. KLASEN, & S. VOLLMER (2019): "Educational gender gaps and economic growth: A systematic review and meta-regression analysis." *World Development* **122(C)**: pp. 199–217.

MONTMARQUETTE, C., N. VIENNOT-BRIOT, & M. DAGENAIS (2007): "Dropout, School Performance, and Working while in School." *The Review of Economics and Statistics* **89(4)**: pp. 752–760.

MORAL-BENITO, E. (2012): "Determinants of economic growth: a bayesian panel data approach." *Review of Economics and Statistics* **94(2)**: pp. 566–579.

Moulin, S., P. Doray, B. Laplante, & M. C. Street (2013): "Work intensity and non-completion of university: longitudinal approach and causal inference." *Journal of Education and Work* **26(3)**: pp. 333–356.

Neyt, B., E. Omey, D. Verhaest, & S. Baert (2019): "Does Student Work Really Affect Educational Outcomes? A Review Of The Literature." *Journal of Economic Surveys* **33(3)**: pp. 896–921.

Oehlert, G. W. (1992): "A note on the delta method." *The American Statistician* **46(1)**: pp. 27–29.

Oettinger, G. S. (1999): "Does high school employment affect high school academic performance?" *ILR Review* **53(1)**: pp. 136–151.

Parent, D. (2006): "Work while in high school in canada: its labour market and educational attainment effects." *Canadian Journal of Economics/Revue canadienne d'économique* **39(4)**: pp. 1125–1150.

Porter, D. (1997): "Higher education: Of costs & value." *Liberal Education* **83(2)**: pp. 55–57.

Post, D. & S.-l. Pong (2000): "Employment during middle school: The effects on academic achievement in the us and abroad." *Educational Evaluation and Policy Analysis* **22(3)**: pp. 273–298.

Quintini, G. (2015): "Working and learning: A diversity of patterns." *OECD Social, Employment and Migration Working Papers 169*, OECD Publishing, Paris.

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

Raftery, A. E. (1995): "Bayesian model selection in social research." *Sociological Methodology* **25**: pp. 111–163.

Reed, W. R. (2020): "A Note on the Use of Partial Correlation Coefficients in Meta-Analyses." *Working Papers in Economics 20/08*, University of Canterbury, Department of Economics and Finance.

Richardson, J. J., S. Kemp, S. Malinen, & S. A. Haultain (2013): "The academic achievement of students in a new zealand university: Does it pay to work?" *Journal of Further and Higher Education* **37(6)**: pp. 864–882.

Riggert, S. C., M. Boyle, J. M. Petrosko, D. Ash, & C. Rude-Parkins (2006): "Student employment and higher education: Empiricism and contradiction." *Review of educational research* **76(1)**: pp. 63–92.

Rochford, C., M. Connolly, & J. Drennan (2009): "Paid part-time employment and academic performance of undergraduate nursing students." *Nurse Education Today* **29(6)**: pp. 601–606.

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

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

Sabia, J. J. (2009): "School-year employment and academic performance of young adolescents." *Economics of Education Review* **28(2)**: pp. 268–276.

Sachar, J. (1980): "Cautions in the interpretation of the partial correlation coefficient." *The Journal of Experimental Education* **48(3)**: pp. 209–216.

Salamonson, Y. & S. Andrew (2006): "Academic performance in nursing students: Influence of part-time employment, age and ethnicity." *Journal of Advanced Nursing* **55(3)**: pp. 342–349.

Schoenhals, M., M. Tienda, & B. Schneider (1998): "The educational and personal consequences of adolescent employment." *Social forces* **77(2)**: pp. 723–761.

Scott-Clayton, J. & V. Minaya (2016): "Should student employment be subsidized? conditional counterfactuals and the outcomes of work-study participation." *Economics of Education Review* **52(C)**: pp. 1–18.

Simón, H., J. M. C. Diaz, & J. L. C. Costa (2017): "Analysis of university student employment and its impact on academic performance." *Electronic Journal of Research in Educational Psychology* **15(2)**: pp. 281–306.

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

Sprietsma, M. (2015): "Student employment: Advantage or handicap for academic achievement?" *ZEW Discussion Papers 15-085*, ZEW - Leibniz Centre for European Economic Research.

Sribney, W. & V. Wiggins (n.d): "Standard errors, confidence intervals, and significance tests for ors, hrs, irrs, and rrrs." Retrieved from https://www.stata.com/support/faqs/statistics/delta-rule/. Accessed: 15/05/2020.

Staff, J. & J. T. Mortimer (2007): "Educational and work strategies from adolescence to early adulthood: Consequences for educational attainment." *Social Forces* **85(3)**: pp. 1169–1194.

Staff, J., J. E. Schulenberg, & J. G. Bachman (2010): "Adolescent work intensity, school performance, and academic engagement." *Sociology of Education* **83(3)**: pp. 183–200.

Stanley, T. D. (2001): "Wheat from chaff: Meta-analysis as quantitative literature review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

Stanley, T. D. (2005): "Beyond publication bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Metaâ€regression methods for detecting and estimating empirical effects in the presence of publication selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Stanley, T. D., H. Doucouliagos, & J. P. Ioannidis (2017): "Finding the power to reduce publication bias." *Statistics in medicine* **36(10)**: pp. 1580–1598.

Stanley, T. D., S. B. Jarrell, & H. Doucouliagos (2010): "Could It Be Better to Discard 90% of the Data? A Statistical Paradox." *The American Statistician* **64(1)**: pp. 70–77.

Steel, L. (1991): "Early work experience among white and non-white youths: Implications for subsequent enrollment and employment." *Youth & Society* **22(4)**: pp. 419–447.

Stern, D. & D. Briggs (2001): "Does paid employment help or hinder performance in secondary school? insights from us high school students." *Journal of Education and Work* **14(3)**: pp. 355–372.

Sterne, J. A. C., B. J. Becker, & M. Egger (2005): *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, chapter The

Funnel Plot, pp. 75–98. The Atrium, Southern Gate, Chichester,West Sussex PO19 8SQ, England: John Wiley & Sons.

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

Tessema, M. T., K. J. Ready, & M. Astani (2014): "Does part-time job affect college students' satisfaction and academic performance (GPA)? The case of a mid-sized public university." *International Journal of Business Administration* **5(2)**: pp. 1–10.

Theune, K. (2015): "The working status of students and time to degree at german universities." *Higher Education* **70(4)**: pp. 725–752.

Tienda, M. & A. Ahituv (1996): "Ethnic differences in school departure: Does youth employment promote or undermine educational attainment?" In G. L. Mangum & S. L. Mangum (editors), "Of Heart and Mind: Social Policy Essays in Honor of Sar A. Levitan," chapter 4, pp. 93–110. Kalamazoo, Michigan: Upjohn Institute for Employment Research.

Torres, V., J. P. Gross, & A. Dadashova (2010): "Traditional-age students becoming at-risk: Does working threaten college students' academic success?" *Journal of College Student Retention: Research, Theory Practice* **12(1)**: pp. 51–68.

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)**: pp. 353–380.

Wang, H., M. Kong, W. Shan, & S. K. Vong (2010): "The effects of doing part-time jobs on college student academic performance and social life in a chinese society." *Journal of Education and Work* **23(1)**: pp. 79–94.

Wang, H., X. Zhang, & G. Zou (2009): "Frequentist model averaging estimation: a review." *Journal of Systems Science and Complexity* **22(4)**: pp. 732–748.

Warren, J. R. (2002): "Reconsidering the relationship between student employment and academic outcomes: A new theory and better data." *Youth & Society* **33(3)**: pp. 366–393.

Warren, J. R. & E. F. Cataldi (2006): "A historical perspective on high school students' paid employment and its association with high school dropout." *Sociological Forum* **21(1)**: pp. 113–143.

Warren, J. R. & J. C. Lee (2003): "The impact of adolescent employment on high school dropout: Differences by individual and labor-market characteristics." *Social Science Research* **32(1)**: pp. 98–128.

Warren, J. R., P. C. LePore, & R. D. Mare (2000): "Employment during high school: Consequences for students' grades in academic courses." *American Educational Research Journal* **37(4)**: pp. 943–969.

Wenz, M. & W.-C. Yu (2010): "Term-time employment and the academic performance of undergraduates." *Journal of Education Finance* **35(4)**: pp. 358–373.

Yanbarisova, D. (2015): "The effects of student employment on academic performance in Tatarstan higher education institutions." *Russian Education & Society* **57(6)**: pp. 459–482.

Zeugner, S. (2011): "Bayesian Model Averaging with BMS for BMS Version 0.3.0." Retrieved from https://cran.r-project.org/web/packages/BMS/vignettes/bms.pdf. Accessed: 23/10/2020.

Zeugner, S. & M. Feldkircher (2009): "Benchmark priors revisited: on adaptive shrinkage and the supermodel effect in bayesian model averaging." *IMF Working Paper WP/09/202*, International Monetary Fund, Finance Department.

Zigraiova, D. & T. Havranek (2016): "Bank competition and financial stability: Much ado about nothing?" *Journal of Economic Surveys* **30(5)**: pp. 944–981.

Zimmermann, C. (2013): "Academic Rankings with RePEc." *Econometrics* **1(3)**: pp. 249–280.

Appendix A

Additional Information for Data Collection

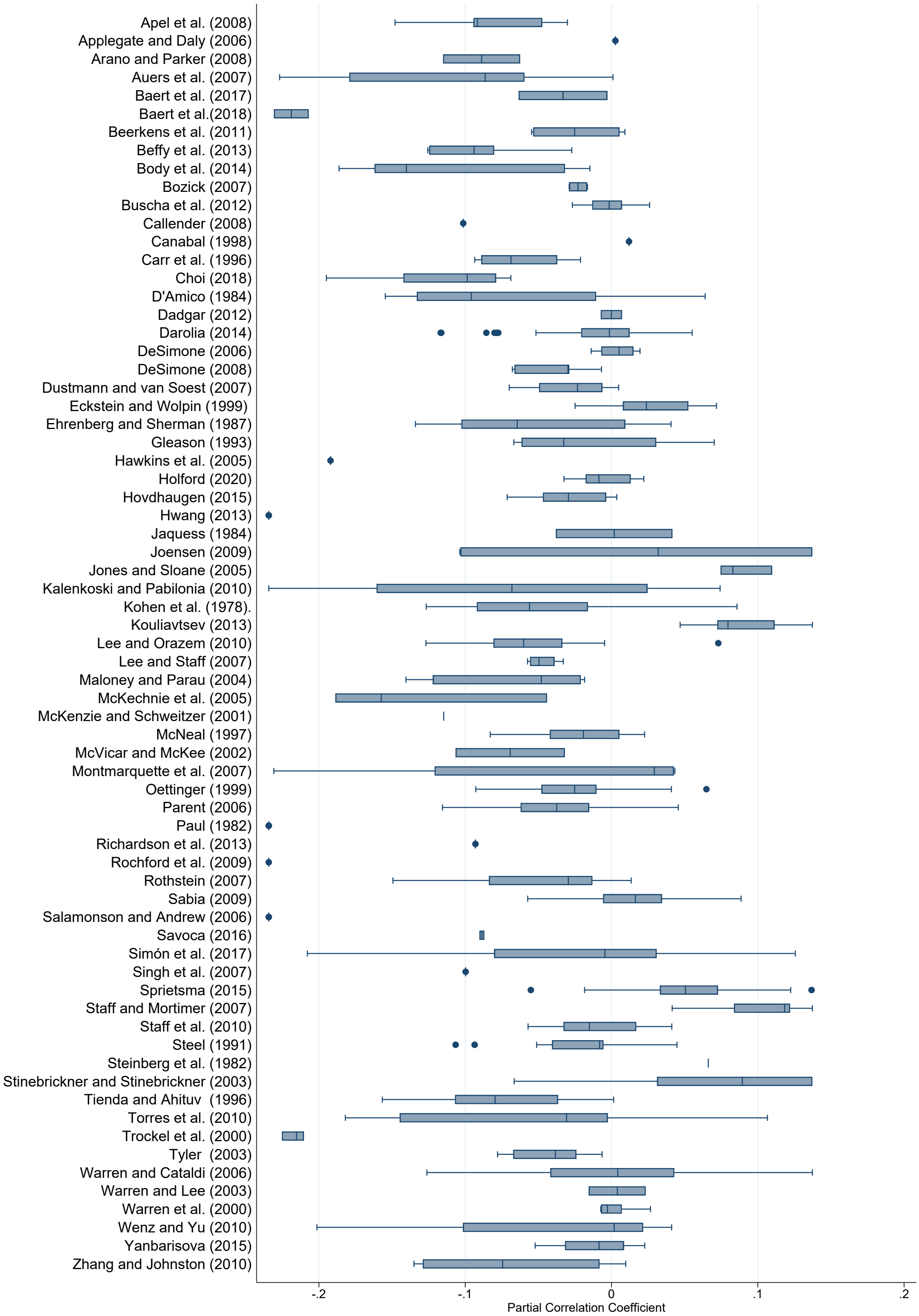
Table A.1: Attempted search queries and reasons for their exclusion

Search query	Reason for excluding the search query
Student work and grades	The search query returns studies focusing on students' grades but not necessarily in relation to students' employment.
Work status and academic outcomes	The search query yields results focusing on the effect of early academic outcomes on employment opportunities (reverse effect).
School employment and academic performance	The search query yields results focusing predominantly on secondary education students.
Youth employment and academic performance	The search combination generates studies with focus on child's laws and child labour legislation.
Effect of youth employment on academic outcomes	The search combination is too specific and generates a low amount of studies.

Table A.2: List of unavailable research papers

Birdwell & Excovitz, 1990: The Relationships between Student Employment during the Academic Year and Academic and NABPLEX Performance
Hammes & Haller, 1983: Making ends meet: Some of the consequences of part-time work for college students
Hobbs, 1993: Part-Time Employment and Schooling
Hood, Craig & Ferguson, 1992: The impact of athletics, part-time employment, and other activities on academic achievement
Howard, 1998: Does Part-Time Employment Affect A-Level Grades Achieved?
Lammers, Onweugbuzie & Slate, 2001: Academic success as a function of gender, class, age, study habits, and employment of college students. Research in the Schools.
Ma & Wooster, 1979: The Effect of Unemployment on the College Student's Academic Performance
Paton-Saltzburg & Lindsay, 1994: The effect of paid employment on the academic performance of full-time students in higher education
Santora, 1994: The relationship of part-time employment and locus-of-control to academic achievement among 12th-grade pupils in selected United States high schools in 1980

Figure A.1: Variation of effect estimates within and across studies



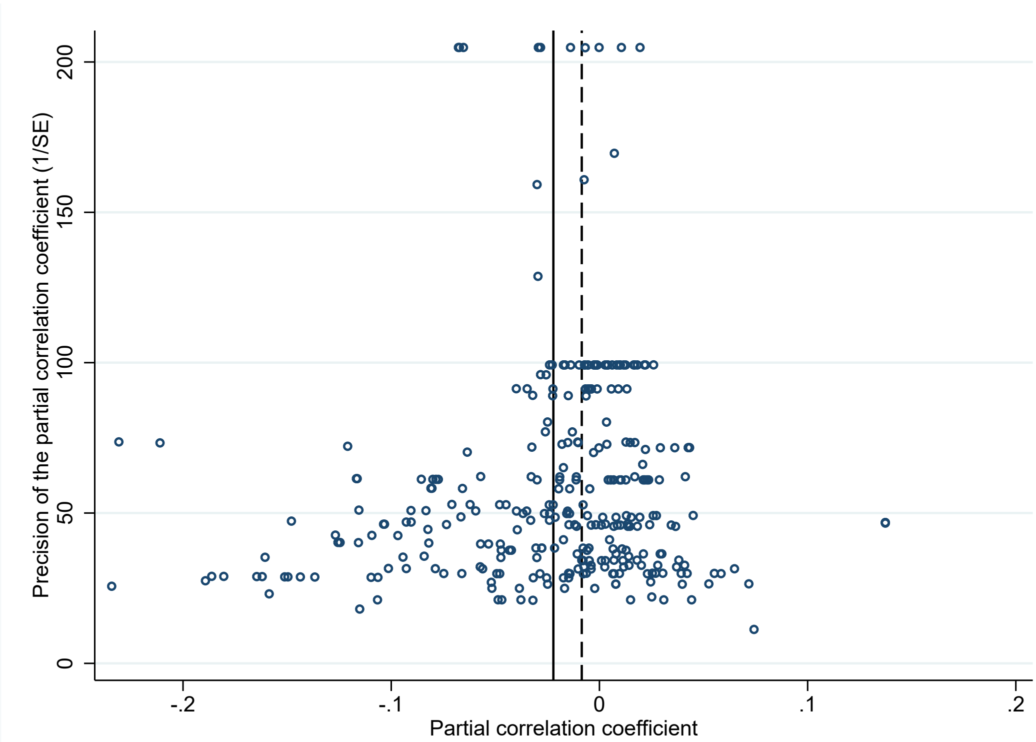
Notes: The figure shows a box plot of PCCs capturing the relationship between student employment and academic achievement across different studies.

Appendix B

Robustness Checks for Publication Bias Tests

Tests of publication bias for studies controlling for endogeneity

Figure B.1: Funnel plot for a subsample of studies controlling for endogeneity



Notes: The figure displays the funnel plot of estimates coming from studies carefully controlling for endogeneity. The solid vertical line represents the mean of PCCs, the dashed vertical line indicates the median.

Table B.1: Linear tests of publication bias for studies controlling for endogeneity

Panel A: Various model specifications				
	OLS	Between effects	Study	Instrument
Standard Error	-0.555	-1.372*	-1.076*	-0.585
(Publication Bias)	(0.57)	(0.57)	(0.51)	(0.58)
	[-2.102, 0.758]		[-2.260, 0.221]	[-2.318,0.780]
Constant	-0.010	-0.002	-0.009	-0.009
(Mean Beyond Bias)	(0.01)	(0.02)	(0.01)	(0.01)
	[-0.035, 0.012]		[-0.035, 0.016]	[-0.033, 0.012]
Observations	307	307	307	307
Studies	29	29	29	29
Panel B: Model specifications weighted by precision				
	WLS	Between effects	Study	Instrument
Standard Error	-0.637	-1.292*	-1.219	-0.659
(Publication Bias)	(0.72)	(0.60)	(0.75)	(0.73)
	[-2.526, 1.092]		[-2.897, 0.017]	[-2.206, 0.876]
Constant	-0.008	-0.004	-0.005	-0.008
(Mean Beyond Bias)	(0.01)	(0.01)	(0.01)	(0.01)
	[-0.039, 0.019]		[-0.044, 0.017]	[-0.034, 0.017]
Observations	307	307	307	307
Studies	29	29	29	29

Notes: The table reports the results of linear regression testing the presence of publication bias among estimates of studies carefully controlling for endogeneity bias. The simple uncorrected mean equals -0.022, the weighted uncorrected mean - 0.035. The standard errors of the regression parameters are clustered at the study level. In panel A we present the following specifications: OLS = ordinary least squares, BE = study-level between effects, Study = weighted by the inverse of the number of estimates reported per study, IV = the inverse of the square root of the number of observations acts as an instrument for the standard error. In Panel B, the same specifications are additionally weighted by the inverse of of PCC's standard errors. Standard errors in parentheses. 95% confidence intervals from wild bootstrap clustering in brackets.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2: Non-linear tests of publication bias for studies controlling for endogeneity

	Stem-based method	Endogenous kink	WAAP	Selection model
Effect Beyond Bias	0.004	-0.013***	0.004	0.007
	(0.022)	(0.004)	(0.007)	(0.003)
Observations	307	307	307	307
Studies	29	29	29	29

Notes: The table reports the results of non-linear tests, showing the magnitude and significance of the true underlying effect corrected for publication bias. The simple uncorrected mean equals -0.022, the weighted uncorrected mean - 0.035. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

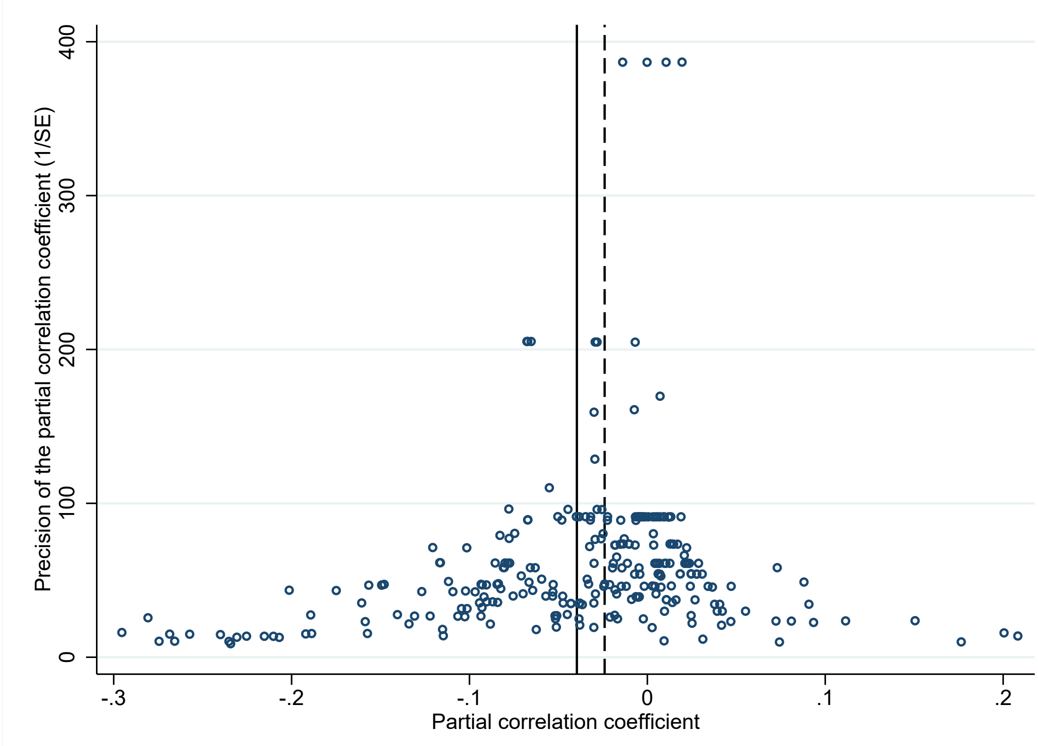
Table B.3: Caliper test for detecting publication bias for studies controlling for endogeneity

Critical threshold of t-statistic	-1.96	0	1.96
Caliper size: 10%			
Share of estimates above the threshold	0.563***	0.455**	0.600*
	(0.13)	(0.16)	(0.24)
Caliper size: 5%			
Share of estimates above the threshold	0.875***	0.667	0.500
	(0.13)	(0.33)	(0.50)
Observations	307	307	307
Studies	29	29	29

Notes: The table reports the share of estimates being above the critical value of t-statistic (in absolute terms) in a 10% and 5% caliper. To illustrate the interpretation of the coefficients, a coefficient of 0.875 means that the ratio of negative significant estimates to non-significant is 87.5% to 12.5%. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tests of publication bias for studies specifying student employment as a continuous variable

Figure B.2: Funnel plot for a subsample of studies specifying student employment as a continuous variable



Notes: The figure displays the funnel plot of partial correlation coefficients capturing the effect between student employment and academic achievement in studies specifying student employment as a continuous variable. The solid vertical line represents the mean of PCCs, the dashed vertical line indicates the median. We use unwinsorized data for constructing this diagram. For quantitative statistical tests we employ winsorized dataset.

Table B.4: Linear tests of publication bias for a subsample specifying student employment as a continuous variable

Panel A: Various model specifications	OLS	Between effects	Study	Instrument
Standard Error	-1.323	-2.480***	-2.219***	-1.376
(Publication Bias)	(0.71)	(0.42)	(0.36)	(0.71)
	[-2.724, 0.561]		[-2.895, -1.356]	[-2.720, 0.518]
Constant	-0.005	0.015	0.005	-0.003
(Mean Beyond Bias)	(0.01)	(0.02)	(0.01)	(0.02)
	[-0.046, 0.028]		[-0.018, 0.026]	[-0.047, 0.029]
Observations	261	261	261	261
Studies	39	39	39	39
Panel B: Model specifications weighted by precision	**WLS**	**Between effects**	**Study**	**Instrument**
Standard Error	-1.233*	-2.281***	-2.248***	-1.265*
(Publication Bias)	(0.54)	(0.45)	(0.40)	(0.55)
	[-2.393, 0.049]		[-3.036, -1.216]	[-2.340, -0.110]
Constant	-0.007	0.004	0.003	-0.007
(Mean Beyond Bias)	(0.01)	(0.01)	(0.01)	(0.01)
	[-0.036, 0.015]		[-0.037, 0.014]	[-0.027, 0.013]
Observations	261	261	261	261
Studies	39	39	39	39

Notes: The table reports the results of linear regression testing the presence of publication bias for a subsample of studies specifying student employment as a continuous variable. The simple uncorrected mean equals -0.039, the weighted uncorrected mean -0.080. The standard errors of the regression parameters are clustered at the study level. In panel A we present the following specifications: OLS = ordinary least squares, BE = study-level between effects, Study = weighted by the inverse of the number of estimates reported per study, IV = the inverse of the square root of the number of observations acts as an instrument for the standard error. In Panel B, the same specifications are additionally weighted by the inverse of of PCC's standard errors. Standard errors in parentheses. 95% confidence intervals from wild bootstrap clustering in brackets.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Non-linear tests of publication bias for a subsample specifying student employment as a continuous variable

	Stem-based method	Endogenous kink	WAAP	Selection model
Effect Beyond Bias	0.005	-0.005	-0.022	-0.041***
	(0.025)	(0.003)	(0.010)	(0.005)
Observations	261	261	261	261
Studies	39	39	39	39

Notes: The table reports the results of non-linear tests, showing the magnitude and significance of the true underlying effect corrected for publication bias for the subsample of studies specifying student employment as a continuous variable. The simple uncorrected mean equals -0.039, the weighted uncorrected mean -0.080. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Caliper test for detecting publication bias for a subsample of studies specifying student employment as a continuous variable

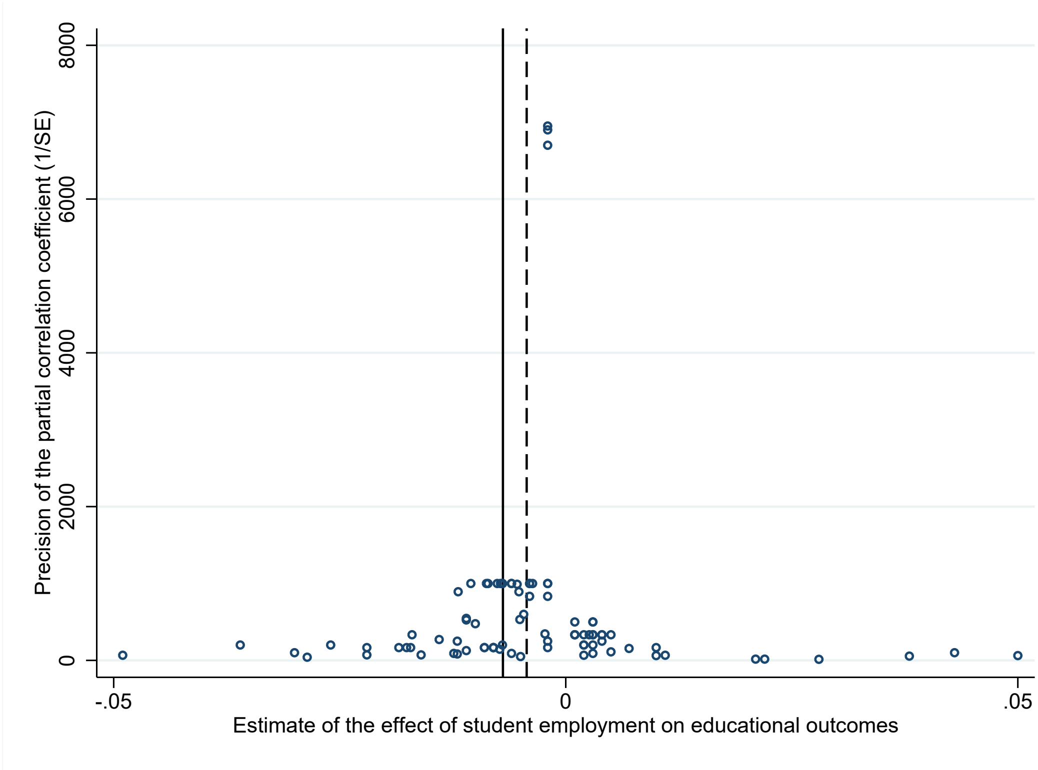
Threshold for t-statistic	-1.96	0	1.96
Caliper size: 10%			
Share of estimates above the threshold	0.700***	0.333	0.500
	(0.16)	(0.21)	(0.50)
Caliper size: 5%			
Share of estimates above the threshold	0.714***	NA	NA
	(0.18)	(NA)	(NA)
Observations	261	261	261
Studies	39	39	39

Notes: The table reports the share of estimates being above the critical value of t-statistics (in absolute terms) in a 10% and 5% caliper. To illustrate the interpretation of the coefficients, a coefficient of 0.714 means that the ratio of negative significant estimates to non-significant is approximately 71% to 29%. NA means that that there is insufficient amount of observations to calculate the statistics. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Tests of publication bias for the subsample of original estimates, not transformed into PCCs

Note: We do not report results of Caliper test for the untransformed sample due to insufficient amount of observations.

Figure B.3: Funnel plot for the untransformed subsample



Notes: The figure displays the funnel plot of original, non-transformed estimates capturing the effect between student employment and academic achievement. The solid vertical line represents the mean of PCCs, the dashed vertical line indicates the median.

Table B.7: Linear tests of publication bias for the untransformed subsample

Panel A: Various model specifications	OLS	Between effects	Study	Instrument
Standard Error	-1.090	-2.015***	-1.887***	0.299
(Publication Bias)	(0.83)	(0.00)	(0.32)	(1.35)
	[-2.266, 1.185]		[-3.120, 1.136]	[WB unfeasible]
Constant	0.006	0.009	0.007	-0.011
(Mean Beyond Bias)	(0.01)	(0.01)	(0.01)	(0.01)
	[-0.008, 0.027]		[-0.003, 0.017]	[-0.039, 0.015]
Observations	92	92	92	92
Studies	16	16	16	16
Panel B: Model specifications weighted by precision				
	WLS	Between effects	Study	Instrument
Standard Error	-0.293*	-1.020	-1.363**	-0.305
(Publication Bias)	(0.71)	(0.10)	(0.50)	(0.77)
	[-2.056, 1.077]		[-2.461, -0.169]	[-1.818, 1.648]
Constant	-0.003**	-0.002***	-0.002***	-0.003**
(Mean Beyond Bias)	(0.00)	(0.00)	(0.00)	(0.00)
	[-0.009, 0.003]		[-0.007, 0.008]	[-0.016, 0.000]
Observations	92	92	92	92
Studies	16	16	16	16

Notes: The table reports the results of linear regression testing the presence of publication bias among original, non-transformed estimates.The simple uncorrected mean equals -0.006, the weighted uncorrected mean -0.029. The standard errors of the regression parameters are clustered at the study level. In panel A we present the following specifications: OLS = ordinary least squares, BE = study-level between effects, Study = weighted by the inverse of the number of estimates reported per study, IV = the inverse of the square root of the number of observations acts as an instrument for the standard error. In Panel B, the same specifications are additionally weighted by the inverse of of PCC's standard errors. Standard errors in parentheses. 95% confidence intervals from wild bootstrap clustering in brackets.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Non-linear tests of publication bias for an untransformed homogenous subsample

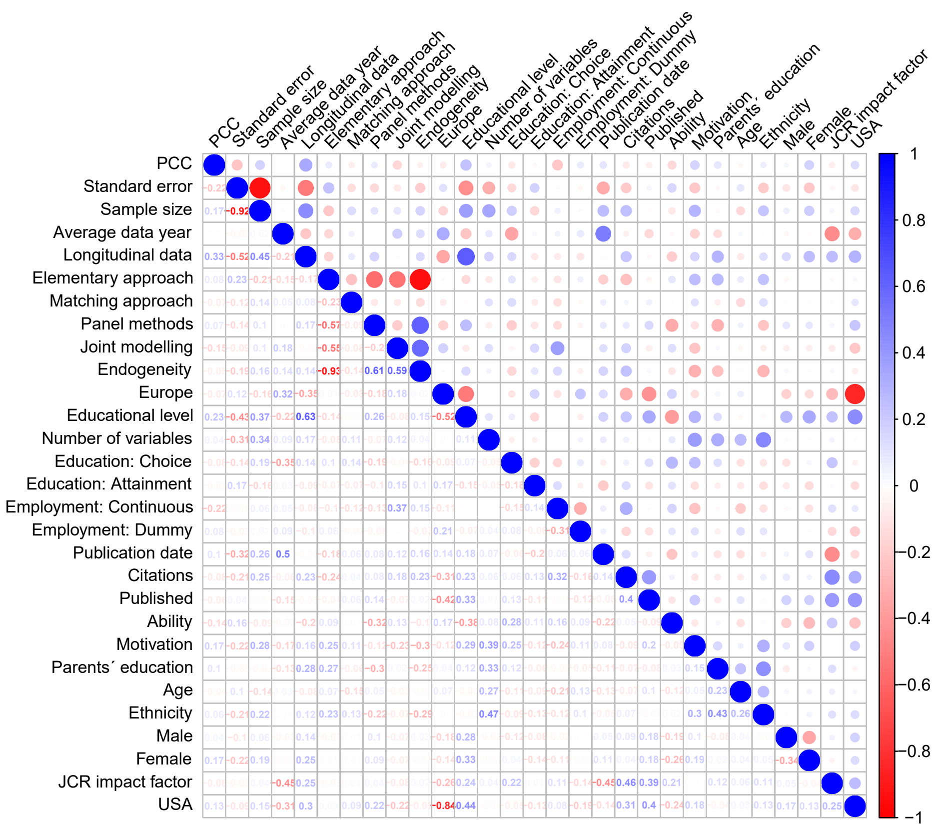
	Stem-based method	Endogenous kink	WAAP	Selection model
Effect Beyond Bias	NA	-0.002	-0.002	-0.000
	(NA)	(0.000)	(0.000)	(NA)
Observations	92	92	92	92
Studies	16	16	16	16

Notes: The table reports the results of non-linear tests, showing the magnitude and significance of the true underlying effect corrected for publication bias for for a subsample of original, untransformed estimates. Standard errors in parentheses. The simple uncorrected mean equals -0.006, the weighted uncorrected mean -0.029. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C

BMA Diagnostics and BMA Robustness Checks

Figure C.1: Correlations between additional variables collected to study heterogeneity among effect estimates



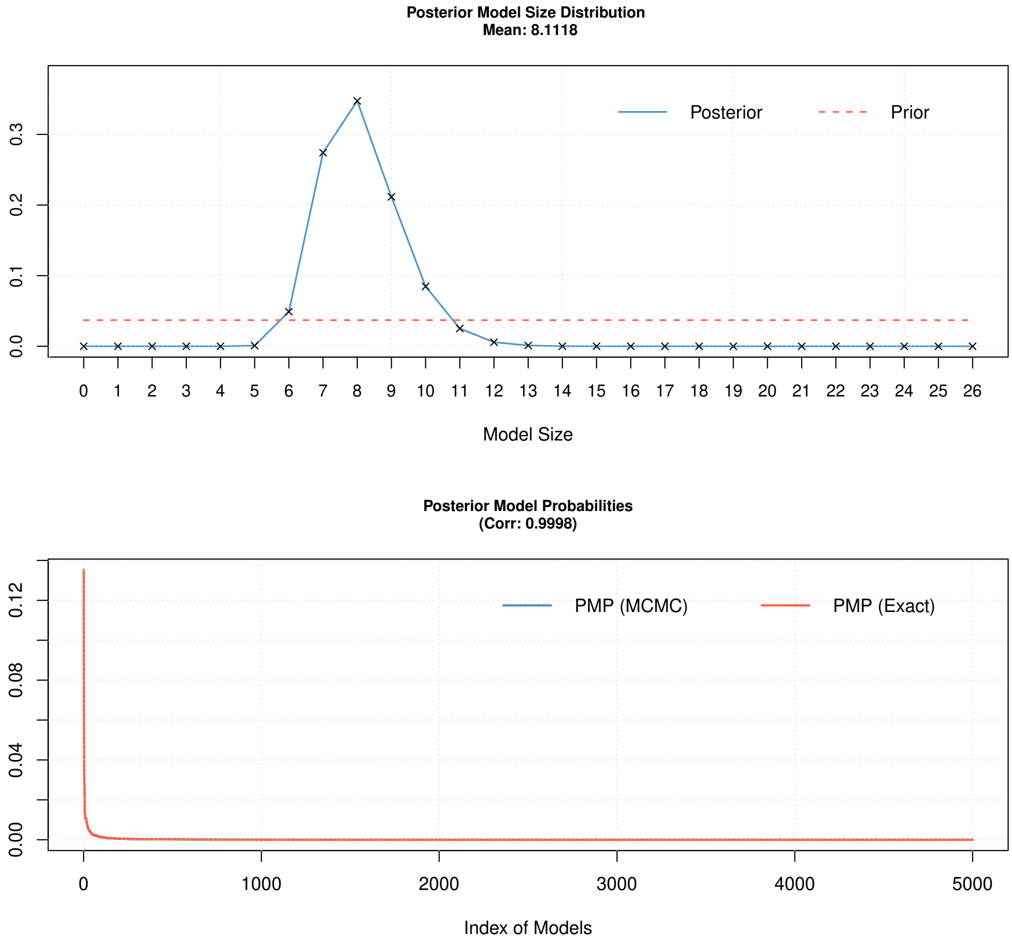
Notes: The figure displays correlation coefficients for variables collected to study heterogeneity in PCCs capturing the effect of student employment on educational outcomes. Due to high correlations we omit *Sample size* and *Endogeneity* from the BMA analysis. We control for the high correlation between *USA* and *Europe* with our choice of model prior.

Table C.1: Summary of BMA estimation for the baseline model

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
8.1118	$2 \cdot 10^6$	$1\ 10^6$	4.362931 mins
Models visited	*Modelspace*	*Models visited*	*Topmodels*
407,048	$6.7 \cdot 10^7$	0.61%	97%
Corr PMP	*No. obs.*	*Model prior*	*g-Prior*
0.9998	861	Random/13	UIP
Shrinkage-Stats			
Av = 0.9988			

Notes: We use the unit information prior (the prior has the same weight as one observation) suggested by Eicher *et al.* (2011) and the dilution prior (accounting for potential collinearity) suggested by George *et al.* (2010). *Draws* is set to 2 million iterations. *Burn-ins* are equal to 1 million and represent the number of iterations that are not stored to compute posterior probabilities. The results of this BMA estimation are reported in Table 6.2.

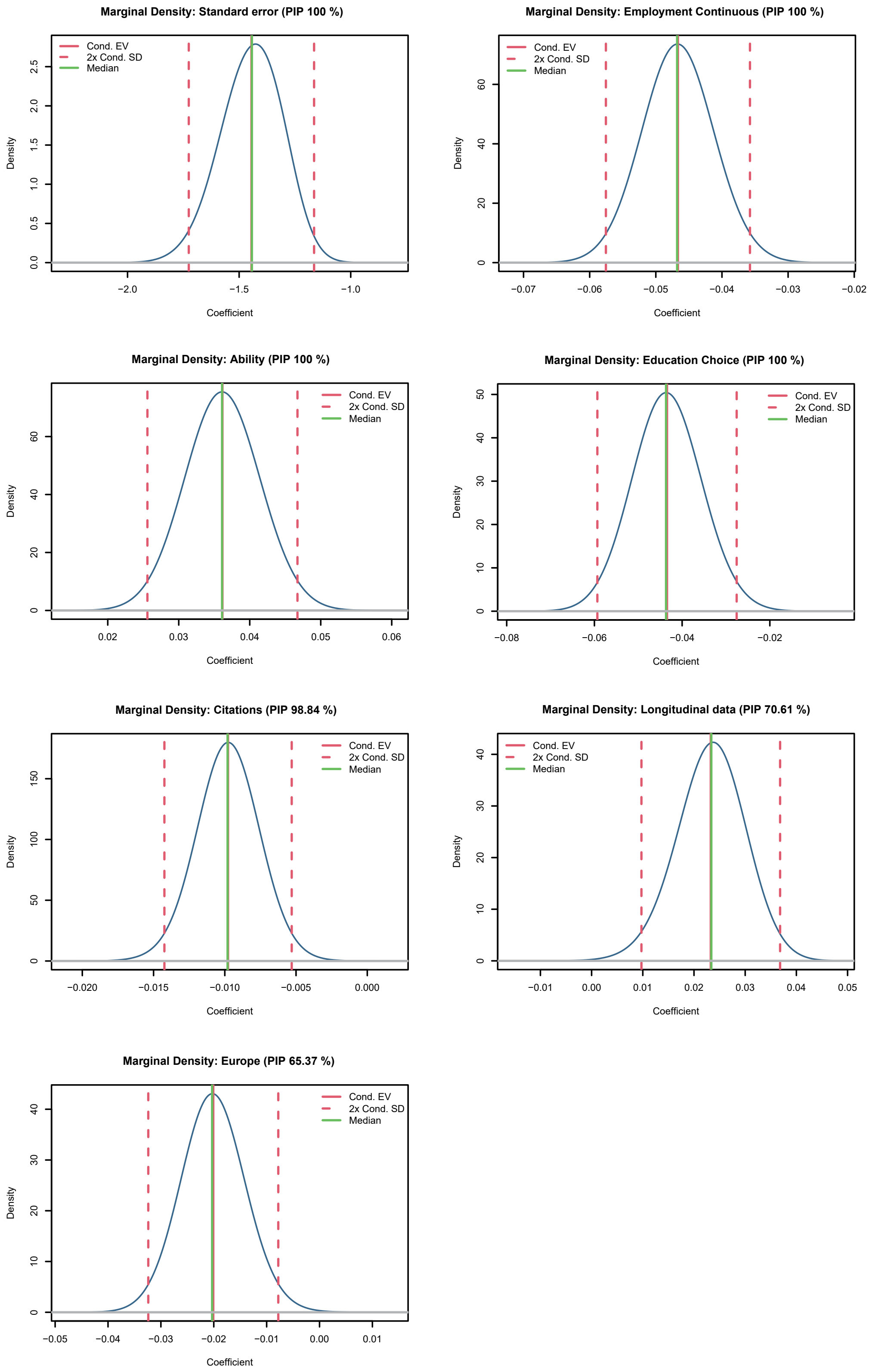
Figure C.2: Model size and convergence for the baseline BMA model



Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA estimation reported in Table 6.2.

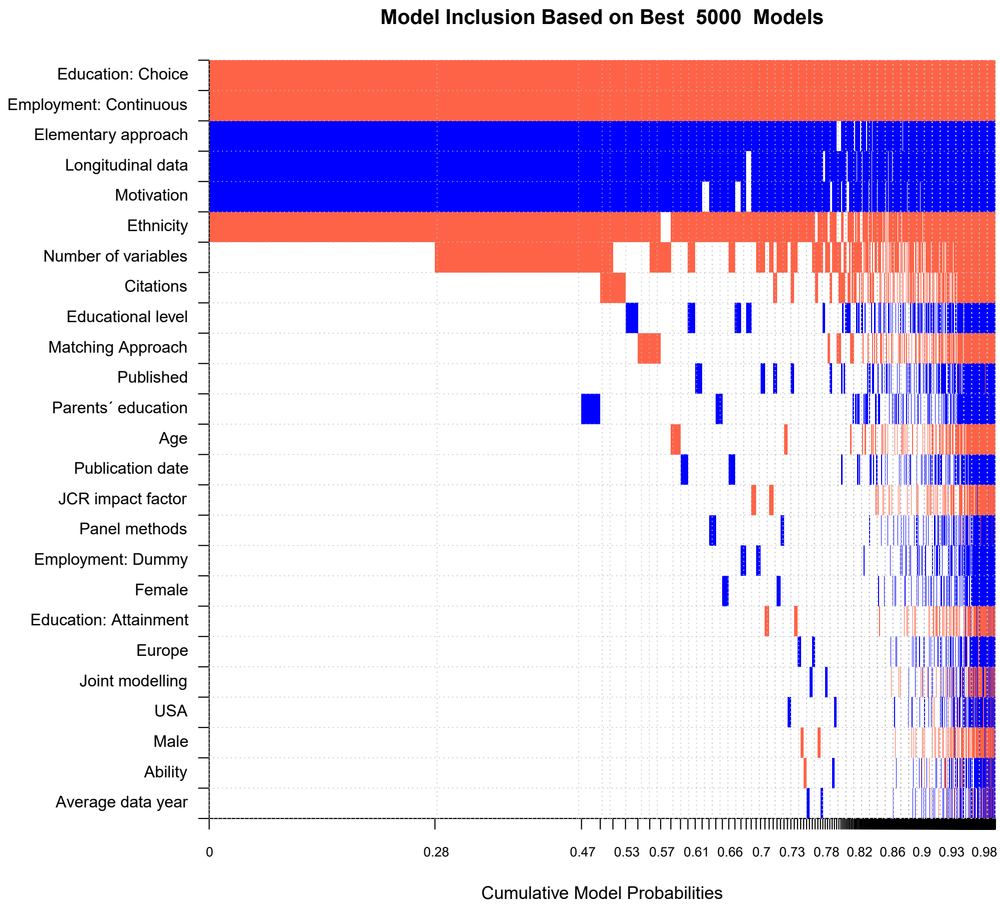
ycentering

Figure C.3: Posterior coefficient distributions for important variables from the baseline BMA model



Notes: The figure depicts the marginal densities of the posterior coefficient distributions of the important variables for the baseline BMA estimation. For instance, we see that the coefficient of *Education: Choice* is positive in all models.

Figure C.4: Model inclusion for our baseline BMA estimation weighted by the precision of estimates



Notes: The figure shows the results of BMA estimation (g-prior = unit information prior, model prior = dilution prior) weighted by the precision (1/SE) of estimates. The vertical axis plots the explanatory variables according to their posterior inclusion probabilities in descending order. The horizontal axis depicts the posterior model probability. White-colored cells indicate that variables is not included in the model. Blue-colored (darker in the greyscale) cells imply that the estimated coefficient of variable is positive, while red-colored cells (lighter in the greyscale) indicate negative estimated coefficient of a variable.

Table C.2: Explaining heterogeneity using a subsample specifying student employment as a continuous variable

	BMA (weighted by No. Obs.)			BMA (weighted by SE)		
	Post. Mean	Post. SD	PIP	Post. Mean	Post. SD	PIP
Standard error	-2.082	0.201	1.000	NA	NA	NA
Data characteristics						
Average data year	-0.002	0.005	0.138	0.000	0.004	0.011
Longitudinal data	0.001	0.004	0.026	0.045	0.050	0.514
Estimation methods						
Elementary approach	0.000	0.001	0.014	0.000	0.002	0.009
Panel methods	0.000	0.004	0.017	0.000	0.003	0.008
Instrumental variable	0.000	0.002	0.014	0.000	0.002	0.009
Design of the analysis						
Europe	-0.001	0.005	0.036	0.000	0.004	0.008
USA	0.000	0.002	0.016	0.000	0.003	0.009
Educational level	0.000	0.003	0.028	0.004	0.016	0.078
Number of variables	-0.001	0.003	0.070	-0.001	0.006	0.053
Variable specification						
Education: Choice	0.000	0.004	0.025	-0.002	0.014	0.038
Education: Attainment	-0.001	0.004	0.032	0.000	0.003	0.008
Publication characteristics						
Publication date	-0.002	0.005	0.162	0.000	0.005	0.018
Citations	-0.011	0.005	0.890	-0.001	0.005	0.060
Published	0.000	0.002	0.016	0.031	0.047	0.329
JCR impact factor	0.000	0.002	0.061	0.000	0.001	0.008
Student characteristics						
Ability	0.014	0.016	0.467	0.000	0.003	0.010
Motivation	0.000	0.001	0.013	0.000	0.004	0.011
Parental education	-0.001	0.005	0.048	-0.008	0.022	0.147
Age	0.000	0.002	0.016	-0.001	0.007	0.029
Ethnicity	0.000	0.001	0.015	0.000	0.004	0.016
Male	0.000	0.004	0.026	0.000	0.002	0.009
Female	0.000	0.002	0.012	0.001	0.006	0.020
Intercept	0.011	NA	1.000	-0.398	NA	1.000
Studies	39			39		
Observations	261			261		

Notes: Post. Mean = Posterior Mean. Post. SD = Posterior standard deviation. PIP = Posterior inclusion probability. BMA estimation is conducted using the unit information prior (the prior has the same weight as one observation) and the dilution prior (accounting for potential collinearity). Compared to the whole sample, this subsample does not include variables: Matching approach, Joint modelling, Employment: Continuous and Employment: Dummy. For explanation, see Section 6.4.

Table C.3: Explaining heterogeneity using a subsample specifying student employment as a categorical variable

	BMA (weighted by No. Obs.)			BMA (weighted by SE)		
	Post. Mean	Post. SD	PIP	Post. Mean	Post. SD	PIP
Standard error	-0.052	0.157	0.128	NA	NA	NA
Data characteristics						
Average data year	-0.008	0.009	0.551	0.000	0.002	0.022
Longitudinal data	0.000	0.004	0.048	0.003	0.012	0.065
Estimation methods						
Elementary approach	0.011	0.023	0.292	0.064	0.052	0.823
Matching approach	0.001	0.024	0.231	0.011	0.037	0.178
Panel methods	0.007	0.022	0.111	0.024	0.044	0.293
Joint modelling	-0.064	0.020	0.956	-0.082	0.071	0.637
Design of the analysis						
Europe	0.003	0.009	0.096	0.000	0.007	0.027
USA	-0.005	0.014	0.132	0.144	0.026	1.000
Educational level	0.037	0.011	0.991	-0.001	0.004	0.069
Number variables	0.000	0.001	0.036	-0.041	0.034	0.666
Variable specification						
Education: Choice	-0.009	0.013	0.353	0.026	0.038	0.362
Education: Attainment	0.001	0.004	0.077	0.000	0.003	0.046
Low intensity	0.047	0.007	1.000	0.030	0.036	0.477
Medium intensity	0.001	0.005	0.084	0.016	0.024	0.366
Publication characteristics						
Publication date	0.002	0.004	0.239	0.001	0.004	0.039
Citations	0.002	0.004	0.169	-0.171	0.036	0.999
Published	-0.070	0.015	1.000	0.079	0.014	1.000
JCR impact factor	-0.002	0.003	0.273	0.014	0.021	0.354
Student characteristics						
Ability	0.034	0.012	0.998	0.006	0.016	0.170
Motivation	0.003	0.007	0.166	0.000	0.004	0.034
Parental education	0.000	0.002	0.032	-0.017	0.024	0.380
Age	-0.006	0.011	0.293	-0.001	0.005	0.040
Ethnicity	0.000	0.002	0.040	0.000	0.003	0.021
Male	0.000	0.003	0.028	-0.013	0.013	0.552
Female	0.000	0.002	0.024	0.003	0.013	0.073
Intercept	0.003	NA	1.000	-0.068	NA	1.000
Studies	29			29		
Observations	442			442		

Notes: Post. Mean = Posterior Mean. Post. SD = Posterior standard deviation. PIP = Posterior inclusion probability. BMA estimation is conducted using the unit information prior (the prior has the same weight as one observation) and the dilution prior (accounting for potential collinearity). Compared to the whole sample, this subsample does not include variables: Employment: Continuous and Employment: Dummy. In contrast, it includes additional variables *Low intensity* and *Medium intensity*. For explanation, see Section 6.4.