CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Master's thesis

2022 **Tereza Navarová**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Tereza Navarová

International gender wage gap: A Meta-analysis

Master's thesis

Prague 2022

Declaration of Authorship

I hereby declare that I compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

I grant to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agree with the thesis being used for study and scientific purposes.

Prague, January 3, 2022

Tereza Navarová

Acknowledgments

I would like to thank my supervisor, prof. PhDr. Tomáš Havránek Ph.D., for his time, guidance and valuable suggestions for this thesis. Most importantly, I wish to thank my family and friends for their support and encouragement throughout my whole studies. 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. 870245.

Bibliographic note

NAVAROVÁ, Tereza. *International gender wage gap: A Meta-Analysis.* 52 p. Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor prof. PhDr. Tomáš Havránek Ph.D.

Abstract

This thesis uses modern meta-analysis methods to produce a systematic quantitative review of the literature on the international gender wage gap. After collecting 661 estimates from 51 peer-reviewed studies, estimates from published studies are subjected to several tests studying the presence of publication bias. To test for publication bias, standard errors of estimates are approximated for studies that do not report standard errors or other confidence measures. Based on both linear and non-linear formal tests, the presence of significant publication bias is not detected. Furthermore, Bayesian model averaging is performed to explain the heterogeneity in the collected data, with robust control performed using frequentist model averaging. The results suggest that omitting variables related to human capital leads to the estimation of higher average gender wage differences. In contrast, authors who focus subsamples of the labour market (such as college graduates or only fulltime workers) estimate on average smaller gender wage differences.

JEL Classification J31, J71, J70, J08
Keywords meta-analysis, gender, gender pay gap, wage differentials, publication bias, model averaging

Author's e-mail navarova.tereza@gmail.com
Supervisor's e-mail tomas.havranek@ies-prague.org

Abstrakt

Tato diplomová práce využívá moderních metod meta-analýzy k vytvoření systematického kvantitativního přehledu literatury o mezinárodních rozdílech v odměňování žen a mužů. Po shromáždění 661 odhadů z 51 odborně recenzovaných studií jsou odhady podrobeny několika testům zkoumajícím přítomnost publikační selektivity. Pro testování přítomnosti publikační selektivity aproximuji standardní chyby odhadů u studií, kde autor neuvádí standardní chyby nebo jiné míry spolehlivosti. Na základě lineárních i nelineárních formálních testů nebyla zjištěna přítomnost významné publikační selektivity. Dále je provedeno bayesiánské průměrování modelů pro vysvětlení heterogenity shromážděných dat, přičemž robustní kontrola je provedena pomocí frekventistického průměrování modelů. Výsledky naznačují, že vynechání proměnných souvisejících s lidským kapitálem vede k odhadu vyšších průměrných mzdových rozdílů mezi muži a ženami. Naopak autoři, kteří se zaměřují na určité podskupiny trhu práce (jako jsou čerství absolventi či jen pracující na plný úvazek), odhadují v průměru menší rozdíly ve mzdách mužů a žen.

Klasifikace JEL	J31, J71, J70, J08
Klíčová slova	meta-analýza, gender, rozdíly v odměňování žen a mužů, mzdové rozdíly, publikační selektivita, průměrování modelů

E-mail autora	navarova.tereza@gmail.com
E-mail vedoucího práce	tomas.havranek@ies-prague.org

Master’s Thesis Proposal

Author	Tereza Navarová
Supervisor	prof. PhDr. Tomáš Havránek Ph.D.
Proposed topic	International gender wage gap: Meta-analysis

Motivation The gender wage gap can be viewed from different perspectives. The one that we read about in newspaper is the non-adjusted wage gap that does not account for differences in occupations, job expertise, or education. One must not view this non-adjusted wage gap as a discrimination measure but rather a comparison of average pay across all occupations, ages, etc. for men and women. A much more precise is the adjusted wage gap that considers all sorts of characteristics of male and female individuals and compares the wages adequately, therefore the adjusted average female salary would be closer to the male salary than its non-adjusted equivalent.

Many researchers have used different methods of quantitatively explaining the gender pay gap (i.e. quantile regression, decomposition technique (Jann, 2008), etc.) A drawback of many papers is that they only cover data from single countries, specific narrow groups or for limited time periods. Many of them also do not use appropriate econometric methods, or they are purely narrative. The aim of this thesis is to collect as many relevant studies as possible that use appropriate econometric models and then to conduct a meta-analysis on the estimates from the collected papers. There are a few meta-analyses on the topic of gender wage differentials (Weichselbaumer and Winter-Ebmer 2005; Stanley and Jarrell 1998) but the literature lacks an up-to-date research using modern techniques and latest data as the meta-analyses are already more than 15 years old.

The method of meta-analysis has been widely used in economics since 1989 (Stanley and Jarrell 1998). Simply put, it is a summary of collected estimates and the subsequent quantitative analysis (Stanley 2005).

Hypotheses

Hypothesis #1: Male researchers report significantly larger estimates of the wage gap.

Hypothesis #2: Papers publishes after the year 2010 report smaller estimates of the wage gap.

Hypothesis #3: The literature estimating the gender wage gap suffers from publication bias.

Methodology The collection of primary studies is a crucial part of a meta-analysis. Since there has been a meta-analysis on this topic in 2005, I will examine its used literature and studies and then I will search the Google Scholar database for most recent studies in the field. In order to employ the meta-analysis, I will only select studies that report standard errors of all estimates. This will also be useful for correction for publication bias (Stanley, 2005). Those who do not satisfy this criterion will be excluded from the sample. After this primary search I will decide which variables to use in my analysis and I will clean the data.

My analysis will be consisting of examining publication bias. This bias is a phenomenon that arises when researchers, for example, decide not to publish their paper due to the insignificance of its estimated values, or in case their results are contradictory to what is generally expected (Rosenthal, 1979). Sometimes what happens is that the researchers use a different specification to achieve the result they were hoping for in the first place. Correcting for publication bias is my first task.

My second task is a correct estimation of my model. A useful method in meta-analysis is Bayesian model averaging (BMA). Since having a large number of variables may be tricky in a model, model averaging allows to assign weights to individually specified models given their data fit and specification (Steel, 2019).

Expected Contribution This thesis should serve as a quantitative overview of research on the gender wage gap throughout years and countries, including most recent studies. An emphasis is placed on the selection of papers to be included in the analysis and correction for publication bias or also other potential bias I will encounter during the work. The results should capture how certain restrictions of datasets, the choice of econometric methods, or other characteristics affect the resulting gender wage gap. I expect to offer a comprehensive summary of gender wage gap studies presenting a genuine and objective decomposition of wage differentials between men and women.

Outline

1. Introduction: I will shortly present my motivation and the shortcomings and findings of existing meta-analyses on the topic.

2. Literature Review: I will provide an overview of different studies and their respective methods and model choices.
3. Data: The data selection is the key point in conducting a meta-analysis. I will describe how I collected my data, what restrictions I chose and why.
4. Methodology: I will explain the modern methods used in meta-analysis, including correction for publication bias.
5. Results: I will discuss my findings and the robustness of my results.
6. CConclusion: I will summarize my work by comparing my findings to the findings of other researchers. I may also suggest further extension of my work for future research.

Core bibliography

Barth, E., & Dale-Olsen, H. (2009). Monopsonistic discrimination, worker turnover, and the gender wage gap. Labour Economics, 16.

Daly, A., Meng, X., Kawaguchi, A., & Mumford, K. (2006). The gender wage gap in four countries. Economic Record, 82.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal, 8.

Menon, N., & Van der Meulen Rodgers, Y. (2009). International trade and the gender wage gap: New evidence from India's manufacturing sector. World Development, 37.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological bulletin, 86, 638.

Stanley, T. D. (2005). Beyond publication bias. Journal of economic surveys, 19, 309-345.

Stanley, T. D., & Jarrell, S. B. (1989). Meta-regression analysis: a quantitative method of literature surveys. Journal of economic surveys, 19.

Stanley, T. D., & Jarrell, S. B. (1998). Gender wage discrimination bias? A meta-regression analysis. Journal of Human Resources.

Steel, M. (2019). Model Averaging and its Use in Economics.

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A meta-analysis of the international gender wage gap. Journal of Economic Surveys, 19.

Contents

List of Tables

List of Figures

Acronyms

BE Between Effects

BMA Bayesian Model Averaging

BO Blinder-Oaxaca

EU European Union

FAT Funnel Asymmetry Test

FE Fixed Effects

FMA Frequentist Model Averaging

GWG Gender Wage Gap

IID Indenpendent and Identically Distributed

IV Instrumental Variable

MRA Meta-Regression Analysis

OECD Organisation for Economic Co-operation and Development

OLS Ordinary Least Squares

PET Precision Effect Test

PIP Posterior Inclusion Probability

PMP Posterior Model Probability

RE Random Effects

US United States

WAAP Weighted Average of Adequately Powered

WLS Weighted Least Squares

Chapter 1

Introduction

The gender differentials in the labour market have been at the center of attention of many policymakers and researchers for decades. Yet, to this day, despite all the steps taken towards equal employment opportunities for men and women, we still face a persistent gender pay gap in developed countries. For several decades, the gender wage differentials have been declining, women have been catching up with men in education and experience, as well as in the level of wages (Goldin 2014). Dating back to the 1960s, the hourly earnings of women reached only the 60% level of men's earnings. When now, the differences have been diminishing not only in wages but also in the attitude towards women at the workplace (Mandel and Semyonov 2014) or the level of attained education and experience (Fortin et al. 2017), leading to a 11.4%[1] adjusted gender pay gap in the year 2019 for EU countries.

Empirical research usually shows a comprehensive analysis of the gender wage gap situation in a chosen country or region, often with different methodological approaches and various options for control variables. Given the importance of the topic, many studies have been done to analyze the development of gender wage differentials of today's society. Also, many studies managed to condense the findings of empirical literature into a narrative review. However, there exist only a few that managed to review the existing literature quantitatively and systematically (Jarrell and Stanley 2004; Stanley and Jarrell 1998; Weichselbaumer and Winter-Ebmer 2005; Konstantinova 2020). This thesis aims to collect and analyze estimates of the gender wage gap from exclusively peer-reviewed studies published in 2003 and later. But not only does this thesis cover the explanation of heterogeneity behind different gender wage gap esti-

[1]Eurostat: The gender pay gap situation in the EU (2019)

mates, it also applies statistical methods to identify potential publication bias which, to the best knowledge of the author, has never been done before.

I collect 661 estimates of the adjusted gender wage gap from 51 studies published between 2003 and 2021. This way, I only aim my attention at the newest findings regarding the wage differential between men and women. The oldest estimate from my dataset dates back to the year 1990 which is the year of the newest estimate in the meta-analysis by Weichselbaumer and Winter-Ebmer (2005).

The collected estimates are subjected to several statistical tests to detect publication bias in the published results of the existing literature. As the techniques of publication bias detection have been introduced relatively recently, to my best knowledge, an analysis of this nature has never been executed on a sample of gender wage gap estimates before. As opposed to the authors of precedent meta-analyses on this topic, this paper manages to deal with the fact that many authors do not report confidence measures and uses weighting techniques to explain heterogeneity in the data to an apparently unbalanced dataset. Unfortunately, due to the nature of reported pay gap decomposition results, only 135 estimates are collected with a respective standard error or other form of uncertainty measure. To tackle this issue in my analysis, the remaining standard errors are approximated to allow extensive testing of the potential presence of publication bias. The visual testing of the presence of publication bias (via funnel plots) is followed by more formal statistical tests including both classical linear and non-linear methods. The applied methods include a baseline OLS followed by weighted least squares to account for heteroskedasticity, as well as using an instrumental variable estimation. Alternatively, the within- and between-study variation is also addressed. Non-linear methods in this analysis address the possible issue with non-linear relationships of the precision measure and the collected estimates. In my thesis, I employ a set of recently introduced non-linear techniques including the stem-based method (Furukawa 2019), the weighted average of adequately powered (WAAP) estimates technique (Ioannidis et al. 2017) and the endogenous kink model (Bom and Rachinger 2019). The application of all of the above mentioned techniques applies both to the whole sample as well as to the restricted sample including only estimates with original standard errors.

While the visual inspection suggests minor presence of publication bias in the state of the art literature, the whole set of formal statistical tests suggest that there is little to none publication bias present which is an uncommon finding in economics. The results are further confirmed by robustness checks employed on the restricted dataset of studies with reported standard errors.

Subsequently, a set of 37 explanatory variables is used to address model uncertainty by employing the Bayesian Model Averaging (BMA) Technique and a robustness check using the alternative frequentist model averaging (FMA). The choice of the BMA method as the baseline model was based mainly on the fact that BMA addresses possible model uncertainty.

I find that in fact authors focusing solely on specific labour market groups, such as only full-time workers, or only fresh college graduates, report a smaller gender wage gap than those focusing on the labour market as a whole. Other study characteristics affecting the estimated gender wage gap are regions of the developed world, the methodological approaches (including controlling for selection bias) or the omission of dependent variables related to the human capital of individual workers (age and education).

The remainder of this paper is organized as follows: Chapter 2 reviews the existing empirical literature of both primary and secondary studies. Chapter 3 provides an introduction into the methodology commonly used in meta-analysis and the nature of collected data that are being analyzed in the empirical part of this thesis. Chapter 4 presents to the reader the empirical procedure and results obtained by the analysis, as well as their discussion. The fifth and final Chapter 5 concludes the main findings, limitations and contribution of this paper.

Chapter 2

Literature Review

The gender wage gap has been a widely studied topic in today's society. This thesis aims to summarize the topic extensively but more importantly quantitatively. Accordingly, the collection and review of both primary and secondary literature are in place. This chapter provides an overview of existing quantitative research surveys on the gender wage gap but also a comprehensive literature synopsis of studies this thesis analyses in its empirical part. The focus will be on applied methodology and the empirical findings.

2.1 History and present of the worldwide gender wage gap

One of the fundamental reasons for an unadjusted gender wage gap is the over-employment of women in jobs and sectors that generally pay less than others. For example, in the OECD countries, 84% of employed women work in the service sector. While for men it is only 61% (OECD 2018). Even though the majority of men also works in the service sector, other sectors are distinctly more represented by men than women such as the industry sector. There the numbers for male and female employment are 33% and 12% respectively (OECD 2018). For the past 10 years, the numbers regarding these proportions have been approximately the same. Although the choice of employment is up to the individual, it is useful to look at the drivers leading to this choice.

The path to one's job partially determines the choice. It can be caused by gender stereotypes in both the labour market and the preceding education. Another driver could be the higher flexibility in the service sector that allows individuals, in this case, females, to manage their household responsibilities with the benefit of shorter and more flexible hours. According to a gallup poll, in the US, still in 2019, women from married and cohabitating couples are in charge of more household responsibilities than men. However, since 1996 the ratio of male and female responsibilities in households has become more equal (Brenan 2021).

One of the most crucial impacts on the wage differences of men and women is parenthood. It is common for women to leave the labour market to take unpaid care of children where they lose momentum in their career path and are more likely than men to return to a part-time job afterward. While in the past, this has not been frequently up to the woman to decide, the society is now more accepting of parental leave which may be a contributing factor to gender equality in not only wages in the future.

2.2 Meta-analyses on GWG

The first meta-analysis studying the gender wage gap (GWG) was published by Stanley and Jarrell (1998). The authors only included 41 observation from a total of 41 studies and managed to explain over 80% of the study-to-study variation in the estimation of GWG. It is crucial to bear in mind, that this meta-analysis only included studies from the United States. On the one hand, their results suggest that the estimated gender wage gaps of the studies are overestimated in case of a male author, or, as expected, by excluding the information on age or experience of the studied subjects. On the other hand, they suggest that if the authors do not consider and treat selection bias, their results are downward biased. They also argue that the estimated GWG is slowly shrinking with time, considering the years of data collection.

Later in 2004, Jarrell and Stanley (2004) extend their analysis with 49 new observations and compare the results to the study from 1998. In general, their results hold and they even accomplished to explain over 90% of the study-to-study variation. The only significant difference is in the fact that the effect of correction for selection bias diminished.

A paper by Weichselbaumer and Winter-Ebmer (2005) builds on the two aforementioned studies by Stanley and Jarrel. In this case, the meta-analysis is upgraded to the international level and does not focus solely on the United States. Authors of this study include more than only one preferred result per study but the analysis includes 1535 observations from 263 studies that are from 63 different countries. The research covers a time period of over 30 years from 1960s to 1990s. The attention here is on the unbalanced dataset, thus the authors attempt to tackle this issue by various weighing techniques. The conclusions of this study are that chosen methodology only has a marginal effect on the estimated gender pay gap, while over 20% of the study-to-study variation is caused by data restrictions. Compared to the results of Jarrell and Stanley (2004), the effect of male authorship is lower and inconsistent over various model specifications.

The thesis by Konstantinova (2020) also builds a meta-analysis on the newly published gender pay gap studies with estimates from the years 1990 to 2015. The author mostly focuses on the meta regressions estimates differing across regions of developed countries. As well as applying the BMA technique, she also constructs a 'meta' gender pay gap estimate suggesting that the gender wage gap under an ideal study would be on average equal to zero over the examined period of time. While the author admits the possibility of publication bias in primary literature, for the case of the gender wage gap she argues that the presence of publication bias is not very likely due to positive motivation of researchers to publish zero (or even negative) estimates as the persisting presence of the gender wage gap is widely accepted. She herself does not collect or construct any measure of uncertainty for the collected estimates. Moreover, the study altogether excludes estimates coming from quantile regression (Machado and Mata 2005) which is a modern method of gender wage gap analysis.

There are more meta-analyses closely connected to the gender wage gap (Cukrowska-Torzewska and Matysiak 2020; Jarrell and Stanley 1990). However, none of them are directly comparable to the above-mentioned ones or to this thesis.

2.3 Primary literature

The study of wage differentials in general has begun in 1950s with Becker's seminal study *The Economics of Discrimination* (Becker 1957). Becker has addressed discrimation of minorities; including race, sex , religion, etc. Ever since then, the topic of gender wage differentials has been widely studied by many and this seminal study has been used as a primary reference with regards to the unexplained GWG.

Since wage differentials are time sensitive and they evolve constantly, the case has been discussed repeatedly through the fields of psychology and economics, both qualitatively and quantitatively. In all cases, the unexplained part of gender wage gap persists even to the present day. Thanks to the evolution of microdata and growing number of data available, the unexplained GWG is closer and closer to zero but still not completely eliminated. We can only assume that the research will continue for many years to come.

Although the existing literature is exhaustive, the parameters of the analyses differ extensively. Many papers focus only on particular subgroups of population, geographical areas instead of focusing on the whole population of a given country. These papers include Francesconi and Parey (2018); Fortin (2008); Chevalier (2007); Triventi (2013); Manning and Swaffield (2008) that focus only on college graduates, or Jagsi et al. (2006); Madden (2012); Melly (2005); Watson et al. (2010) focusing only on certain occupation groups or sectors.

2.3.1 The estimated gender wage gap

Thanks to the literature published in the last few decades, the empirical literature studying gender wage gaps is quite rich. The case of the United States has been widely studied by Blau and Kahn (2017) over the period of 1955-2014. Many researchers circle around particularly defined causes for the persisting gender wage differentials. Two basic effect groups have been identified as influential on the gender wage gap. First, the role of changing individual human capital (O'Neill and Polachek 1993) and second, the role of job related characteristics of men and women (Black et al. 2008; Bayard et al. 2003). It is far from unusual that male and female workers whose human capital accumulation differs, tend to occupy different job positions and are therefore remunerated on the basis of these endowments.

After controlling for endowments of this type, a part of the gender wage gap remains unexplained. Discrimination may be behind some of these unexplained factors. But what percentage of the unexplained gender wage gap it makes up remains unknown to us. In order to be clear, while some authors distinguish between the explained and unexplained gender wage gap differential, others may refer to them as endowments and remuneration effects. These two types of terminology refer to the same type of effects, respectively.

One wonders whether the inclusion all possible variables explaining worker and employer characteristics in the regression would lead to results explaining the entire gender pay gap. However, most studies lack the feasible number of variables with the power to do that.

Another input effect to explain the entire gender wage gap is the selection effect. Fundamentally, selection effect arises when the sub-sample of workers is non-random and differs from the sub-sample of individuals who do not participate in the labour market. Various methodologies correct for selection bias. The effect of selection is mostly confirmed in studies that work with data from the US (Blau and Kahn 2006; Neal 2002). In Europe (Olivetti and Petrongolo 2008), selection effect shows to be highly negative in Southern European countries, while in Northern countries in shows to be positive. Nevertheless, the labour participation rate of men and women in Northern Europe is comparable. Moreover, Nicodemo (2009) shows that the resulting gender wage gap differs for selection-corrected and non-corrected decomposition. She also argues that the most part of the gender wage gap was explained by the discrimination effect rather that differences in characteristics of male and female workers.

A promising variable has been added to recent literature that can partially explain another part of the unexplained wage gap, however, it is still not quite common. Risk aversion, applicable both in relation to earnings or safety, has shown to be higher for female workers (Le et al. 2017; Black et al. 2008; Dohmen et al. 2005). Research has shown that risk aversion affects choices of occupation or human capital investment that both have an influence on wage differentials.

As many authors who study the gender pay gap focus only on a specific sector of labor market or a subgroup of population, an attractive case of these choices is the analysis of differences of GWG in public and private sectors (Castagnetti and Giorgetti 2019; Arulampalam et al. 2007; Andrews and Kasy 2019; Barón and Cobb-Clark 2010).

In public sector, the average GWG tends to be typically significantly lower compared to private sector, while the earnings distribution varies greatly across the two sectors (Arulampalam et al. 2007). Although many authors draw to this conclusion, it is crucial to mention that the considerably smaller GWG is restricted only to developed countries (Andrews and Kasy 2019).

Blau and Kahn (2017) in accordance with Barón and Cobb-Clark (2010) argue that a large proportion of the GWG remains unexplained especially in higher paying jobs but is comprehensively explained in low paying jobs. Their findings suggest that the unexplained part of the wage differential increases with the wage level in public sector with respect to private sector. This is also confirmed by the analysis of Castagnetti and Giorgetti (2019) using the standard techniques of decomposition. In the public sector, there is evidence found on a glass ceiling. At the same time, regarding the private sector GWG the unexplained part is larger at the bottom of the wage distribution, suggesting a sticky floor.

An important aspect of looking into the reasons for a persistent gender wage gap is the difference of attitudes and acquired human capital of men and women. While many studies typically include variables such as *education*, *experience*, or *tenure*, the differences lay in especially in the incentives of male and female market participants and the consequential models the studies decide to employ. Traditionally, women have a lower labour market participation rate than men (Reimer and Schröder 2006). It is usually women who have more work experience disruptions, which, by the way, is a variable that is often omitted in research, while it may be very important for true and correct estimation of the gender wage gap. The return of a person after a work interruption may cause their income to lower in spite having the same amount of years in experience as a colleague with no experience disruptions (Blau and Kahn 2017). This is common for women in jobs with lower expectations on built experience and other human capital related characteristics. Consequently, the effect on wage of the work disruptions is minimal with these jobs (Polachek 1981). Women will also tend to choose jobs without firm-specific requirements because a firm-specific training would not allow them to transfer their knowledge to a more general settings in case of a work disruption such as the arrival of children.

There are analyses focusing especially on work interruptions and their effect on lowering female wages. Goldin (2014) analyzes the effect of work interruptions as a part of an analysis of the temporal flexibility (or inflexibility) impact on the gender wage gap. Her main focus is on the possible disproportion in rewards for working longer hours. While she focuses mainly on the hours worked, she is able to also analyze the interruptions in the same context. Goldin (2014) argues that the differences in wages are mainly due to the differences in longer work hours instead of differences in the requirements on human capital. However, according to recent literature, we have observed a steady decline in the gender wage differences that is attributed to changes in labour-market experience of men and women that have been decreasing as well (Gayle et al. 2012; O'Neill and Polachek 1993; Blau and Kahn 2006). Others have also found negative effects of work interruptions (Blau and Kahn 2013; Mincer and Polachek 1974) but the effect, same as gender wage gap itself, is diminishing with time.

Adamchik et al. (2003) focuses on the Polish labor market between the years 1993 and 1997 and confirms that there is still a substantial part on unexplained wage differential. However, the findings regarding the explained part suggest that occupational and industrial segregation accounts for its considerable part. And while some studies focus only on one country, Christofides et al. (2013) work with data from 26 different European countries using a 2007 dataset provided by the EU. The wage gaps in the countries differ substantially but still the differences in workers' characteristics nor the differences in countries are able to explain the whole GWG. Furthermore, thanks to quantile regression they also observe existence of sticky floors and glass ceilings in a number of countries.

The persistence of gender wage gap over time may be caused by reasons that no quantitative study can capture within an analysis that has no kind of behavioral or psychological overlap. Some of the reasons that are hard to quantify could be investments in training, choice to work from home or any other reasons coming from generally voluntary choices made by men and women (Weichselbaumer and Winter-Ebmer 2005) that then make a part of the unexplained residual of the gender pay gap.

2.3.2 Empirical approaches

This section will explore different methodological approaches to the investigation of unexplained gender pay gap. It will describe methods from Blinder-Oaxaca decomposition and its extensions to the quantile regression method.

Blinder-Oaxaca Decomposition

The vast majority of studies build on the classical decomposition developed by Blinder (1973). The decomposition is based on a simple regression equation:

$$Y_i = \beta_0 + \sum_{j=1}^{n} \beta_i X_{ij} + u_i, \tag{2.1}$$

where Y_i denotes the natural logarithm of wage, and $X_{ij}, j = 1, ..., n$ are explanatory variables belonging to a worker i used to explain Y. In order to compare the two groups, in this case, men and women, it makes sense to introduce two separate regressions for each group (M - male, F - female):

$$Y_i^M = \beta_0{}^M + \sum_{j=1}^{n} \beta_j{}^M X_{ij}{}^M + u_i{}^M \tag{2.2}$$

$$Y_i^F = \beta_0{}^F + \sum_{j=1}^{n} \beta_j{}^F X_{ij}{}^F + u_i{}^F \tag{2.3}$$

The two regressions form the first step of the decomposition method. As a results of separating the two groups, it is possible to examine the relationship between wages and the characteristics specific to the female and male sample.

The second part of the process is the actual decomposition analysis of the structure of wages. The two regressions can be reorganized and further divided into an explained part of the wage differential (due to characteristics differences) and its unexplained part.

It is up to the author to define what structure of wage makes the non-discriminatory benchmark using which they are able to decompose the wage differential of male and female population. In correspondence with literature on gender wage gap, the male wage structure is used as a reference wage. Further, I will show what the decomposition looks like basing it on Equation 2.2 and 2.3.

$$\begin{aligned} \bar{Y_i}^M - \bar{Y_i}^F &= \hat{\beta_0}^M + \sum_{j=1}^{n} \hat{\beta_j}^M \bar{X_j}^M - \hat{\beta_0}^F - \sum_{j=1}^{n} \hat{\beta_j}^F \bar{X_j}^F = \\ &= \left(\hat{\beta_0}^M + \sum_{j=1}^{n} \hat{\beta_j}^M \bar{X_j}^M - \left(\hat{\beta_0}^M + \sum_{j=1}^{n} \hat{\beta_j}^M \bar{X_j}^F \right) \right) \\ &+ \left(\left(\sum_{j=1}^{n} \hat{\beta_j}^M \bar{X_j}^F \right) - \hat{\beta_0}^F - \sum_{j=1}^{n} \hat{\beta_j}^F \bar{X_j}^F \right) = \\ &= \underbrace{\left(\sum_{j=1}^{n} \hat{\beta_j}^M (\bar{X_j}^M - \bar{X_j}^F) \right)}_{\text{Explained part}} + \underbrace{\left((\hat{\beta_0}^M - \hat{\beta_0}^F) + \sum_{j=1}^{n} \bar{X_j}^F (\hat{\beta_j}^M - \hat{\beta_j}^F) \right)}_{\text{Unexplained part}} \end{aligned} \tag{2.4}$$

The term $\left(\hat{\beta_0}^M + \sum_{j=1}^{n} \hat{\beta_j}^M \bar{X_j}^F\right)$ can be simply interpreted as the wage of a female worker if she was to be paid as an equivalent of a male worker (Leythienne and Ronkowski 2018).

In order to explain, the first part fo the Equation Equation 2.4 is the difference between the mean of the natural log of male wage and the 'counterfactual" mean of the natural log of female wage. It represents the explained part of the wage differential as it takes into account the differences in average characteristics of men and women weighted by the average characteristics of men. The second, unexplained, part is the difference between the actual and counterfactual means of the natural log of female wages. It measures the part of the wage differential that is caused by the difference in both estimated coefficients and constants for men and women weighted by the average characteristics of women (Leythienne and Ronkowski 2018).

Both the explained and unexplained part of the wage differential can be expressed as a proportion of the total wage difference of male and female. Moreover, the explained part can be further decomposed into subcomponents that could also be expressed as proportions of the total wage differential in order to give value to individual characteristics describing the differences in wages.

The original Blinder-Oaxaca decomposition has been subject to rather considerable critique since its publication. Most of this criticism revolves around the choice of explanatory varaibles and consequently the model specification (Ospino et al. 2010). First, Rosenzweig and Morgan (1976) criticizes the choice of variable regarding age and age squared instead of using a variable for work experience and its squared term. He claims that given this misspecification, differential bias is created in estimated returns for male education and therefore the education effect that explains the wage differential between men and women may be overestimated. Overall, the critics suggest that the decomposition of the residual term that represents discrimanation cannot be determined precisely the same way for every decomposition, since the difference in intercept values depends on the decision of measurement (Ospino et al. 2010).

Another popular critique of the Blinder-Oaxaca decomposition is that the discrimination is only measured in the labor market. It does not consider the possibility of differences in access to education or the possibility of one gender being more prone to work than the other (Madden 2000). Furthermore, Atal et al. (2009) suggest that the decomposition only considers the average wage gap, but completely omits the possibility of various distributions among individuals belonging to the same group. Another flaw is hidden in the lack of restrictions to comparable individuals.

Extensions

The original Blinder-Oaxaca decomposition method from 1973 has gained some extensions over the years. The perhaps most used and adopted alternative was introduced by Neumark (1988). He bases his method of decomposition on a pooled regression without using group specific intercepts. His distinct technique analyzes what part of the unexplained wage differentials represents discrimination, when its already cleaned of differences in productivity. He focuses on the relationship between employers' discriminatory tastes and the final estimate of discrimination regarding wages.

Cotton (1988) argues that in absence of discrimnation on the labor market, the wage structure of men nor women would not exist. In conflict with Neumark (1988), he suggests that both male and female wages should have as a reference earnings structure applied weighted average.

$$\beta^* = \frac{n^M}{N}\beta^M + \frac{n^F}{N}\beta^F \tag{2.5}$$

Neumark (1988) states that in absence of discrimination on the labour market, this constitution is unsatisfactory and that pooled regression estimates should be used.

Nonetheless, the wage differential is suggested to be decomposed into three parts, with the first one representing the explained gender wage gap, and the unexplained GWG being divided into two separate residuals. The first term of the unexplained part is attributable to the male advantage and the second one to the female disadvantage (Chevalier 2007).

The main reason for diversified estimates of the gender wage gap are mainly due to the inconsistency in the choice of a reference wage structure in decomposition techniques.

Recently, the quantile regression methods have been used to estimate the gender wage gap. In order to incorporate quantile regression in the B-O decomposition framework, authors make use of a generalization of the Blinder-Oaxaca decomposition introduced by Machado and Mata (2005) that makes quantile regression applicable. The benefit of this method is that the estimates can be interpreted across the whole earnings distribution, at any of its quantile (Castagnetti and Giorgetti 2019). The difference to the classical Blinder-Oaxaca is that here the GWG is based on the construction of a distribution of what would be the wage of women given that the earnings structure is the same as the one of men.

According to Elder et al. (2010), an appealing method to get a single coefficient for the unexplained gender wage gap is a pooled OLS regression. They further argue that compared to including a group estimator in OLS, the pooled Blinder-Oaxaca strategy understates unexplained differences by overstating the observables' role in the explanation of the mean outcome. Their conclusion is that pooled OLS regression estimates are biased because of omitting group specific intercepts, causing overstating of the observables' role.

Chapter 3

Methodology and Data

3.1 Methodology

The methodology related to the topic of meta-analysis is otherwise known as a quantitative literature review. Gathering information from the narrative point of view is believed to be insufficient in some cases where a quantitative assessment is in place. Narrative reviews may be highly influenced by the author and their beliefs while empirical reviews should be highly objective. The use of the word 'should' is in order as even empirical studies might be biased. In order to estimate the true effect it is necessary to proceed with a quantitative literature review.

Meta-analysis offers researchers to systematically and quantitatively overview empirical literature. With using this tool to integrate the results of empirical research, the researchers aim to increase its statistical power. More attention can be paid to the factors that are identified to have influenced the true effect and therefore the results are not as sensible to human influence as the results of a supposed qualitative survey.

The history of meta-analyses dates back to the beginning of the 20th century. It was commonly used in the medical sphere (Pearson 1904) and it steadily spread across all empirical research. In 1990, Jarrell and Stanley (1990) introduced the method of meta analysis into the field of economics. In 2020, Havránek et al. (2020) states that more than 2000 meta analysis had been conducted. In 2012, it had only been 656 (Stanley et al. 2013). Both the methods of meta-analysis and the numbers of observations studied within a meta analysis have increased dramatically. The meta-analysis by Xue et al. (2019) includes 12,788 estimates on social capital and health from more than

450 studies, with identifying 71 variables that may explain the heterogeneity among them (Havránek et al. 2020).

Stanley (2001) introduces the procedures to conduct the quantitative research synthesis. He mentions the important steps in the beginning. First, the researcher should include all the relevant studies, published or not, in order to eliminate any sort of non-randomness in the collected data. Also, one should incorporate in his paper the details of the search he conducted so it can be easily replicated by others if they wish to do so.

Then, it is important to reduce the sample to studies that have empirical value, i.e. they report empirical results, standard errors, etc. Afterwards, the dependent variable should be chosen such that it is comparable across the relevant studies . Usually, results of statistical tests, elasticities and regression coefficients are used as a comparable metric.

The next stage would be identifying the independent variables. Those are the characteristics specific to each study that will serve for explaining the differences and variations between studies.

In meta-analysis, instead of collecting data points regarding individuals, the author collects reported estimates from already existing studies that analyse the topic at hand. The data are collected with regards to the methodology and the data they use. Only that way can the author employ a sensitivity analysis of the estimated coefficients to the different datasets and applied methods, or model specifications. While the goal of a meta-analysis is not to specify the perfect model for estimation of gender wage gap, it can easily and comprehensively show which of the study's or author's characteristics or choices affect the resulting estimates of the effect they are studying, in this case *gender wage gap*. It is important that researchers take the conclusions of meta analyses into account in order to avoid needless variation among their estimates.

Meta-analyses usually consist of two crucial parts. One of them is the estimation of the effect various studies have on the outcome of those studies. This part may also include an estimation of the true effect. Also, apart from the common tests for heteroscedasticity, misspecification or autocorrelation, a researcher should employ test for the presence of publication bias which tends to be the second part of a modern meta-analysis. In this thesis, I will follow the widely used and modern procedures and will employ both of these parts as they are considered crucial for a worthy meta analysis.

3.1.1 Publication bias

In the first part of the analysis, I will analyse the publication bias. In order to detect and quantify the publication bias, I will use funnel plots. Funnel plots are the most common approach to detecting publication bias. Then, I will follow up with meta regression tests which will serve as a formal objective test of publication bias.

It has been believed that science is suffering from a credibility crisis. Many researchers use questionable methods and sometimes even fabricate their results (Ioannidis et al. 2017). Detecting this kind of manipulation can be very hard. A common reader would not recognize fabricated results from the correct ones, especially in cases where they are unable to replicate the analysis, that is essentially applying the same econometric methods and collecting a sample of data equivalent to the original study (Andrews and Kasy 2019). There might be some indicators even without the need to replicate the study (i.e. important coefficients are all statistically significant, only some of the control variables are reported, etc.). Still, many are left undetected. Especially in meta-analyses, it is crucial to take this into consideration.

The key to assessing the validity of a scientific paper and its results is firstly publication or misspecification bias and the statistical power of the results (Ioannidis et al. 2017).

However, if an author decides not to publish his study (and consequently the results), we are looking at yet another type of publication bias. This problem is known as the "file drawer problem" and was first presented by Rosenthal (1979). This phenomena has been acknowledged by researchers in many different fields and can be caused by various reasons. First, there is political influence that can have this consequence, this can include sponsorship that motivates author to publish, or not to publish in some cases.

Another explanation could be the motivation to publish statistically significant results that are viewed as more valuable to research. Some may even doubt their contribution when their results are insignificant causing them not to publish their papers. According to Thornton and Lee (2000) psychological theses and dissertations are more likely to be published with positive and statistically significant results than with negative ones. This may be driven by authors themselves who do not even try to publish once they recognize that their results are not as they expected them to be.

Suppose an author wants to publish a paper supporting a hypothesis that has been statistically supported on samples from other countries but is yet to be studied in the country of their interest. Suppose the results are expected to describe a negative effect. However, in the case of the country of their interest, the results reveal themselves to be positive and, moreover, statistically significant. One the one hand, the decision not to publish might be simply because of the undesirable sign of the effect itself. On the other hand, these results may have an effect on the future studies this author planned to write or even already published. But once they publish the undesirable effects it may be irrelevant to pursue the future research as planned leaving the author in a difficult situation.

Of course, the resulting outcomes may be affected by the methods the authors choose or an inefficient sample size causing large standard errors and hence insignificant results.

So while narrative literature can be viewed as non-objective, not even empirical literature is exempt from bias. At the same time, this holds for quantitative literature surveys also. The collection of studies to be included in a meta-analysis is a work of the author. Even though the mechanisms are designed to be as much objective as possible, there is space for bias. Some authors include unpublished papers that can bring in some unreliability of the study. In this thesis, only papers written in English are included. This being said, the exclusion of relevant papers in different languages may also have consequences on the final outcome.

Having mentioned the reasons for publication bias, this thesis will aim to reduce its publication bias to the minimum, by for example including only published and peer-reviewed papers, and detecting publication bias in the already existing literature. To obtain relevant results I will be using a graphical method (funnel plots) and statistical regression tests.

3.1.2 Bayesian model averaging

The structural heterogeneity remains to be explained after resolving the issue of publication bias. It is expected, that estimates of adjusted gender wage gap will differ across age, occupation, regions and other. However, some variation can be caused by different estimation approaches.

To investigate the differences in reported estimates it is important to regress them on a set of explanatory variables that are a potential source of heterogeneity.

The above mentioned regression model can be expressed by Equation 3.1.

$$gwg_i = gwg_0 + \beta_j \times \sum_j X_{ij} + u_i, \tag{3.1}$$

where gwg_{ij} is the considered effect estimate; gwg_0 is the constant, β_j is the vector of coefficients, and X_{ij} is the set of explanatory variables of study characteristics (including the standard error); u_i is a normal IID error term. Following this equation, any potential publication bias is expected to be varied randomly across all studies. The only modeled variation is the systematic variation.

The next step would be to select a model. Ordinarily, one by one the insignificant explanatory variables would be excluded from the model until there was left a model with only significant regressors. But because the choice to include some variables might be complicated due to their high number, it is advisable to consider model averaging techniques.

Bayesian model averaging is a technique for dealing with model uncertainty by combining all possible model specifications with the use of the variables available as it calculates a weighted average over all of the combinations. There are 2^N model combinations where N is the number of available explanatory variables (Havránková 2015). It is more appealing than e.g. frequentist meta-regression where one has to specify the key explanatory variables and the set of control variables themselves which could bring subjectivity into the analysis. Since subjectivity is the thing we are trying the hardest to avoid in meta-analyses, it would be infeasible to follow this path as a baseline estimation. We employ frequentist techniques only as a robustness check. Another benefit of following the BMA method is its fairly straight-forward interpretation that is not misleading in the way in which the frequentist technique often is. In the frequentist procedure, coefficients and their respective p-values are reported. In has been shown (Raftery 1995) that in large samples, p-values may be misleading and lead to inconsistency because there is no account for model uncertainty.

For the frequentist technique, the set of key explanatory variables would need to be identified as well as a set of controls that are considered to be weak. Consequently, the the insignificant variables would be removed one by one using t-tests. Such a method could be statistically invalid due to ignoring model uncertainty but also quite inconvenient with the number of available variables. The ignored model uncertainty could lead to a biased estimation. Bayesian model averaging can overcome this limitation by estimating all possible combinations of explanatory variables and calculating a weighted average over all of them. I will use frequentist meta regression onyl as a form of robustness check to the preferred BMA method.

Another reason to choose BMA instead of a frequentist procedure is the model interpretation. The standard frequentist method reports coefficients together with their p-values indicating the evidence strength against the null hypothesis (=a coefficient is zero). However, especially in large samples (such as ours) p-values can be misleading and can lead to inconsistent conclusions by not taking model uncertainty into consideration (Raftery 1995). BMA, on the other hand, takes model uncertainty into account and enables the researcher to distinguish between the reasons to reject the null hypothesis: i. due to insufficient data; ii. due to evidence for the null (Hoeting et al. 1999), both due to returning the posterior effect probabilities.

The aim of BMA in this thesis is to asses the effect of each explanatory variable from the constructed dataset on the outcoming gender wage gap estimate, which can serve as a tool for designing future studies but also allows us to correctly predict the gender pay gap ourselves. The BMA uses Bayesian theorem and the law of total probability while looking at the unknown parameters as random variables to obtain the results.

3.2 Data

3.2.1 Data collection

The process of data collection for this meta-study began with an inspection of a widely known research database - Google Scholar by a search for the most distinguished papers on the topic of GWG. The aim was to find at least 5 papers that have been commonly cited and were, ideally, published in one of the prominent journals.

After this primary identification, a search query on Google Scholar was constructed such that they included studies on the topic at hand and the primarily chosen papers were identified among the first hits. The search was conducted in English and only papers written in English will be included in further examination. It is acknowledged that due to this restriction, the analysis may be biased due to non-random sampling.

In order to include also the most recent studies on gender wage gap, the search query was first restricted to years between 2005 and 2021, and then 2015 to 2021 to be sure to incorporate even the newest papers. This meta analysis is building on the existing meta study by Weichselbaumer and Winter-Ebmer (2005) which is why the data collection is restricted to papers that were published after the publication of the aforementioned study.

The constructed search query is the following:

"gender wage gap" OR "gender pay gap" OR "wage differentials" gender pay gap differential discrimination estimate regression empirical

Then, the abstracts of the selected papers are examined to determine whether they carry any empirical results. Those will be further analysed and will be included in the final dataset in case they use adequate econometric methods of estimating the gender wage gap.

Some of the identified studies were only of a qualitative character without any implementation of econometric methods. They usually only referred to other quantitative studies which may have caused they inclusion in the results of the specified search query. Those were automatically excluded from any further review. Studies implementing econometric methods remained for further examination.

Customarily, the distinction between relevant and irrelevant results can be made by identifying the presence of standard errors with the estimates. Unfortunately, that is not the case of studies of gender wage gaps. Since the most common method only reports standard errors occasionally, it can be harder to distinguish relevant results. This issue will be further discussed and resolved in the chapter regarding Publication Bias. To overcome this issue in the data collection I focus solely on peer reviewed research papers and examine the used econometric methods more deeply than would have been necessary if the standard errors had been reported. All studies using methods whose relevancy was unverifiable were excluded from the final dataset.

The studies had to follow other certain rules to be included. I only included studies from year 2003 forward as the aim of this thesis is to work with most recent data, the underlying data cover only the period from the year 1990. Another rule was to collect estimates from studies that analyse surveys or markets constructed throughout all professions without fixating on certain profession or industry subsamples.

After the data collection I compared the resulting literature findings with literature of the most recent found meta-analysis on gender wage gap by Konstantinova (2020). The vast majority of found studies was identical. I examined the remaining studies (Konstantinova 2020) and decided to include those that were relevant for this study (peer reviewed and working with data from developed countries only).

This process yielded 51 relevant studies summarized in Table 3.1 comprising 661 estimates of the adjusted gender wage gap.

Table 3.1: Studies included in the Meta-Analysis

Addabbo and Favaro (2011)	Francesconi and Parey (2018)	Livanos and Pouliakas (2009)
Albrecht et al. (2004)	Fransen et al. (2012)	Manning and Swaffield (2008)
Arulampalam et al. (2007)	Gallen et al. (2019)	McDonald and Thornton (2007)
Avlijaš et al. (2013)	Gardeazabal and Ugidos (2005)	Meara et al. (2020)
Barnard (2008)	Grove et al. (2011)	Meng (2004)
Bayard et al. (2003)	Healy and Ahamed (2019)	Miller (2009)
Bensidoun et al. (2018)	Chevalier (2007)	Mueller and Plug (2006)
Bertrand et al. (2010)	Chevalier (2004)	Mysíková et al. (2012)
Black et al. (2008)	Christofides et al. (2013)	Nicodemo (2009)
Boll and Lagemann (2018)	Jones et al. (2018)	Nyhus and Pons (2012)
Boraas and Rodgers III (2003)	Joy (2003)	Picchio and Mussida (2011)
Campos-Soria and Ropero-García (2016)	Jurajda (2005)	Redmond and McGuinness (2019)
Castagnetti et al. (2017)	Kee (2006)	Reimer and Schröder (2006)
Fernandes and Ferreira (2021)	Korkeamäki and Kyyrä (2006)	Russell et al. (2010)
Flinn et al. (2018)	Le et al. (2017)	Simón (2012)
Fortin et al. (2017)	Lechmann and Schnabel (2012)	Tharp et al. (2019)
Fortin (2008)	Litman et al. (2020)	Zajkowska et al. (2013)

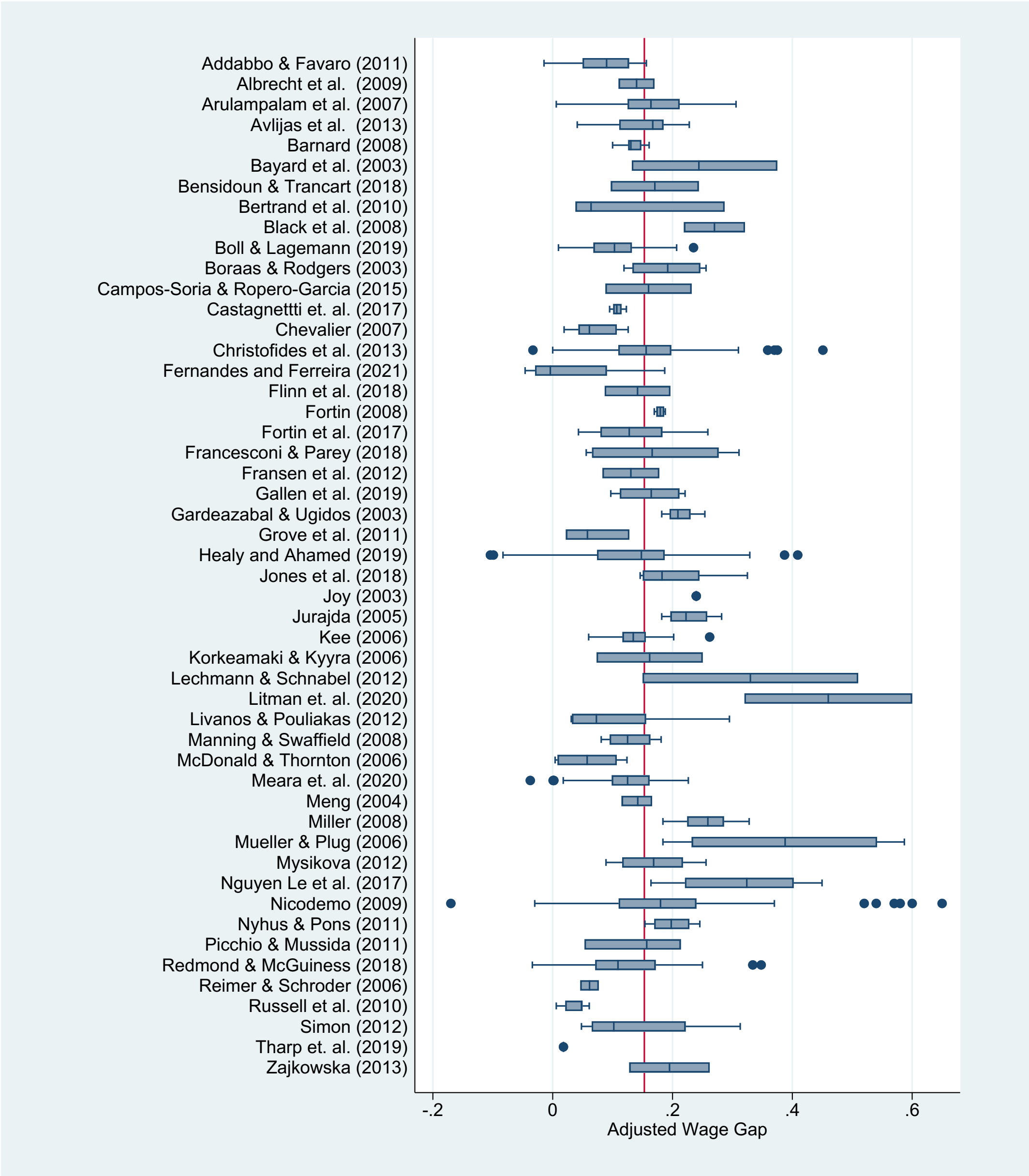
To visually represent the findings of the collected studies for a meta-analysis, I present a forest plot in Figure 3.1 that shows how the adjusted gender wage gap estimates are heterogeneous and different both within and across studies.

3.2.2 Data characteristics

For each of the colected studies, I extract the adjusted gender wage gap (referred to as GWG) then, if available, its standard error, and other additional characteristic variables that will serve to help us reveal the underlying heterogeneity of the estimates. The majority of variables are, as expected, dummy variables. Compared to the meta-analysis of Konstantinova (2020), I decided to reduce the number of dummy variables used in any of the analysis. While I collected a larger number of variables than subsequently used, the decision not to include came soon after realizing that their reporting value was poor. The reason for exclusion of some of the dummy variables was similar. In the final dataset, I do not include dummies whose average is very close to 0 or 1, as their added value to the analysis would be immaterial.

I consider three types of variables for the analysis. First, variables regarding the author's or the study's characteristics are included. Then, I also collect variables related to the methods applied in underlying analyses and lastly, I

Figure 3.1: Forest Plot



study the approaches (i.e. included or omitted dependent variables) to individual regressions that produce the independent variable of the adjusted gender wage gap. The last mentioned type of variables may, and most likely will, vary within a single study.

The decision to include some study specific variables arose from two different aspects. The potential added value to the analysis, i.e. the frequency of inclusion or omission of variables in regressions, or economic intuition coming either from the typically collected variables in a meta-analysis or potential influential aspects of a characteristic. Specifically, meta-analysis do not tend to analyze the gender of the author. However, in a paper focusing solely on gender differences, a characteristics such as this one can be perceived as possibly important.

Many analysis on wage differentials focus only on segments on the labour market or are based on data from specific sources. In my thesis, I try to incorporate as many different approaches to the evaluation of the adjusted gender wage gap as possible, so the inclusion of variables that capture differences between underlying data was natural. These include variables such as *data, private, public, etc.* or the division into region categories of Europe, North America, and the rest of the developed world.

The information extracted from each study is summarized in Table 3.2.

Table 3.2: Description of regression variables

Variable	Description
gwg	The adjusted gender wage gap obtained directly from a study
Standard error	The standard error reported directly in a study.
Study specific variables	
College graduates	A dummy variable equal to one if a study only works with data of new college graduates
Data	A dummy variable equal to one if the data come from a demographic statistics
Female	A dummy variable equal to one if the study has solely female author or authors
Fulltime	A dummy variable equal to one if the study works with data of only full-time workers
Policy	A dummy variable equal to one if the topic is of high importance in the political debate of a country or region of the original underlying data
Private	A dummy variable equal to one if the study focused on only private sector workers

To be continued on the next page

Table 3.2: Description of regression variables - continued

Variable	Description
Public	A dummy variable equal to one if the study focused on only public sector workers
Central & Eastern Europe	A dummy variable equal to one if the data come from the Central & Eastern Europe region
North America	A dummy variable equal to one if the data come from Canada or the USA
Northern Europe	A dummy variable equal to one if the data come from the Northern Europe region
Southern Europe	A dummy variable equal to one if the data come from the Southern Europe region
Western Europe	A dummy variable equal to one if the data come from the Western Europe region
Survey (base group)	A dummy variable equal to one if the data come from a survey
Methodological aspect	
Blinder-Oaxaca	A dummy variable equal to one if the study does not use Blinder-Oaxaca decomposition to obtain the adjusted gender wage gap
Quantile Regression	A dummy variable equal to one if the study employs quantile regression
Selection	A dummy variable equal to one if the study does not correct for selection bias
Regression specific variables	
Age	A dummy variable equal to one if the study omits workers' age
College Major	A dummy variable equal to one if a study omits the college major of a worker
Contract	A dummy variable equal to one if the study omits the type of workers' employment contracts
Education	A dummy variable equal to one if the study omits workers' level of education
Experience	A dummy variable equal to one if the study omits workers' years of experience in the relevant field
Female Share	A dummy variable equal to one if the study omits the share of female in a workplace
Firm size	A dummy variable equal to one if the study omits the size of the worker's firm
FT or PT	A dummy variable equal to one if the study does not distinguish between a part-time and a full-time worker
Hours worked	A dummy variable equal to one if the study omits the number of hours worked by a worker

To be continued on the next page

Table 3.2: Description of regression variables - continued

Variable	Description
Children	A dummy variable equal to one if the study omits if a worker has children
Children Number	A dummy variable equal to one if the study omits the number of children the worker has
Industry	A dummy variable equal to one if the study omits the industry of the workers' job
Marital status	A dummy variable equal to one if the study omits one's marital status
Occupation	A dummy variable equal to one if the study omits workers' occupation
Race	A dummy variable equal to one if the study omits workers' race
Region	A dummy variable equal to one if the study omits the region of a worker
Salary (base group)	A dummy variable equal to one if the study works with a worker's weekly, monthly or annual salary.
Sector	A dummy variable equal to one if the study omits the sector of the workers' job
Tenure	A dummy variable equal to one if the study omits if a worker has tenure
Wage	A dummy variable equal to one if the study works with an hourly wage of a worker

Note: The dummy variables ommited from the analysis are denoted as a "base group".

As reviewed in the literature review in Chapter 2, most papers on gender wage differentials use Blinder-Oaxaca decomposition. That made it a natural choice as one of my dependent variable, same as the use of quantile regression. I also consider important to include a variable indicating whether the underlying study corrected for selection bias.

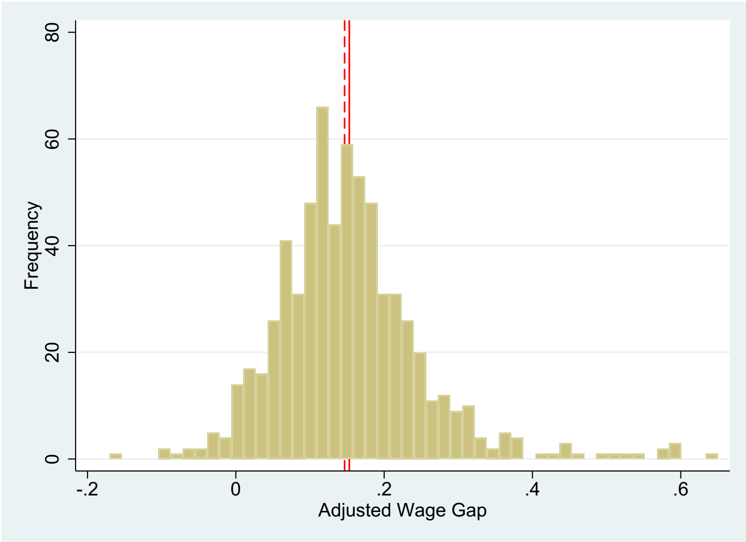
The selection of the remaining variables was made after the identification of the most frequently used variables in analyses dealing with this topic. The regression specific explanatory variables carry information mostly about omission of a given variable, or about how the underlying data are applied.

3.2.3 Data description

The estimated gender wage gaps have a simple mean of 0.149 and a standard deviation of 0.081. A histogram of the the estimates is presented in Figure 3.2.

The distribution seems symmetrical with a small skewness to the right. The median of the GWG estimates is 0.15 as well as the unweighted mean.

Figure 3.2: Histogram of *gwg* estimates



With accordance to other sources, the adjusted gender wage gap diminishes with passing years. It is clear that in the case of my data the decrease is only very slight but detectable. Since the data covers only the last two decades, the nature of the slight decrease is understandable. The depicted scatter points in Figure 3.3 represent an average gender wage gap per study. We must bear in mind that the year of publishing a study does not necessarily mean that the single estimates come from the same year. A study can contain estimates throughout years.

To compare, I attach Figure 3.4 depicting all collected estimates against the year of data origin. There, the decrease illustrated by the trend line is more apparent than in the previous case. On the other hand the graph is not as neat as Figure 3.3 due to the fact that the studies' estimates may differ in years.

Looking at the data by regions, divided into Western Europe (WE), Southern Europe (SE), Northern Europe (NE), Central and Eastern Europe (CEE), North America (NA) and the rest of the developed world, the country specific averages are all very close to each other, ranging from 0.136 in the CEE region to 0.183 in North America as a weighted average value of all collected estimates from the regions regardless of the original studies, the year of publishing or the year of data collection. The weigh is set to be the number of estimates reported per study. On the other hand, in the region of CEE there is also the highest difference between the lowest and highest estimated GWG. Most collected es-

Figure 3.3: Study average *gwg* Estimates in Time

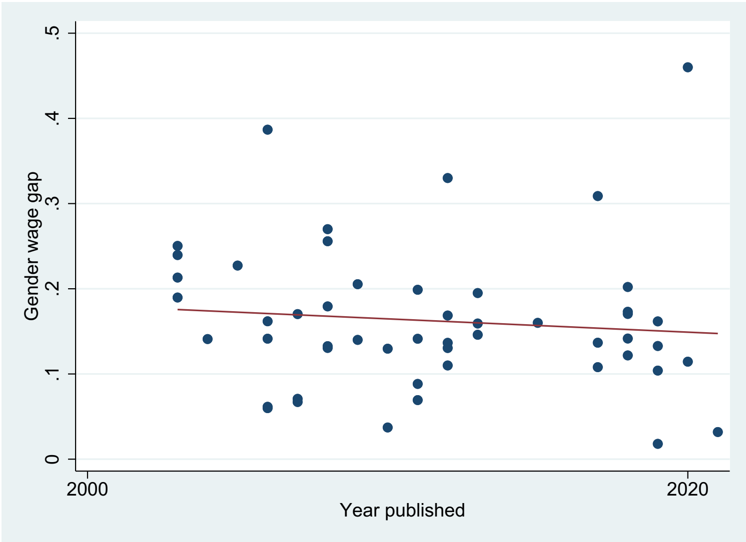
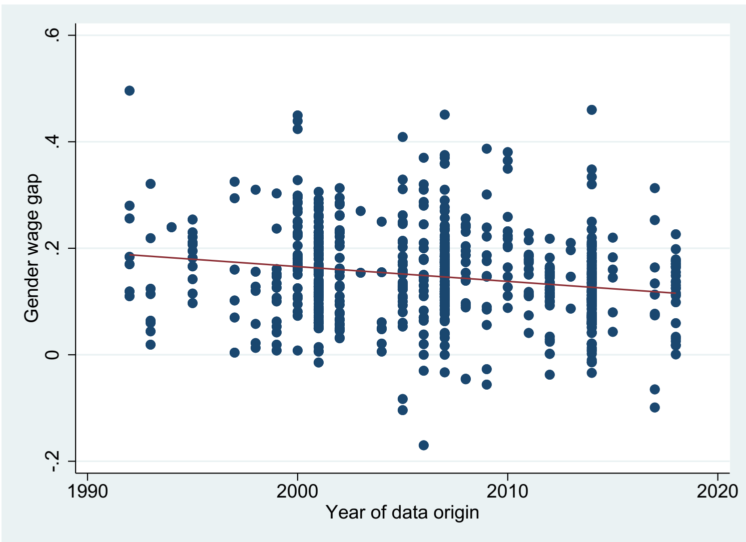


Figure 3.4: *gwg* Estimates in Time - Full Sample



timates are based on data from Northern Europe, they form one third of all collected *gwg* estimates. Only seventeen estimates come from Australia, which is the only country that falls under the base category of *'rest of the developed world'*. Out of the seventeen collected estimates, the average is 0.141. Since the US labour market is quite large, studies that report estimates based on US labour market data do not analyze other countries but rather focus only on US. Hence many of these estimates serve only as controls, supposedly with restricted number of regression variables. While for the case of Europe, many studies report estimates to compare the cross-country situation. I suppose that studies reporting estimates from Europe would be hesitant to report as many robustness checks for every single country as the ones based on US data.

The Table 3.3 lists all the collected variables along with their summary statistics.

Table 3.3: Summary statistics of regression variables

Variable	Mean	St. d.
gwg	0.149	0.081
Standard Error	0.014	0.010
Study specific variables		
College graduates	0.123	0.328
Data	0.729	0.445
Female	0.175	0.381
Fulltime	0.263	0.441
Government	0.141	0.348
New Entrants	0.071	0.257
Policy	0.817	0.387
Private	0.086	0.281
Public	0.051	0.221
Central & Eastern Europe	0.106	0.308
North America	0.077	0.267
Northern Europe	0.118	0.323
Southern Europe	0.104	0.306
Western Europe	0.442	0.497
Methodological aspect		
Blinder-Oaxaca	0.251	0.434
Quantile Regression	0.159	0.366
Selection	0.755	0.430
Regression specific variables		
Age	0.460	0.499
College Major	0.870	0.337

To be continued on the next page

Table 3.3: Summary statistics of regression variables - continued

Variable	Description	St. d.
Contract	0.688	0.464
Education	0.342	0.475
Experience	0.604	0.490
Female Share	0.950	0.218
Firm Size	0.664	0.473
FT-PT	0.590	0.492
Hours Worked	0.717	0.451
Children	0.793	0.406
Children Number	0.915	0.279
Industry	0.666	0.472
Marital Status	0.625	0.485
Occupation	0.551	0.498
Race	0.778	0.416
Region	0.699	0.459
Salary	0.098	0.298
Sector	0.617	0.486
Tenure	0.694	0.461

As mentioned before, only dummy variables with their mean value between 0.05 and 0.95 were included in both the overview of summary statistics and the analysis itself.

Authors usually base their analysis on data coming from a statistical database of a country or a region of interest. There are, however, some who rather use data from surveys. The mean of the estimates gender wage gap weighted by the number of observations per study based on data from a statistical source is lower by 0.05 than the average estimate based on survey data. Survey data face a great disadvantage, as the respondents are not obliged to respond truthfully. They are often not motivated to provide honest and accurate answers. Also, the sample is more prone to being nonrandom as it is very hard to find a representative sample of respondents. The average reported estimate coming from survey data is 0.19 which is higher than the simple average reported gender wage gap estimate.

With studies concentrating only on the public sphere, the reported gender wage gap is on average larger than in studies focused on the private sector only. This comparison between the two statistics is unexpected as I would anticipate that in the public sector salaries are managed on a tabular basis disregarding the gender of a labour market participant. However, only a small number of authors deal directly with either the public or private sector alone.

What I consider surprising in the nature of the applied regressions is that quite a lot of authors did not include *age* in the explanatory variables of all their regressions. I would consider this variable to be of great importance when addressing this issue as the question of age is closely connected to the level of wages which is a crucial underlying aspect of these analyses. While some of this can be explained by the fact that a large number of collected estimates may had been used as robustness checks with excluded estimates, it is not uncommon for studies to omit these variables completely.

Only 47 of my observations are classified as the outcomes of an analysis based on data of only new labour market entrants. Their adjusted gender wage gap is on average much smaller than the mean adjusted wage gap of all of our observations, 0.11. People who enter the labour market for the first time tend to have the same opportunities so the lower average wage gap is not unexpected.

Chapter 4

Empirical results

4.1 Publication bias

The study of the presence of publication bias will start with the analysis of the funnel plot and will be followed by other formal testing methods.

A funnel plot is the most commonly used graph designated to identify the existence of publication bias in research literature. It is widely used in meta-analyses and systematic reviews (Sutton 2000). On the vertical axis of the funnel plot is the measure of precision, while on the horizontal axis the size of the effect is measured, usually represented by a regression coefficient or elasticity.

The precision can be measured differently. A basis for its measurement is the inverse of a standard error of the measured effect. In this case, where the standard errors are missing in the majority of studies, it may make things a little harder but there is a rather elegant solution to this.

According to Havranek et al. (2015) it is possible to compute the approximate standard error for studies that do not report the actual measures of uncertainty. The computation is straightforward. After identifying the studies without reported standard errors, I make the assumption that estimates of each study are normally distributed. Then, I take the median estimate of the GWG and use the difference between the 50th and 16th percentile of the distribution estimates. Intuitively, it only makes sense to use studies reporting multiple GWG estimates. With the assumption of the presence of publication bias it is expected that authors would tend to report a smaller standard error (in this case represented by the approximate standard error) for smaller GWG estimates. Same as Havranek et al. (2015), I assume this to be a great possibility for us to analyse a larger portion of data. Without it, I would be limited to GWG studies using only methods that generally include standard errors with their reported estimates. Since the majority of studies does not report measures of uncertainty, I would be limited to only 135 estimates.

However, another way to detect publication bias could also be in using the sample size of a study or its square root (Stanley 2005).

The funnel plots are interpreted as follows: in the absence of publication bias and in presence of some expected heteroskedasticity, the plots should be distributed randomly but symmetrically in the shape of an inverted funnel, hence the name. The shape of the *ideal* plot is given by higher standard errors in smaller samples.

Nonetheless, the crucial aspect for identification of publication selection is the symmetry of the plot. Simply put: the more asymmetrical, the more likely is the publication selection present. The fundamental belief is that in case of a publication bias, one direction is preferred to the other one. But in case we suspect publication bias in the significance size, that is harder to detect from the funnel plot. In general it would mean that the funnel plot is very wide and hollow.

As was described earlier, the funnel plot with approximate standard errors may look a little unconventional with estimates from one common study (without a reported standard error) sharing the approximate standard error.

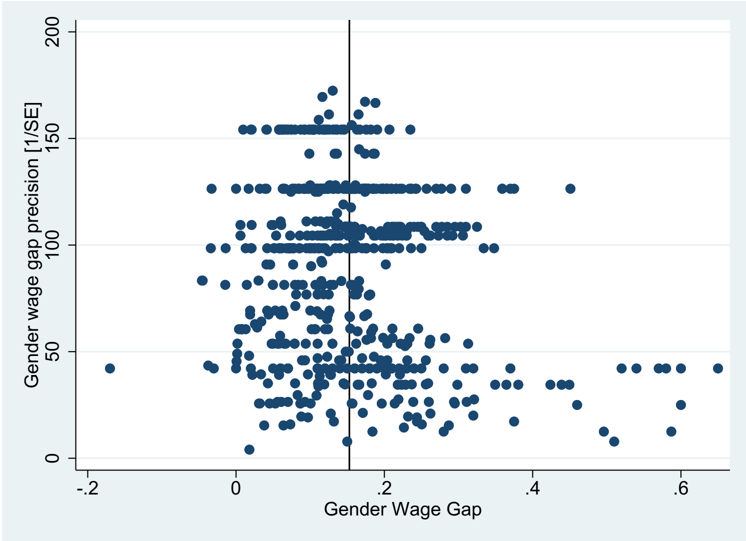
Figure 4.1 represents the individual *gwg* estimates portrayed alongside with the precision measure - the inverse of standard error. Intuitively, with a higher standard error, the precision measure is smaller. This can be observed in the estimates at the bottom of the plot that have a very wide distribution of gender wage gap estimates and come from one study. Their approximate standard error is quite large. Estimates with extreme values were excluded from the visualization for better orientation in the presented funnel plots.

Considering the reasons behind large constructed approximate standard errors, the funnel plots suggests that the publication bias may be present but only to a limited extent. Otherwise, the funnel plot appears to be rather symmetrical in a shape of an inverted funnel. Therefore, the shape of the funnel suggests that there is little if some publication bias in the underlying studies.

Moreover, the funnel plots suggest a true effect closer to 0.2 than the simple mean of the collected estimates (0.15).

It is appropriate to compare the funnel plot of all estimates and the ap-

Figure 4.1: Funnel Plot

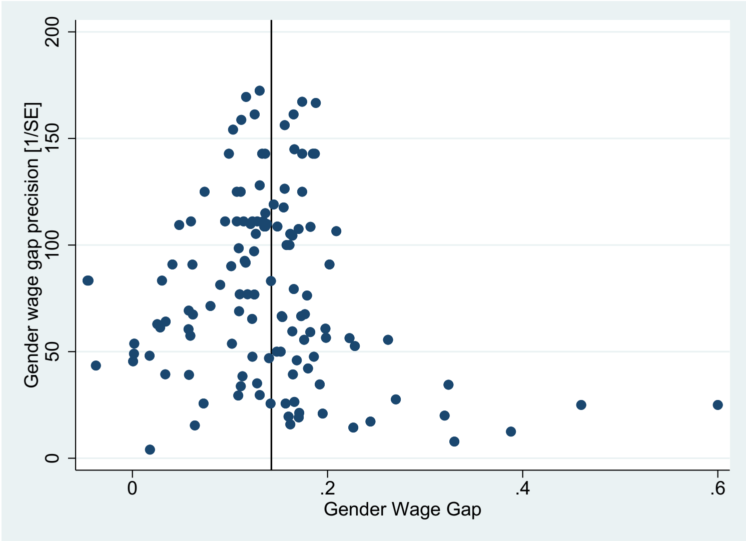


Note: The studies without reported standard errors have a common constructed approximate standard error per study which is the reason for the presence of horizontal lines of this funnel plot, as many estimates 'share' a standard error within a study.

proximated standard errors with a plot that only takes into account the median estimate for the studies with no reported standard error. That way, the approximate standard errors are visualized only once in the funnel plot. While the funnel plot in Figure 4.2 is more scattered due to a lower number of estimates, it still resembles the shape of the inverted funnel quite well. Unfortunately, due to the fact that the funnel plot combines median values from studies without reported standard errors with all the values that report a standard error, to comment on the frequences of occurence of negative and positive estimates would be inadequate. We can notice that the plot is slightly skewed to the right. This sort of asymmetry can be an indication of publication bias. A robustness check in the form of a funnel plot of estimates with reported standard errors is attached in Appendix A. A symmetrical funnel plot would be pictured with the most precise estimates at the top of the graph without any indication of preference towards positive or negative estimates. Both of the presented funnel plots are asymmetrical to a certain level and thus indicate publication bias. To ascertain that assumption, I need to employ a more rigorous method with straightforward results to interpret.

The interpretation of the funnel plots, while intuitive, may be subjective. That is why the implementation needs a more quantitative and objective method of detecting publication bias is in place. And with that, meta

Figure 4.2: Funnel Plot: Median *gwg* estimates



regression tests will help. Meta-Regression Analysis (MRA) can be used for this particular purpose (Stanley 2005) but it is more often used to study the variation of econometric results that have been reported as was introduced by Krueger and Card (1995).

The following methodology description draws mainly from a study by Stanley (2005) — *'Beyond Publication Bias'*. The simplest approach in MRA testing is to model the relation between the reported effect and its standard error.

$$x_i = \beta_0 + \beta_1 \times SE_i + \epsilon_i, \tag{4.1}$$

where x_i is denoting the *i-th* reported effect and SE_i is its corresponding standard error, both reported and approximated (if not available). The coefficient of β_0 is representing the 'true value' of the effect around which the observed effects should vary independently. I assume that the errors ϵ_i will be heteroscedastic as the studies differ in many aspects, including methods and sample sizes. Particularly, the coefficient β_1 brings us the information about the potential size or direction of publication bias. The formal name of the first test is Funnel Asymmetry Test (FAT). It tests the null hypothesis that publication bias is not present, i.e. if the coefficient of β_1 is statistically not different from 0, we assume there is no publication bias present. In the case of its detection, it will be proportional to the standard error. The second formal test, Precision Effect Test (PET), investigates the 'true value' of the adjusted gender wage gap with the null hypothesis of $\beta_0 = 0$.

The disproportions of the sample sizes and model specifications causes the variance of error terms ϵ_i to not be constant. As this violates the assumptions of OLS and as a consequence yields heteroscedastic results, I need to look for alternatives. I will subject the Equation 4.1 to five different methods of correcting heteroscedasticity. First, I use standard errors clustered at study level to account for potential within-study correlation. Second, I apply weighted least squares estimation (WLS). Third, I use instrumental variable (IV) estimation. And last but not least, I engage in a fixed effects (FE) estimation.

When using standard errors clustered at study level, we assume that standard errors are correlated when they are a part of the same study but uncorrelated otherwise. In other words, for error terms ϵ_{ij} denoting the effect i from study j we assume zero conditional mean , i.e.

$$E[\epsilon_{ij}|x_{ij}] = 0.$$

With clustering, for all $j \neq j'$:

$$E[\epsilon_{hj}\epsilon_{hj'}|x_{hj}, x_{hj'}] = 0,$$

where the vector of independent variables is x_{hj}.

The application of weighted least squares (WLS) will allow us to correct standard errors with weights and therefore help us to obtain efficient estimates. In WLS, I first weigh the formula given by Equation 4.1 by the inverse of standard errors. That way the weights are smaller for observations with larger error term variance and larger for observations with small error term variance as those observations carry more information. For the purpose of this analysis I call this method precision weighted least squares. The estimated model will be a transformation of the one given by Equation 4.1 as follows:

$$\frac{x_i}{SE_i} = \frac{\beta_0}{SE_i} + \beta_1 + u_i, \tag{4.2}$$

where $u_i = \frac{\epsilon_i}{SE_i}$ and $u_i \sim N(0, \sigma^2)$. While here the coefficients are reversed, β_1 still serves as a coefficient for publication bias as well as β_0 represents the true value of the effect. Although I already worked to eliminate heteroskedasticity, I will still estimate the model using clustering of standard errors to remove any remaining heteroskedasticity. The study weighted least squares estimation brings us the benefit of correcting heteroskedasticity in the baseline regression where the standard error of the estimate of gwg measures the degree of dispersion of the estimated adjusted gender wage gap(gwg), and the benefit of giving more weight to results with smaller standard errors.

Another approach to WLS would be to weigh the estimates by the inverse of the number of estimates reported per study. The disadvantage of non weighted regression is that the studies with a higher number of reported estimates are excessively represented compared to studies with few estimates. In order to give each study the same power to influence the results it makes sense to weigh the estimated by the inverse of the number of estimates reported per study. For the purpose of this analysis I call this method study weighted least squares.

Another useful method to tackle heteroskedasticity is to use instrumental variable estimation. The instrumental variable is set to be equal to the inverse of the square root of the number of reported estimates per study. That way the instrument is set to be correlated with the standard error but should be uncorrelated with the error terms. The chosen instrument is shown to be a strong one as the robust F-statistics in the first stage is 31.

The Panel A of Table 4.1 reports the publication bias tests results for the full sample of 661 estimates and both their bootstrapped and reported standard errors. The first employed test was an ordinary least squares estimation of the estimated adjusted gender wage gap on the approximate and reported standard errors. The slope of the coefficient is positive and greater than one, which according to publication selectivity classification by Doucouliagos and Stanley (2011) suggests 'substantial' selectivity. However, the coefficient is only significant at a 10% significance level and faces the drawbacks of OLS. The constant estimate in the publication bias tests can be interpreted as the mean estimate of the adjusted gender wage gap when corrected for any potential selective reporting. In this case, the constant coefficient in the OLS estimation is positive and statistically significant at even a 1% significance level. It is still smaller than the simple mean of all gwg estimates (0.153).

In contrast with other linear methods for publication bias testing, the OLS estimation yields the only statistically significant coefficients suggesting 'substantial' publication bias. While the instrumental variable estimation and study weighted least squares yield positive results of similar magnitude they are both statistically insignificant. Moreover, the results of the precision WLS yield very different results that the other linear methods. The estimated coefficient is still non-significant but negative, and the estimated constant coefficient yields a higher, statistically significant, estimate than what is the simple mean of the

Table 4.1: Publication Bias Tests - full sample

Panel A: Linear techniques				
	OLS	IV	Study	Precision
Standard error	1.171*	1.159	0.794	-1.509
	(0.661)	(1.212)	(1.208)	(1.963)
Constant	0.133***	0.135***	0.132***	0.162***
	(0.0130)	(0.0207)	(0.0305)	(0.0231)
Observations	661	661	661	661
Panel B: Between and Within				
	BE	FE	RE	
Standard error	1.228*	-0.0302	0.875	
	(0.694)	(0.651)	(0.568)	
Constant	0.128***	0.150***	0.136***	
	(0.0175)	(0.00897)	(0.0144)	
Observations	661	661	661	
Panel C: Nonlinear techniques				
	Stem	WAAP	Kink	
Effect beyond bias	0.183	0.153***	0.162***	
	(0.018)	(0.003)	(0.006)	
Observations	661	661	661	

adjusted gender wage gap (gwg). This non-intuitive result is possibly caused by the fact that the precision measure is bootstrapped. Apart from the precision weighted least squares, all of the constant estimates yield very similar results, both in magnitude and direction, suggesting that the mean estimate of the adjusted gender wage gap is somewhere around 0.13.

Alternatively, I also employ estimation to address both between- and within-study variation.

The fixed effects estimation captures the aspects that observations have in common within the same study but allows for variation across studies. That being said, I modify the model from Equation 4.1 accordingly:

$$x_{ij} = \beta_0 + \beta_1 \times SE_{ij} + \epsilon_{ij} + \nu_j, \tag{4.3}$$

where ν_j represents a vector of study specific characteristics. Also I assume that $\nu_j \sim N(0, \tau^2)$ and that it is independent of both ϵ_{ij} and SE_{ij}. Another assumption of the fixed effects model is, of course, the existence of one single true effect size that is estimated in all studies the same way - both population and variables used in estimation. The true effect is therefore a weighted average with accurate estimates being preferred to those that are inaccurate. The use of within-study variation can be problematic in meta-analyses because there are studies that report many estimates (possibly only as robustness checks) while other studies only report a few estimates. That way the identification of publication bias would be based on the studies with a larger number of estimates which does not make common sense. Next, I perform tests for between study variation and study-level random effects that combine both between and within study variation. The results suggest some (according to Doucouliagos and Stanley (2011) substantial) publication bias for between effects estimation while the fixed-effects and random-effects yield non-significant standard error coefficient estimates. The constant coefficient is positive and similar in size and direction to the results of the rest of the linear techniques.

I also employ several non-linear alternatives to the testing of publication bias. According to Stanley et al. (2010) there exist cases when the lienar relationship between the standard error and publication bias is defied. When looking back to the funnel plot in Figure 4.1, he suggest that the estimates portrayed at the top part of the funnel plot are less likely to be affected by publication bias than the ones at the bottom, due to their very small standard errors and high significance levels.

First, I perform a nonlinear technique called Stem-based method introduced by Furukawa (2019). It is a non-parametric method robustly selecting a number of the most precise studies from the 'stem of the funnel' based on a specific algorithm that minimizes the mean standard error: publication bias is diminshed in case only the most precise studies are used. Its wide confidence intervals make it a more conservative method in approaching the detection of publication bias. This method yields a relatively high, yet nonsignificant, mean gwg estimate compared to the linear methods.

The second nonlinear method I perform is the weighted average of adequately powered (WAAP) estimates by Ioannidis et al. (2017). This technique is based on the weighting the estimates by their average with proportional weights to the estimate's precision (inverse of standard error). This method focuses only on estimates with suitable statistical power. The designated statistical power for the test was chosen to be at 80%. Using this method I obtain a mean adjusted gender wage gap coefficient equal to 0.153 which is fairly close to the results of the linear techniques.

Lastly, the final nonlinear method I performed is the Endogenous kink model by Bom and Rachinger (2019). Same as the stem-based method by Furukawa (2019) it bases its calculations on estimating Monte Carlo simulations. Bom and Rachinger (2019) estimates the precision level at which studies are to be reported. The model assumes the highly precise estimates not to be biased and uses two linear segments that fit the model and refers to the joining place as the kink. The results suggest that the true mean estimate of gwg is 0.162 at a 1% significance level.

A new method p-uniform* an extension of p-uniform was developed by van Aert and van Assen (2018) and is mostly used in psychology but can be also applied in other areas. The core concept is the belief of p-values being uniformly distributed at the mean effect size. In other words, the estimated coefficient is equal to the underlying effect value when testing the hypothesis (Gechert et al. 2020). Only some values of p-values are affected with publication bias, usually regarding under-representation of large p-values or over-representation of p-values close below 0.05. The aim of this method is for the distribution of the p-values to be approximately uniform, this is achieved by finding a corresponding coefficient. While acknowledging this new method for publication bias detection, it is not feasible for the collected data and will not be employed for my analysis as it works with variance among standard errors within a study which, in the case of approximation of standard errors, is equal to zero in most cases.

In order to confirm that the approach of using bootstrapped standard errors is substantiated, I employ the same methods for publication bias detection on only the 135 estimates of the adjusted gender wage gap with reported standard errors obtained directly from the respective studies. From the linear methods, non of the standard error coefficients is significant. They also differ substantially in size but also in direction. As for the constant coefficient, all coefficients are statistically significant (at least at a 5% significance level) and are comparable in terms of both size and direction. This allows us to conclude that on the restricted dataset of 135 estimates I do not find evidence for the presence of publication bias and that the mean *gwg* estimate is positive and varies between 0.122 to 0.136 respective to the method used. The summarized results can be found in Appendix A Table A.2.

Both panel B and panel C of Table A.2 yield comparable results without any strong indication of publication bias but roughly the same estimated mean constant coefficient.

4.2 Why estimates vary

The next step of this thesis is to employ Bayesian model averaging which will help to indicate the best suitable model to estimate what variables have the most effect on the outcome of an analysis focusing on GWG. As I collected many different variables, many of which may only serve as controls, this tool will be very helpful.

Let us start with defining the key equation variables for an explanation of the BMA technique. Let D represent the data, N the number of explanatory variables and K the number of models. Also, $K = 2^N$. before observing any data, the beliefs about the model M_k are represented by $p(M_k)$, a marginal probability, which is the prior probability that M_k is the true model. The posterior probability, after observing the data, is derived from an extended Bayes' theorem:

$$p(M_k|D) = \frac{p(D|M_k)p(M_k)}{p(D)} = \frac{p(D|M_k)p(M_k)}{\sum_{m=1}^{K=2^N} p(D|M_m)p(M_m)}, \tag{4.4}$$

where the following Equation 4.5 represents the integrated likelihood of model M_k:

$$p(D|M_k) = \int p(D|\beta_k, M_k)p(\beta_k|M_k)d\beta_k. \tag{4.5}$$

β_k is the parameters vector of M_k, $p(\beta_k|M_k)$ is the β_k prior density under model M_k and the probability $p(D|\beta_k, M_k)$ is the conventional form of the likelihood. The likelihood of the model M_k in Equation 4.4 occurring given data D is represented by the conditional probability $p(M_k|D)$ that can also be called the posterior model probability because it serves as metric of the degree of belief in the model after accounting for data. After having considered all possibilities for M_k, the posterior distribution of GWG sizes given data D is:

$$p(gwg|D) = \sum_{k=1}^{K} p(gwg|M_k, D)p(M_k|D), \tag{4.6}$$

where the vector $M_1, ..., M_k$ represents the considered models. Equation 4.6 represents the overall posterior distribution showing the average of posterior distributions per each of the considered models weighted by the posterior model probability.

We can also derive the weighted expected value of gwg (the posterior mean) such that:

$$E[gwg|D] = \sum_{k=0}^{K} g\hat{w}g_k p(M_k|D), \tag{4.7}$$

where $g\hat{w}g_k$ is the expected gwg given data D and model M_k, i.e. $E[gwg|D, M_k]$. The derived posterior variance of the gwg is therefore:

$$Var[gwg|D] = \sum_{k=0}^{K} (Var[gwg|D, M_k] + g\hat{w}g_k{}^2)p(M_k|D) - E[gwg|D]^2. \tag{4.8}$$

In a few words, while the goodness-of-fit of the model is indicated by the posterior model probability, the prior model probability captures the prior beliefs of a researcher concerning the model probability before having considered the data (Zeugner 2011). To calculate a variable's posterior inclusion probability (PIP) we use the sum of all the posterior model probabilities (PMP) of models including that particular variable. The PIP is interpreted as a probability of that particular variable being a useful predictor for the dependent variable of the model.

In order for the BMA model to be complete, one needs to be able to calculate or approximate all the integrals, explore the model space and define priors on it and on the distribution of the coefficients β as well.

Before implementing the Bayesian estimation several issues in computation may have to be addressed.

The first issue, or rather a challenge, is the computation of of integrals in the presented likelihood function (Hoeting et al. 1999). Second issue lies in the excessively large model space that might be problematic to be computed on a standard computer. In this case, there are 2^{37} possible regressions, due to 37 identified explanatory variables. A solution to this computation challenge offers the application of the Metropolis-Hastings algorithm of the Markov chain Monte Carlo method. That way only the models with the highest PMP are estimated and this the computational load is diminished. The Metropolis-Hastings algorithm compares the benchmark model's PMP with another competing model 's PMP and thus determines the adequate models. If a competing model is accepted, it automatically becomes a benchmark model and is being compared with other competing models until another one is accepted (Zeugner 2011).

Next, one needs to define the prior distributions on the model probabilities and the regression parameters. Since my prior knowledge and beliefs regarding the model priors is rather small, I set a uniform prior probability for each model based on Eicher et al. (2009) as well as opting for the unit information prior (UIP). That way the information provided by UIP is corresponding to information of one observation in the dataset.

In order to keep the power of influence the same for all estimates, I weigh all of the estimates by the number of observations per study, that way none of the studies in over-represented. I run the model using the *bms* package in R.

In line with Steel (2020) I also implement frequentist model averaging (FMA). FMA does not require any prior specification. According to Hansen (2007) I estimate the estimator that by minimizing the Mallows criterion (Amini and Parmeter 2012) determines the weights. A smaller Mallows criterion can be interpreted as a better goodness-of-fit to a model. Moreover, this method reduces the number of explanatory variables from 2^{37} to 37.

4.2.1 Results

The results of the BMA analysis are reported in Figure 4.3. The figure visually represents the model inclusion of different explanatory variables. In the columns of Figure 4.3, there are the potential combinations of explanatory variables in a model scaled by their PMP on the horizontal axis. Going from the left hand side the columns are wider, suggesting a more likely model. The explanatory variables with the highest PIP are depicted at the top of the Figure 4.3, while the ones with the lowest PIP are at the bottom. In case the cells are non-coloured, it means that the particular variable would be excluded from the model. The coloured cells would be, on the other hand, included. The red cells (lighter in grayscale) suggest that a variable bears a negative coefficient direction in the regression, while blue cells (darker in grayscale) suggest a negative sign of a coefficient in the regression.

We can see from Figure 4.3 that the signs of the significant variables remain the same throughout top models which signalizes robust parameters in terms of inclusion of other variables. For BMA I weigh all the whole dataset by the inverse of number of observations per study to control for the unbalanced data where the number of observations per study ranges from 2 to 104.

Figure 4.3: Bayesian Model Averaging: Model Inclusion

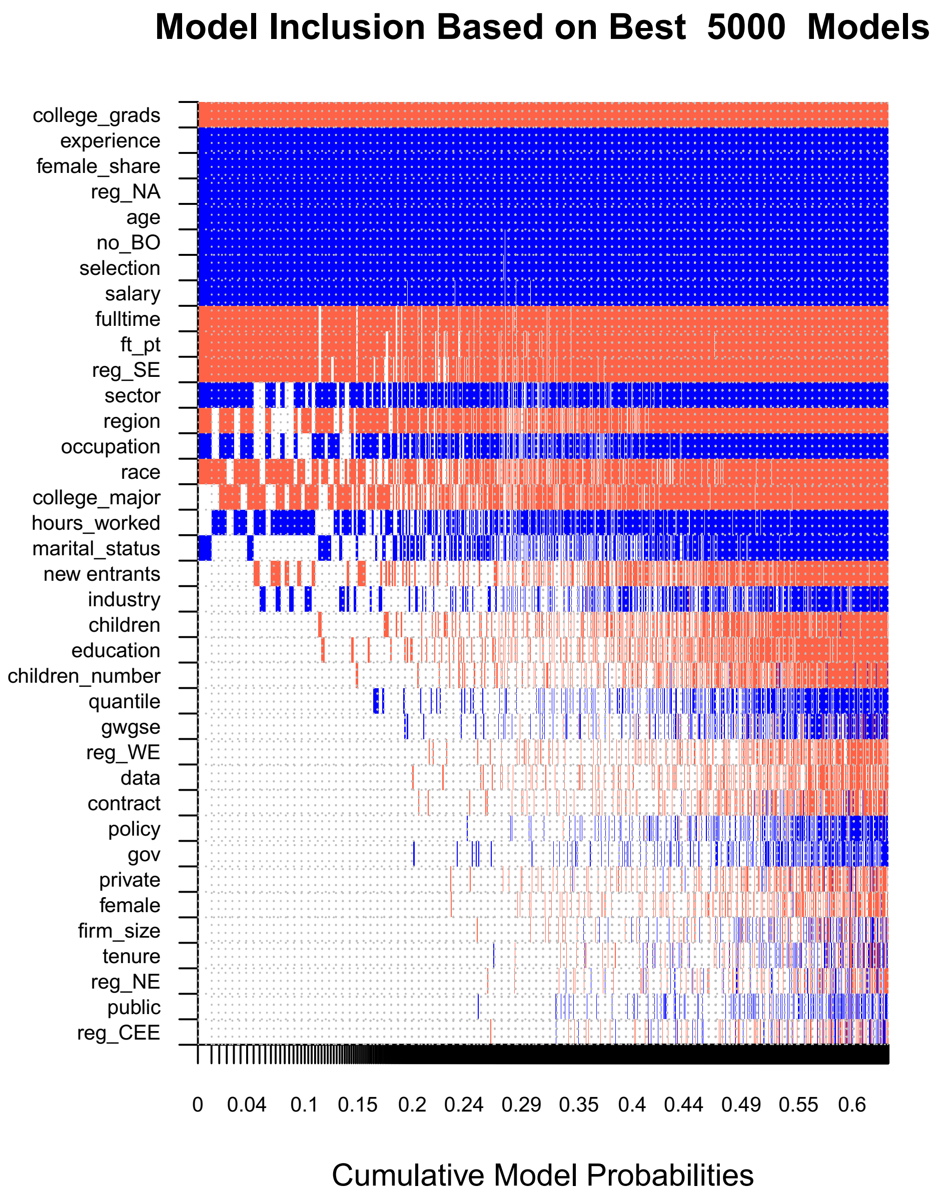


Table 4.2: Coefficient Estimates

Variable	PIP	Post. Mean	Post. SD	Cond. Pos. Sign
Standard Error	0.054	0.011	0.068	0.838
Study specific variables				
College graduates	1.000	-0.071	0.014	0.000
Data	0.044	-0.001	0.004	0.000
Female	0.027	0.000	0.002	0.017
Fulltime	0.910	-0.04	0.018	0.000
Government	0.033	0.001	0.010	1.000
New Entrants	0.236	-0.008	0.016	0.000
Policy	0.034	0.001	0.004	0.939
Private	0.031	0.000	0.003	0.073
Public	0.019	0.000	0.006	0.968
Central & Eastern Europe	0.018	0.000	0.003	0.339
North America	1.000	0.082	0.014	1.000
Northern Europe	0.019	0.000	0.003	0.312
Southern Europe	0.801	-0.045	0.027	0.000
Western Europe	0.044	-0.001	0.004	0.004
Methodological aspect				
Blinder-Oaxaca	0.997	0.045	0.009	1.000
Quantile Regression	0.082	0.003	0.013	1.000
Selection	0.987	0.048	0.013	1.000
Regression specific variables				
Age	1.000	0.051	0.012	1.000
College Major	0.582	-0.02	0.02	0.000
Contract	0.038	-0.001	0.004	0.047
Education	0.132	-0.004	0.011	0.000
Experience	1.000	0.099	0.013	1.000
Female Share	1.000	0.065	0.014	1.000
Firm Size	0.021	0.000	0.002	0.409
FT-PT	0.855	-0.041	0.021	0.000
Hours Worked	0.563	0.023	0.023	1.000
Children	0.144	-0.005	0.014	0.007
Children Number	0.095	-0.003	0.011	0.015
Industry	0.178	0.005	0.012	0.99
Marital Status	0.352	0.011	0.017	1.000
Occupation	0.625	0.017	0.016	1.000
Race	0.599	-0.022	0.021	0.000
Region	0.682	-0.023	0.018	0.000
Salary	0.978	0.042	0.014	1.000
Sector	0.733	0.024	0.018	1.000

To be continued on the next page

Table 4.2: Coefficient Estimates - continued

Variable	PIP	Post. Mean	Post. SD	Cond.Pos. Sign
Tenure	0.020	0.000	0.002	0.402

The numerical results of BMA are shown in Table 4.2. The numerical results provide information about posterior inclusion probability, the posterior mean and standard deviation and conditional posterior sign for all the considered variables. While the interpretation of the PIP needs more context, the interpretation of the latter three is intuitive just as an interpretation of standard regressions. For the interpretation of the posterior inclusion probability, we follow the guidelines offered by Kass and Raftery (1995) to be able to evaluate the importance of each explanatory variable:

- $0.5 < PIP < 0.75$: weak effect,
- $0.75 < PIP < 0.95$: substantial effect,
- $0.95 < PIP < 0.99$: strong effect,
- $0.99 < PIP$: decisive effect.

I will follow these guidelines in the interpretation of estimated results of our BMA analysis.

As in Chapter 3, I divide the variables into three main categories: *Study specific variables, Methodological aspects*, and *Regression specific variables.* In Table 4.2, the variables are sorted alphabetically within the respective category. Each of the three categories contains variables that do have at least a substantial effect (Kass and Raftery 1995) on the individual estimates and therefore help us with explaining the variation among the collected gender wage gap estimates.

In my analysis, I am mainly interested in variables that have at least a substantial effect on the *gwg* estimates. therefore, I focus on variables whose PIP is greater than 0.75. From the *study specific variables* it is the case of 4 variables whose estimated effect is further described in the following paragraphs:

College Graduates. Conducting a study with a focus on solely fresh college graduates has a negative effect on the outcome of the gender wage gap. In particular, researchers of such studies estimate a 0.07 smaller *gwg* than others. The intuition behind this results is quite straightforward. We would expect that college graduates are more likely to have the same pay due to same opportunities, same age, and same amount of experience (which is most likely to be zero at this point of life).

Geographical variables. From Figure 4.3 we may notice that two of the geographical variables ended up being important to the estimation of the gender wage gap: *Northern America* and *Southern Europe*. Each effect has a different sign. While the coefficient for the region of Northern America is positive, making estimates of the gender pay gap originating in this region larger, the coefficient for Southern Europe has a negative sign, making the Southern European estimates smaller. A smaller gender wage estimate for Southern Europe does not necessarily imply a more equal labour market. It can be also caused by the fact, that the women employment rate in countries with a lower *gwg* is also lower. In such countries women with smaller earning potential tend to leave or to never enter the labour market at all.

Fulltime. Another variable that has a substantial effect on the gender pay gap estimates is the author's choice to focus solely on full-time workers, rather than controlling for other forms of employment. Omitting part-time workers means omitting a large portion of work force that decided to either lower their working hours or simply divide them between more employers. The resulting coefficient has a negative sign, signaling that focusing only on full-time workers obtains estimates that are on average lower by 0.04.

Methodology. In terms of methodological approach, there are two variables that account for variation between estimates of the gender wage gap. First, when researchers use other methods than Blinder-Oaxaca decomposition for decomposing the gender pay gap, they obtain higher estimates than researchers sticking to this method or its extensions. Then, also the researchers who do not control for selection bias obtain 0.05 higher gender wage gap estimates.

In the regression specific variables I identify 5 variables with PIP higher than 0.75 as per Kass and Raftery (1995). All those variables are specific to different regressions within studies, so they may have been used as robustness checks. Nevertheless, the serve us to analyze what variables have the most effect on the gender wage gap estimate.

Age and Experience. Considering the variables *Age* and *Experience* to be very close in nature, I dedicate to them a shorter common paragraph. Omitting each of these variables in a regression brings higher resulting gender wage gap estimates. The explanation behind this is rather straightforward. Without considering the age or the years of experience of a worker, it is rather infeasible to compare their earnings. Since women and men of the same age tend to have different years of experience (most commonly due to maternal leaves) it makes common sense to include these variables for a more accurate gender wage gap estimate.

Other regression specific variables. The regressions that do not take into account female shares within a firm estimate almost 0.08 higher adjusted gender wage gaps than those that do. In firms with higher female share, the gender pay gap tends to be smaller (Simón 2012). Most competitive positions in competitive firms are occupied by men, which intuitively makes space for a larger gender pay gap. Overall, regressions that omit some control variables do naturally affect the resulting estimates. In this case, the aim si to explain the most part of the gender wage gap, omitting any variable would therefore indicate a space for potential higher estimated wage gap. However, the omission of distinguishing between full-time and part-time workers leads to gender wage gap estimates that are almost 0.04 lower.

A robustness check in the form of frequentist model averaging was applied to the data as well. As the Mallows method (Hansen 2007) reduces the model space to only 37 different models, the first model is only the estimate regressed on the first dependent variable. The order in which variables enter the model is determined by the category-level groupings I established in Chapter 3: Methodology and Data: : *Study specific variables, Methodological aspects*, and *Regression specific variables.*

Table 4.3: Robustness check: Coefficient Estimates from FMA

Variable	Coef.	Sd. E.	p-value
Standard Error	0.480	0.147	0.016
Study specific variables			
College graduates	-0.066	0.014	0.000
Data	-0.014	0.017	0.397
Female	0.011	0.014	0.446

To be continued on the next page

Table 4.3: Robustness check: Coefficient Estimates from FMA - continued

Variable	Coef.	Sd. E.	p-value
Fulltime	-0.045	0.013	0.001
Government	0.059	0.040	0.143
New Entrants	-0.011	0.016	0.513
Policy	0.063	0.022	0.050
Private	-0.016	0.016	0.296
Public	-0.014	0.046	0.757
Central & Eastern Europe	0.019	0.025	0.444
North America	0.059	0.018	0.001
Northern Europe	-0.018	0.021	0.379
Southern Europe	-0.059	0.020	0.004
Western Europe	-0.044	0.015	0.003
Methodological aspect			
Binder-Oaxaca	0.033	0.011	0.002
Quantile Regression	0.029	0.011	0.020
Selection	0.033	0.016	0.836
Regression specific variables			
Age	0.0441	0.014	0.001
College Major	-0.031	0.014	0.027
Contract	-0.005	0.017	0.785
Education	-0.022	0.016	0.148
Experience	0.123	0.014	0.000
Female Share	0.070	0.016	0.000
Firm Size	-0.027	0.017	0.123
FT-PT	-0.069	0.017	0.000
Hours Worked	0.043	0.016	0.008
Children	0.012	0.038	0.755
Children Number	-0.031	0.041	0.444
Industry	0.017	0.014	0.231
Marital Status	0.040	0.015	0.005
Occupation	0.039	0.013	0.002
Race	-0.036	0.015	0.013
Region	-0.041	0.012	0.001
Salary	0.041	0.014	0.003
Sector	0.047	0.014	0.001
Tenure	0.010	0.015	0.504

While there are some estimates that are not in terms of significance in line with the BMA approach. The majority of the FMA estimates suggest results with similar significance, magnitude, and direction of the variables effects on the estimated gender wage gap (as per Table 4.3). The FMA considers a larger number of dependent variables to have an effect on the estimated gender wage gap than the BMA technique. Other than that, the results seem to be robust.

Chapter 5

Conclusion

This thesis applies the modern methodology of meta-analysis on a sample of collected estimates and parameters from literature that evaluates the gender pay gap in countries of the developed world. Meta-analysis is a convenient tool to assess large portions of published data quantitatively with the aim of bias reduction to a minimum.

The available literature for the topic of gender wage differences has been quite thorough in the past decades. However, the underlying data, used methods and approach tend to be very diverse. While a group of authors focuses on one segment of the labour market, other authors focus on another. The aim of this thesis is to evaluate and summarize the diverse state of the art literature predominantly in a quantitative manner. This would not be the first attempt to summarize the literature on gender wage gap differential quantitatively. The two most recent works were written by Weichselbaumer and Winter-Ebmer (2005) and Konstantinova (2020). However, the meta-analysis in this thesis is constructed based on the most modern methods and is practically unprecedented in terms of publication bias detection and modern methods and weighing techniques for model averaging.

None of the existing meta-analysis have employed a test for the presence of publication bias in the available papers. To perform a publication bias analysis, it is ideal to obtain the standard errors together with the estimates. While a large number of authors does not report a measure of uncertainty along with their adjusted gender pay gap estimates, I decided to approximate them in order to utilize the largest possible number of reported estimates in the existing literature.

Moreover, in the data collection of meta-analysis, the presence of standard errors usually serves as an indicator of a suitable estimate. Since such a judgment could not be used in this case, I was compelled to find an alternative approach. As opposed to the existing meta-analyses, I decided to include only studies that were peer-reviewed and placed more emphasis on the methodology used to decompose the gender pay gap (including the recent method of quantile regression (Machado and Mata 2005)). The final dataset contains 661 estimates of the adjusted gender wage gap from 51 different studies. The included papers were published between the years 2003 and 2021, all of the included estimates are based on data from developed countries.

After the data collection and the subsequent approximation of the potentially missing standard errors, I was able to execute a number of tests to detect publication bias. My first step was to start with tests using the funnel plot, which is a visual tool that serves to detect the presence of bias in literature. Nonetheless, since this technique can be interpreted in a subjective manner, I proceeded with more formal tests (with robustness checks on the restricted dataset of estimates with originally reported standard errors), in particular I carry out estimation techniques of OLS, WLS, instrumental variable estimation, and within- and between-study variation. Moreover, I employ some adequate and up-to-date non-linear techniques to confirm that in fact, the literature does not show any significant publication bias, which is a very rare discovery in economics.

The apparent absence of publication bias can be justified. In the general public, it is possible to meet advocates of non-existent gender wage gap, but also advocates of the view that gender wage differentials are high and that the situation should be dealt with accordingly. There is nothing to suggest that the professional community should not be undivided in this regard either. From my point of view, advocates of persisting gender pay gap need research that confirms their beliefs while the other group needs to confirm the opposite. In this manner, the publication bias would remain undetected even if it was in fact present.

Consequently, in order to determine the main factors of heteregeneity of gender wage gap estimates, I employ the Bayesian model averaging method and a robustness check in the form of frequentist model averaging. The results of both of these methods are essentially in line with each other.

So what causes estimates to vary across studies? Rationally, with the exclusion of certain variables that describe the human capital of workers, the reported gender wage gap increases. such variables include *age, experience,* *sector*, etc. Moreover, authors who do not control for selection bias or employ other methods that the widely used Blinder-Oaxaca decomposition in their papers report larger gender wage gap estimates.

Still, there are aspects that cause the gender wage gap estimates to be on average lower. For the most part, such results are typical of studies that focus on only one industry or a particular group of employees. In this thesis, I identified that focusing solely on full-time workers and college graduates yields on average lower gender wage gap estimates.

The scope of this thesis has its limitations. For further research I would suggest that more attention is given to collected quantile regression estimates and their inclusion in the meta-analysis. As the use of quantile regression is becoming more and more common, I would expect that any future meta-analysis on this topic will consist predominantly of quantile regression estimates. However, since quantile regression estimates only form a nonsignificant part of the collected data, not as much attention has been dedicated to them here.

Bibliography

Adamchik, V., Hyclak, T. and King, A. (2003), 'The wage structure and wage distribution in poland, 1994-2001', *International Journal of Manpower* **24**(8), 916–946.

Addabbo, T. and Favaro, D. (2011), 'Gender wage differentials by education in italy', *Applied Economics* **43**(29), 4589–4605.

Albrecht, J., Van Vuuren, A. and Vroman, S. (2004), 'Decomposing the gender wage gap in the netherlands with sample selection adjustments', *Available at SSRN 621064* .

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

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

Arulampalam, W., Booth, A. L. and Bryan, M. L. (2007), 'Is there a glass ceiling over europe? exploring the gender pay gap across the wage distribution', *ILR Review* **60**(2), 163–186.

Atal, J., Ñopo, H. and Winder, N. (2009), 'New century, old disparities: gender and ethnic wage gaps in latin america'.

Avlijaš, S., Ivanović, N., Vladisavljević, M. and Vujić, S. (2013), *Gender pay gap in the Western Balkan countries: evidence from Serbia, Montenegro and Macedonia*, FREN-Foundation for the Advancement of Economics.

Barnard, A. (2008), 'Modelling the gender pay gap in the uk: 1998 to 2006', *Economic & Labour Market Review* **2**(8), 18–24.

Barón, J. D. and Cobb-Clark, D. A. (2010), 'Occupational segregation and the gender wage gap in private-and public-sector employment: a distributional analysis', *Economic Record* **86**(273), 227–246.

Bayard, K., Hellerstein, J., Neumark, D. and Troske, K. (2003), 'New evidence on sex segregation and sex differences in wages from matched employee-employer data', *Journal of labor Economics* **21**(4), 887–922.

Becker, G. S. (1957), *The economics of discrimination*, University of Chicago press.

Bensidoun, I., Trancart, D. and Dutreuilh, C. (2018), 'Career choices and the gender pay gap: The role of work preferences and attitudes', *Population* **73**(1), 35–60.

Bertrand, M., Goldin, C. and Katz, L. F. (2010), 'Dynamics of the gender gap for young professionals in the financial and corporate sectors', *American economic journal: applied economics* **2**(3), 228–55.

Biewen, M., Fitzenberger, B. and Seckler, M. (2020), 'Counterfactual quantile decompositions with selection correction taking into account huber/melly (2015): An application to the german gender wage gap', *Labour Economics* **67**, 101927.

Black, D. A., Haviland, A. M., Sanders, S. G. and Taylor, L. J. (2008), 'Gender wage disparities among the highly educated', *Journal of human resources* **43**(3), 630–659.

Blau, F. D. and Kahn, L. M. (2006), 'The us gender pay gap in the 1990s: Slowing convergence', *Ilr Review* **60**(1), 45–66.

Blau, F. D. and Kahn, L. M. (2013), 'Female labor supply: Why is the united states falling behind?', *American Economic Review* **103**(3), 251–56.

Blau, F. D. and Kahn, L. M. (2017), 'The gender wage gap: Extent, trends, and explanations', *Journal of Economic Literature* **55**(3), 789–865.

Blinder, A. S. (1973), 'Wage discrimination: reduced form and structural estimates', *Journal of Human resources* pp. 436–455.

Boll, C. and Lagemann, A. (2018), 'Gender pay gap in eu countries based on ses (2014)', *Publications Office of the European Union* **10**, 978935.

Bom, P. R. D. and Rachinger, H. (2019), 'A kinked meta-regression model for publication bias correction', *Research Synthesis Methods* **10**(4), 497–514.

Boraas, S. and Rodgers III, W. M. (2003), 'How does gender play a role in the earnings gap-an update', *Monthly Lab. Rev.* **126**, 9.

Brenan, M. (2021), 'Women still handle main household tasks in u.s.'.
URL: *https://news.gallup.com/poll/283979/women-handle-main-household-tasks.aspx*

Campos-Soria, J. A. and Ropero-García, M. A. (2016), 'Occupational segregation and the female–male wage differentials: Evidence for spain', *Gender Issues* **33**(3), 183–217.

Castagnetti, C. and Giorgetti, M. L. (2019), 'Understanding the gender wage-gap differential between the public and private sectors in italy: A quantile approach', *Economic Modelling* **78**, 240–261.

Castagnetti, C., Rosti, L. and Töpfer, M. (2017), 'The convergence of the gender pay gap: An alternative estimation approach'.

Chevalier, A. (2004), 'Motivation, expectations and the gender pay gap for uk graduates', *Available at SSRN 527125* .

Chevalier, A. (2007), 'Education, occupation and career expectations: determinants of the gender pay gap for uk graduates', *Oxford Bulletin of Economics and Statistics* **69**(6), 819–842.

Christofides, L. N., Polycarpou, A. and Vrachimis, K. (2013), 'Gender wage gaps,â€sticky floorsâ€™ and â€glass ceilingsâ€™ in europe', *Labour Economics* **21**, 86–102.

Cotton, J. (1988), 'On the decomposition of wage differentials', *The review of economics and statistics* pp. 236–243.

Cukrowska-Torzewska, E. and Matysiak, A. (2020), 'The motherhood wage penalty: A meta-analysis', *Social Science Research* **88-89**, 102416.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G. G. (2005), 'Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey', *SSRN Electronic Journal* .

Doucouliagos, C. and Stanley, T. (2011), 'ARE ALL ECONOMIC FACTS GREATLY EXAGGERATED? THEORY COMPETITION AND SELECTIVITY', *Journal of Economic Surveys* **27**(2), 316–339.

Eicher, T. S., Papageorgiou, C. and Raftery, A. E. (2009), 'Default priors and predictive performance in bayesian model averaging, with application to growth determinants', *Journal of Applied Econometrics* **26**(1), 30–55.

Elder, T. E., Goddeeris, J. H. and Haider, S. J. (2010), 'Unexplained gaps and oaxaca–blinder decompositions', *Labour Economics* **17**(1), 284–290.

Fernandes, A. P. and Ferreira, P. (2021), 'Executivesâ€™ gender pay gap and financing constraints', *Journal of Economic Behavior & Organization* **192**, 381–404.

Flinn, C. J., Todd, P. E. and Zhang, W. (2018), 'Personality traits, intra-household allocation and the gender wage gap', *European Economic Review* **109**, 191–220.

Fortin, N. M. (2008), 'The gender wage gap among young adults in the united states the importance of money versus people', *Journal of Human Resources* **43**(4), 884–918.

Fortin, N. M., Bell, B. and Böhm, M. (2017), 'Top earnings inequality and the gender pay gap: Canada, sweden, and the united kingdom', *Labour Economics* **47**, 107–123.

Francesconi, M. and Parey, M. (2018), 'Early gender gaps among university graduates', *European Economic Review* **109**, 63–82.

Fransen, E., Plantenga, J. and Vlasblom, J. D. (2012), 'Why do women still earn less than men? decomposing the dutch gender pay gap, 1996–2006', *Applied Economics* **44**(33), 4343–4354.

Furukawa, C. (2019), 'Publication bias under aggregation frictions: From communication model to new correction method'.

Gallen, Y., Lesner, R. V. and Vejlin, R. (2019), 'The labor market gender gap in denmark: Sorting out the past 30 years', *Labour Economics* **56**, 58–67.

Gardeazabal, J. and Ugidos, A. (2005), 'Gender wage discrimination at quantiles', *Journal of population economics* **18**(1), 165–179.

Gayle, G.-L., Golan, L. and Miller, R. A. (2012), 'Gender differences in executive compensation and job mobility', *Journal of Labor Economics* **30**(4), 829–872.

Gechert, S., Havranek, T., Irsova, Z. and Kolcunova, D. (2020), 'Measuring capital-labor substitution: The importance of method choices and publication bias', *An online appendix* pp. 1–49.

Goldin, C. (2014), 'A grand gender convergence: Its last chapter', *American Economic Review* **104**(4), 1091–1119.

Grove, W. A., Hussey, A. and Jetter, M. (2011), 'The gender pay gap beyond human capital heterogeneity in noncognitive skills and in labor market tastes', *Journal of Human Resources* **46**(4), 827–874.

Hansen, B. E. (2007), 'Least squares model averaging', *Econometrica* **75**(4), 1175–1189.

Havranek, T., Irsova, Z., Janda, K. and Zilberman, D. (2015), 'Selective reporting and the social cost of carbon', *Energy Economics* **51**, 394–406.

Havránek, T., Stanley, T. D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, J., Iwasaki, I., Reed, W. R., Rost, K. and Aert, R. C. M. (2020), 'REPORTING GUIDELINES FOR META-ANALYSIS IN ECONOMICS', *Journal of Economic Surveys* **34**(3), 469–475.

Havránková, Z. (2015), 'Six essays on meta-regression analysis'.

Healy, G. and Ahamed, M. M. (2019), 'Gender pay gap, voluntary interventions and recession: the case of the british financial services sector', *British Journal of Industrial Relations* **57**(2), 302–327.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), 'Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and e. i. george, and a rejoinder by the authors', *Statistical Science* **14**(4).

Ioannidis, J. P. A., Stanley, T. D. and Doucouliagos, H. (2017), 'The power of bias in economics research', *The Economic Journal* **127**(605), F236–F265.

Jagsi, R., Guancial, E. A., Worobey, C. C., Henault, L. E., Chang, Y., Starr, R., Tarbell, N. J. and Hylek, E. M. (2006), 'The â€śgender gapâ€ť in authorship of academic medical literatureâ€"a 35-year perspective', *New England Journal of Medicine* **355**(3), 281–287.

Jarrell, S. B. and Stanley, T. D. (1990), 'A meta-analysis of the union-nonunion wage gap', *ILR Review* **44**(1), 54–67.

Jarrell, S. B. and Stanley, T. D. (2004), 'Declining bias and gender wage discrimination? a meta-regression analysis', *Journal of Human Resources* **39**(3), 828–838.

Jones, M., Makepeace, G. and Wass, V. (2018), 'The uk gender pay gap 1997–2015: What is the role of the public sector?', *Industrial Relations: A Journal of Economy and Society* **57**(2), 296–319.

Joy, L. (2003), 'Salaries of recent male and female college graduates: Educational and labor market effects', *ILR Review* **56**(4), 606–621.

Jurajda, Ŝ. (2005), 'Gender segregation and wage gap: An east-west comparison', *Journal of the European Economic Association* **3**(2-3), 598–607.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kee, H. J. (2006), 'Glass ceiling or sticky floor? exploring the australian gender pay gap', *Economic Record* **82**(259), 408–427.

Konstantinova, E. (2020), 'A meta-analytic approach to gender pay gap: A case of discrimination?', *Master's Thesis, Charles University, Faculty of Social Sciences* p. 50p.

Korkeamäki, O. and Kyyrä, T. (2006), 'A gender wage gap decomposition for matched employer–employee data', *Labour Economics* **13**(5), 611–638.

Krueger, A. and Card, D. (1995), 'Time-series minimumwage studies: A meta-analysis', *American Economic Review* **85**, 238–43.

Le, T. A. N., Sasso, A. T. L. and Vujicic, M. (2017), 'Trends in the earnings gender gap among dentists, physicians, and lawyers', *The Journal of the American Dental Association* **148**(4), 257–262.

Lechmann, D. S. and Schnabel, C. (2012), 'What explains the gender earnings gap in self-employment? a decomposition analysis with german data'.

Leythienne, D. and Ronkowski, P. (2018), *A decomposition of the unadjusted gender pay gap using structure of earnings survey data: 2018 edition.*, Publications Office.

Litman, L., Robinson, J., Rosen, Z., Rosenzweig, C., Waxman, J. and Bates, L. M. (2020), 'The persistence of pay inequality: The gender pay gap in an anonymous online labor market', *PloS one* **15**(2), e0229383.

Livanos, I. and Pouliakas, K. (2009), 'The gender wage gap as a function of educational degree choices in an occupationally segregated eu country'.

Machado, J. A. F. and Mata, J. (2005), 'Counterfactual decomposition of changes in wage distributions using quantile regression', *Journal of Applied Econometrics* **20**(4), 445–465.

Madden, D. (2000), 'Towards a broader explanation of male-female wage differences', *Applied Economics Letters* **7**(12), 765–770.

Madden, J. F. (2012), 'Performance-support bias and the gender pay gap among stockbrokers', *Gender & Society* **26**(3), 488–518.

Mandel, H. and Semyonov, M. (2014), 'Gender pay gap and employment sector: Sources of earnings disparities in the united states, 1970–2010', *Demography* **51**(5), 1597–1618.

Manning, A. and Swaffield, J. (2008), 'The gender gap in early-career wage growth', *The Economic Journal* **118**(530), 983–1024.

McDonald, J. A. and Thornton, R. J. (2007), 'Do new male and female college graduates receive unequal pay?', *Journal of Human Resources* **42**(1), 32–48.

Meara, K., Pastore, F. and Webster, A. (2020), 'The gender pay gap in the usa: a matching study', *Journal of Population Economics* **33**(1), 271–305.

Melly, B. (2005), 'Public-private sector wage differentials in germany: Evidence from quantile regression', *Empirical Economics* **30**(2), 505–520.

Meng, X. (2004), 'Gender earnings gap: the role of firm specific effects', *Labour Economics* **11**(5), 555–573.

Miller, P. W. (2009), 'The gender pay gap in the us: does sector make a difference?', *Journal of Labor Research* **30**(1), 52–74.

Mincer, J. and Polachek, S. (1974), 'Family investments in human capital: Earnings of women', *Journal of political Economy* **82**(2, Part 2), S76–S108.

Mueller, G. and Plug, E. (2006), 'Estimating the effect of personality on male and female earnings', *Ilr Review* **60**(1), 3–22.

Mysíková, M. et al. (2012), 'Gender wage gap in the czech republic and central european countries', *Prague economic papers* **21**(3), 328–346.

Neal, D. (2002), The measured black-white wage gap among women is too small, Technical report.

Neumark, D. (1988), 'Employers: Discriminatory behavior and the estimation of wage discrimination', *The Journal of Human Resources* **23**(3), 279.

Nicodemo, C. (2009), 'Gender pay gap and quantile regression in european families'.

Nyhus, E. K. and Pons, E. (2012), 'Personality and the gender wage gap', *Applied Economics* **44**(1), 105–118.

OECD (2018), 'Distribution of employment by aggregate sectors, by sex'.
URL: *https://stats.oecd.org/index.aspx?queryid=54755*

Olivetti, C. and Petrongolo, B. (2008), 'Unequal pay or unequal employment? a cross-country analysis of gender gaps', *Journal of Labor Economics* **26**(4), 621–654.

O'Neill, J. and Polachek, S. (1993), 'Why the gender gap in wages narrowed in the 1980s', *Journal of Labor economics* **11**(1, Part 1), 205–228.

Ospino, C. G., Roldan Vasquez, P. and Barraza Narvaez, N. (2010), 'Oaxaca-blinder wage decomposition: Methods, critiques and applications. a literature review', *Revista de economía del Caribe* (5), 237–274.

Pearson, K. (1904), '3. mathematical contributions to the theory of evolution.-12. on a generalised theory of alternative inheritance, with special reference to mendel's laws', *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **203**(359-371), 53–86.

Picchio, M. and Mussida, C. (2011), 'Gender wage gap: A semi-parametric approach with sample selection correction', *Labour Economics* **18**(5), 564–578.

Polachek, S. W. (1981), 'Occupational self-selection: A human capital approach to sex differences in occupational structure', *The review of Economics and Statistics* pp. 60–69.

Raftery, A. E. (1995), 'Bayesian model selection in social research', *Sociological methodology* pp. 111–163.

Redmond, P. and McGuinness, S. (2019), 'The gender wage gap in europe: Job preferences, gender convergence and distributional effects', *Oxford Bulletin of Economics and Statistics* **81**(3), 564–587.

Reimer, D. and Schröder, J. (2006), 'Tracing the gender wage gap: Income differences between male and female university graduates in germany', *Zeitschrift für ArbeitsmarktForschung-Journal for Labour Market Research* **39**(2), 235–253.

Rosenthal, R. (1979), 'The file drawer problem and tolerance for null results.', *Psychological bulletin* **86**(3), 638.

Rosenzweig, M. R. and Morgan, J. (1976), 'Wage discrimination: A comment', *The Journal of Human Resources* **11**(1), 3–7.

Russell, H., Smyth, E. and O'Connell, P. J. (2010), 'Gender differences in pay among recent graduates: private sector employees in ireland', *Journal of Youth studies* **13**(2), 213–233.

Simón, H. (2012), 'The gender gap in earnings: an international comparison with european matched employer–employee data', *Applied Economics* **44**(15), 1985–1999.

Stanley, T. D. (2001), 'Wheat from chaff: Meta-analysis as quantitative literature review', *Journal of economic perspectives* **15**(3), 131–150.

Stanley, T. D. (2005), 'Beyond publication bias', *Journal of economic surveys* **19**(3), 309–345.

Stanley, T. D., Jarrell, S. B. and Doucouliagos, H. (2010), 'Could it be better to discard 90% of the data? a statistical paradox', *The American Statistician* **64**(1), 70–77.

Stanley, T., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. S. and Rost, K. (2013), 'META-ANALYSIS OF ECONOMICS RESEARCH REPORTING GUIDELINES', *Journal of Economic Surveys* **27**(2), 390–394.

Stanley, T. and Jarrell, S. B. (1998), 'Gender wage discrimination bias? a meta-regression analysis', *The Journal of Human Resources* **33**(4), 947.

Steel, M. F. (2020), 'Model averaging and its use in economics', *Journal of Economic Literature* **58**(3), 644–719.

Sutton, A. J. (2000), 'Empirical assessment of effect of publication bias on meta-analyses', *BMJ* **320**(7249), 1574–1577.

Tharp, D. T., Lurtz, M., Mielitz, K. S., Kitces, M. and Ammerman, D. A. (2019), 'Examining the gender pay gap among financial planning professionals: A blinder-oaxaca decomposition', *Financial Planning Review* **2**(3-4), e1061.

Thornton, A. and Lee, P. (2000), 'Publication bias in meta-analysis: its causes and consequences', *Journal of clinical epidemiology* **53**(2), 207–216.

Triventi, M. (2013), 'The gender wage gap and its institutional context: a comparative analysis of european graduates', *Work, Employment and Society* **27**(4), 563–580.

TuáşĄn, A. T. T. (2018), 'Investigating the gender wage gap in vietnam by quantile regression: Sticky floor or glass ceiling', *Journal of Asian Business and Economic Studies* **25**(S01), 04–23.

van Aert, R. C. and van Assen, M. (2018), 'Correcting for publication bias in a meta-analysis with the p-uniform* method', *Manuscript submitted for publication Retrieved from: https://osfio/preprints/bitss/zqjr92018.[Google Scholar]* .

Watson, I. et al. (2010), 'Decomposing the gender pay gap in the australian managerial labour market', *Australian Journal of Labour Economics* **13**(1), 49.

Weichselbaumer, D. and Winter-Ebmer, R. (2005), 'A meta-analysis of the international gender wage gap', *Journal of Economic Surveys* **19**(3), 479–511.

Xue, X., Reed, W. R. and Menclova, A. K. (2019), Social Capital and Health: A Meta-Analysis, Working Papers in Economics 19/18.
URL: *https://ideas.repec.org/p/cbt/econwp/19-18.html*

Zajkowska, O. et al. (2013), 'Gender pay gap in poland–blinder-oaxaca decomposition', *Metody Ilościowe w Badaniach Ekonomicznych* **14**(2), 272–278.

Zeugner, S. (2011), 'Bayesian model averaging with bms', *Tutorial to the R-package BMS 1e30* .

Appendix A

Additional tests and regressions

Table A.1: Summary statistics of regression variables: weighted

Variable	Mean	St. d.
gwg	0.155	0.088
Standard Error	0.022	0.013
Study specific variables		
College graduates	0.278	0.448
Data	0.885	0.319
Female	0.233	0.423
Fulltime	0.244	0.43
Government	0.059	0.236
New Entrants	0.148	0.355
Policy	0.878	0.328
Private	0.128	0.334
Public	0.03	0.169
Central & Eastern Europe	0.079	0.27
North America	0.203	0.403
Northern Europe	0.071	0.257
Southern Europe	0.082	0.274
Western Europe	0.408	0.492
Methodological aspect		
Blinder-Oaxaca	0.368	0.483
Quantile Regression	0.096	0.295
Selection	0.803	0.398
Regression specific variables		
Age	0.573	0.495
College Major	0.859	0.348
Contract	0.744	0.437
Education	0.392	0.489

To be continued on the next page

Table A.1: Summary statistics of regression variables: weighted - continued

Variable	Description	St. d.
Experience	0.55	0.498
Female Share	0.884	0.321
Firm Size	0.736	0.441
FT-PT	0.642	0.48
Hours Worked	0.793	0.405
Children	0.782	0.413
Children Number	0.848	0.359
Industry	0.724	0.447
Marital Status	0.684	0.465
Occupation	0.631	0.483
Race	0.779	0.415
Region	0.698	0.459
Salary	0.233	0.423
Sector	0.642	0.48
Tenure	0.709	0.454

Figure A.1: Funnel Plot: *gwg* Estimates: Restricted Dataset

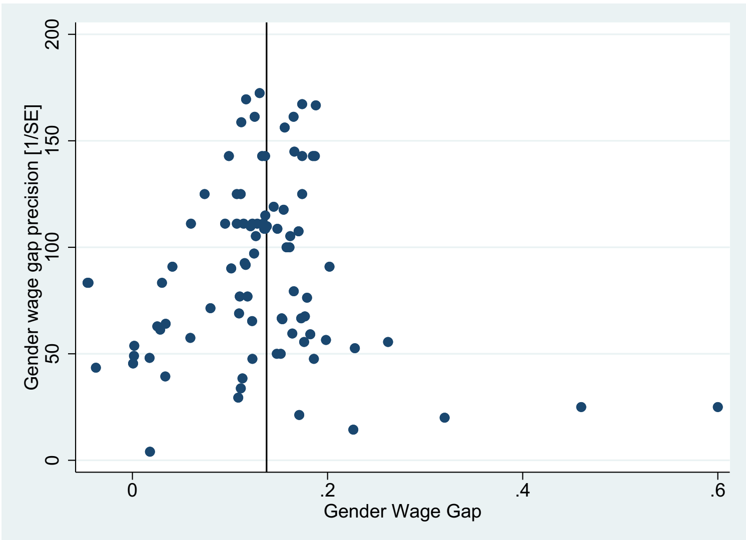


Table A.2: Publication Bias Tests: Restricted Dataset

Panel A: Linear techniques				
	OLS	IV	Study	Precision
Standard Error	-0.130	1.340	-1.140	2.269
	(1.846)	(5.313)	(3.177)	(5.228)
Constant	0.136***	0.122**	0.132***	0.132***
	(0.0236)	(0.0528)	(0.0432)	(0.000453)
Observations	135	135	135	135
Panel B: Between and Within				
	BE	FE	RE	
Standard Error	0.753	-0.346	-0.168	
	(3.027)	(0.828)	(1.037)	
Constant	0.130**	0.139***	0.146***	
	(0.0548)	(0.00817)	(0.0362)	
Observations	135	135	135	
Panel C: Nonlinear techniques				
	Stem	WAAP	Kink	
Effect beyond bias	0.026	0.131***	0.132***	
	(0.104)	(0.014)	(0.014)	
Observations	135	135	135	