CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Investigating the Relationship between Industry Gender Imbalance and Gender Pay Gap: The Case of UK

Bachelor's thesis

Author: Štěpánka Máchová
Study program: Economics and Finance
Supervisor: doc. PhDr. Julie Chytilová, Ph.D.
Year of defense: 2024

Declaration of Authorship

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

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

During the preparation of this thesis, the author used OpenAI's ChatGPT to assist in developing the R-code, creating tables in LaTeX, and refining the writing style. After using this tool, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication.

Prague, April 29, 2024

Stepanka Machova

Abstract

Since the early 21st century, the issue of gender pay inequality has increasingly captured public and academic attention. This topic covers a broad spectrum of aspects, among which the industry-specific influence on gender pay disparities is of particular interest to our study. This thesis specifically examines the extent to which the Gender Pay Gap (GPG) in the United Kingdom is influenced by industry affiliation. Additionally, it evaluates the average proportion of women employed within these sectors, utilizing this information as an indicator of industrial segregation. Our analysis relies on data from 4908 companies operating in the United Kingdom and reporting consistently throughout the 2017/18 - 2022/23 period. We employ the Ordinary Least Squares estimation technique, incorporating corrections for heteroskedasticity detected in our models, to explore these relationships. The first important finding confirms a link between the GPG and industry affiliation, revealing significant differences across industries. Furthermore, our research demonstrates that industries with a higher average proportion of women tend to have larger GPGs. Although this finding is highly statistically significant, its practical significance is weakened by the modest magnitude of the estimated effects.

Keywords Gender Pay Gap, Wage discrimination, Inequality, Segregation, United Kingdom

Title Investigating the Relationship between Industry Gender Imbalance and Gender Pay Gap: The Case of UK

Abstrakt

Od počátku 21. století se problematika genderových rozdílů v odměňování stále více dostává do popředí veřejného i akademického zájmu. Toto téma zahrnuje širokou škálu aspektů, z nichž zvláštní pozornost v naší studii věnujeme vlivu průmyslových odvětví na rozdíly v odměňování mezi pohlavími. Tato práce konkrétně zkoumá, do jaké míry je Gender Pay Gap (GPG) ve Spojeném království ovlivněn průmyslovou příslušností. Kromě toho posuzuje průměrný podíl žen zaměstnaných v těchto odvětvích, přičemž využívá tuto informaci jako ukazatel průmyslové segregace. Naše analýza vychází z údajů 4908 společností působících ve Spojeném království, které konzistentně vykazovaly své statistiky v období od roku 2017/18 do roku 2022/23. K prozkoumání těchto vztahů využíváme metodu nejmenších čtverců, do které zahrnujeme korekce pro identifikovanou heteroskedasticitu. První důležité zjištění potvrzuje spojitost mezi GPG a průmyslovou příslušností, odkrývající výrazné rozdíly napříč průmyslovými sektory. V druhé řadě náš výzkum ukazuje, že odvětví s vyšším průměrným podílem žen mají tendenci mít také větší GPG. Ačkoliv je toto zjištění vysoce statisticky významné, jeho praktický význam je oslaben skromnou velikostí pozorovaných efektů.

Klíčová slova Gender Pay Gap, Mzdová diskriminace, Nerovnost, Segregace, Spojené království

Název práce Analýza vztahu mezi genderovou nerovnováhou v odvětvích a rozdílem ve mzdových podmínkách: Případ Spojeného království

Acknowledgments

I would like to express my gratitude especially to doc. PhDr. Julie Chytilová, Ph.D., my thesis supervisor, for her continuous support from the beginning of my thesis topic selection to its completion. Her professional guidance, valuable insights, and the time and patience she dedicated to my thesis have been essential in shaping my work. I genuinely appreciate her willingness to supervise my thesis and her irreplaceable contribution throughout this process.

Special thanks are also extended to my family for their endless support and belief in me, not only during my academic journey but also in every aspect of my life.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Máchová, Štěpánka: *Investigating the Relationship between Industry Gender Imbalance and Gender Pay Gap: The Case of UK*. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2024, pages 78. Advisor: doc. PhDr. Julie Chytilová, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

aGSIF Adjusted Generalized Standard Error Inflation Factor

EEC European Economic Community

EHRC Equality and Human Rights Commission

EOC Equal Opportunities Commission

EU European Union

GEO Government Equalities Office

GPG Gender Pay Gap

IP Karmel-MacLachlan

MLR Multiple Linear Regression

NMW National Minimum Wage

OLS Ordinary Least Squares

ONS Office for National Statistics

SIC Standard Industrial Classification

UK United Kingdom

VIF Variance Inflation Factor

Chapter 1

Introduction

Gender inequality in remuneration remains a significant challenge in labour economics, with women often facing different pay conditions compared to men. One aspect contributing to this persistent issue is the existence of societal biases resulting in a tendency to undervalue work performed by women. According to the devaluation theory, attributes associated with women are typically assigned lower value, applying to work-related matters as well. Subsequently, this leads to the reality that jobs predominantly occupied by women tend to be generally less compensated than those typically performed by men (Leuze & Strauß, 2016). Furthermore, it is essential to note that men and women often receive different treatment even within the same sector or identical job roles. This issue is quantified through the GPG, assessing the wage disparity between men and women, typically expressed as a percentage of men's earnings (WGEA, 2024a). In response, the European Union (EU) has encouraged its member states, including the United Kingdom (UK) at that time, to adopt legislation to monitor gender pay equality. Under the Equality Act 2010 Regulations 2017, larger companies and public sector bodies in the UK are required to annually disclose their GPG-related statistics, providing us with a rich foundation for our study.

The general interest in workplace equality is growing, and a notable amount of scholars have devoted their research efforts to related topics. While much of the existing research focused on occupational segregation, we shifted our attention to industrial segregation, explicitly examining the link between industry gender imbalance and the GPG. The study is carried out on a sample of 4908 UK companies over the period from 2017/18 to 2022/23, applying the Ordinary Least Squares (OLS) estimation as our chosen method. Our methodology initially utilizes industry dummies to identify 16 distinct industry categories based on the SIC classification. Secondly, it considers the average female representation within industries, chosen as an indicator of industrial segregation. This second approach builds on the strategy used by Olsen *et al.* (2018), who applied it in the occupational context. The motivation behind this work is to determine, firstly, whether the GPG is strongly dependent on the industry affiliation, and secondly, whether there is an existent relationship between the GPG and the average proportion of women within these industries. The objective of this thesis is to broaden the existing knowledge by analyzing new data made available through the adoption of EU-based legislative framework and, further, by altering the dependent variable to move away from the traditional focus on wages.

The thesis is structured as follows. Chapter 2 and Chapter 3 provide background information. The discussion begins with the development of equality law in Chapter 2, specifically focusing on legislative pieces relevant to this thesis. This is followed by Chapter 3, which outlines the key aspects associated with the GPG, emphasizing their current state within the UK. Chapter 4 reviews existing literature, introducing the GPG as a traditional measure for assessing the degree of inequality in financial remuneration and concludes with the formulation of our research hypotheses. In Chapter 5, the focus shifts to the data sources used, detailing all the steps done to prepare our final dataset. This chapter also introduces our variables together with tables showing their distribution to provide the audience with a basic understanding of them. Chapter 6 presents the selected methodology, states the reasons for its selection and verifies underlying assumptions to ensure valid statistical inference. At the beginning of Chapter 7, graphical visualizations are employed to illustrate our data before proceeding with the estimation process and the subsequent presentation of results. Finally, Chapter 8 concludes the whole thesis with a summary of the findings and implications for potential future research.

Chapter 2

Gender Equality Framework

The following two chapters are designed to provide the reader with background information regarding the issue of gender equality in remuneration and the specific setting of the UK. Firstly, we delve into the development of equality law since the post-war period, specifically focusing on the legislation pieces giving a framework to this thesis.

2.1 Development of UK's Equality Laws and Policies

Early Roots of Equality Legislation in the UK

Equal pay for women is now recognized as one of the most significant labour market reforms of the twentieth century, and **the Equal Pay Act of 1970** was an important step in this evolutionary process. The Act mandated that employers must offer equal pay, terms, and conditions for work of the same or substantially similar nature. Moreover, it required parity in compensation for roles that, although differing, were assessed to hold equal value under a job evaluation scheme (Scott, 2023). Both Zabalza & Tzannatos (1985) and Joshi & Owen (1987) report a direct and positive impact on women's relative pay during the 70s, with reductions in the GPG of 19% and 15%, respectively. While the primary aim of this Act was to address wage-related matters, **the Sex Discrimination Act**, coming into force on December 29th 1975, alongside the Equal Pay Act, was designed to eliminate discriminatory practices in non-pay aspects, such as hiring, promotion opportunities or job transfers (University of Kent, 2020; Scott, 2023). Applicable to both men and women, this Act prohibited discrimination 'on the grounds of' sex and married status, including both direct and indirect forms of discrimination (Dickens, 1995). A major impact of this new legislation was the establishment of the Equal Opportunities Commission (EOC), tasked with overseeing compliance with anti-discrimination measures (Scott, 2023). Although now operating as part of the Equality and Human Rights Commission (EHRC), this organization persists in its role to this day. Additionally, the Act provided individuals with direct access to courts and tribunals, allowing them to challenge cases of sex discrimination directly. Despite theoretical accessibility, proving direct discrimination remained challenging, as employers rarely admitted to such practices, and discriminatory actions were often concealed (Atkins, 1986).

European Union Membership and Legislative Adjustments

In January 1973, the UK became a member of the European Economic Community (EEC), which later gave rise to the contemporary EU, committing to align with its principles (Böök *et al.*, 2021; Guerrina & Masselot, 2018). The principle of equal pay for equal work between men and women was, in fact, one of the fundamental principles of the 1957 Treaty establishing the EEC (Böök *et al.*, 2021). Despite the implementation of the Equal Pay Act, many employers found loopholes to get around its requirements (Scott, 2023). Instances of abuses of the Equal Pay Act, along with the UK's failure to meet its obligations under the Treaty of Rome regarding equal pay for work of equal value, as determined by the European Court of Justice in 1982, led to the introduction of **the Equal Pay Amendment Regulations of 1983**. These Regulations helped to translate the EU's Work of Equal Value Directive into practical measures (Bryson *et al.*, 2020), granting women equal compensation for work of equal value. To prevent misuse, this legislation replaced the previous definition of equal pay with a more robust one, now described as having similar demands in terms of effort, skill, and decision-making (Scott, 2023).

Supplementary Policies for Addressing Wage Inequality

Further initiatives to reduce wage inequality led to the implementation of a statutory **National Minimum Wage (NMW)** in April 1999. The expectation of reducing inequality by compressing the entire wage distribution stemmed from the USA evidence, where this step indeed served as a tool for reducing wage disparities (Dex *et al.*, 2000; Dickens & Manning, 2004). The similarity of trends between wage inequality in the USA and the UK has further contributed to optimism regarding its implementation. Yet, the outcomes in the UK fell short of expectations. Although it has indeed successfully boosted the earnings of the lowest-paid workers, the overall effect on wage inequality was minimal. One of the reasons identified was that the minimum wage had been set at a relatively low level, affecting less than 10% of workers (Dickens & Manning, 2004). Therefore, the GPG was reduced only at the lower end of the wage distribution, resulting in a shortfall of the intended effect of the legislation (Bryson *et al.*, 2020).

Additional policies aimed at increasing women's participation in the labour market have been implemented between and after the introduction of the aforementioned ones. Notable examples include legislation on statutory paid maternity leave in 1975 *(the Employment Protection Act)*, the initiation of free nursery education in 1999 *(the National Childcare Strategy)*, and the establishment of regulations concerning part-time workers in 2000 *(Part-time Workers' Regulations)*. Although these policies did not directly address equal pay, they contributed to a rise in the number of mothers with young children participating in the workforce (Bryson *et al.*, 2020).

Modernizing Equality Legislation: The Equality Act 2010

As a former EU member until January 1st 2021, the UK operated under both European and national anti-discrimination regulations in the labour market. **The Equality Act 2010** simplified the law by unifying all existing legislation in this field into one comprehensive Act, covering England, Wales, and Scotland; however, not Northern Ireland. In labour relations, Northern Ireland maintains autonomy to formulate its laws (Mosakova & Kizilova, 2021).

The Equality Act 2010 prohibits discrimination based on nine protected characteristics:

- Age
- Disability
- Gender reassignment
- Marriage and civil partnership
- Pregnancy and maternity
- Race
- Religion and belief
- Sex
- Sexual orientation

As illustrated by these categories, the Act protects citizens against discrimination across a spectrum of attributes, including appearance, beliefs, and behaviour. Notably, this protection extends to situations where discrimination stems from association with individuals possessing protected characteristics. As a result, traditionally disadvantaged individuals are expected to have better inclusion in society and encounter less discrimination, or at least have legal resources when faced with such challenges (Fell & Dyban, 2017).

Further Advancements: The Equality Act 2010 Regulations 2017

There are various regulations associated with the Equality Act 2010, which are evolving continuously. Of particular importance for this thesis are **the Equality Act 2010 (Gender Pay Gap Information) Regulations 2017** and **the Equality Act 2010 (Specific Duties and Public Authorities) Regulations 2017**. The 2017 Regulations established obligations for private and public sector employers with 250 or more employees to annually disclose their GPG status on a specific date, known as the snapshot date. Notably, the law also considers situations where the employee count fluctuates around 250. If, on the snapshot day, the company has fewer than 250 employees, but the director anticipates an increase during the year, it is recommended to voluntarily submit a report and prepare for the following year when compliance becomes mandatory. The snapshot date is March 31st for most public employers and April 5th for private, voluntary, and all other public employers. The decision on whether the company is subject to reporting obligations is made based on this date, and within a year from that point, the report must be available in the required form (Maříková & Volejníčková, 2022).

Entities subject to this legislation have to calculate six metrics related to gender equality in the workplace:

- Mean and median GPG for hourly pay
- Mean and median GPG for bonus pay
- Percentage of men and women receiving bonus pay
- Percentage of men and women in each hourly pay quarter

A summary of these metrics must be uploaded to the website designated by the Government Equalities Office (GEO). Following this, all GPG reports become publicly available on the GPG Service page. Furthermore, private and public companies have to publish these reports on their respective websites or their intranet or parent company's website if they do not have their own, where they should be maintained for a minimum of three years. These reports are intended to be publicly available, primarily to benefit the employees of the given company or organization, ensuring transparency in communicating this information. Non-compliance is considered unlawful and is subject to enforcement by the EHRC, which oversees deadlines and accuracy and holds the authority to impose sanctions (Maříková & Volejníčková, 2022).

2.2 Brexit: Impact on Gender Equality Legislation

In January 2020, the UK formally ended its membership within the EU, beginning a transition period until December 2020. Following the end of the transition period, the UK was no longer bound by EU law. To avoid potential challenges of Brexit, **the European Union (Withdrawal) Act 2018** was introduced. Before the Exit Day, this Act eased the transition by replacing EU law with its domestic equivalent, known as the retained EU law (Brett Taylor & Wilson, 2023). **The Equality Act 2010 (Amendment) Regulations 2023** later addressed the preservation of the equality framework derived from the EU law, including the reporting obligation for relevant employers. These legislative adjustments have ensured that complete information is available for companies required to report since the initial reporting year 2017/18 (Department for Business and Trade, 2023).

Throughout the four decades of EU membership, the EU's equality framework has had a positive impact on the development of relevant legislation in the UK. Fagan & Rubery (2018) observed that the UK synchronized the introduction of its laws on equal pay and sex discrimination with EU legislation in response to the offer of EU membership. The EU's initiatives and directives, mostly preceding advancements in the UK, thus acted as a primary catalyst for adopting these measures. Sanders & Flavell (2023) stress the historical struggles of gender equality advocates in the UK in their fight for policy change at the national level. Nevertheless, for many of them, the EU emerged as a promising platform for initiating such changes, as subsequently demonstrated.

Chapter 3

Gender Dynamics in the UK Labour Market: Key Indicators

This chapter focuses on four aspects commonly cited together with gender pay inequality. The specific objective of this chapter is to explore these characteristics within the UK context. We aim to provide a brief overview of their historical development and current status, helping us to determine whether these aspects continue to pose challenges.

3.1 Activity and Employment Rates

The role of women in the UK labour market has changed significantly as time has progressed. Prior to the Second World War, merely around one-third of women were economically active, but by the 1980s, this number had nearly doubled. Furthermore, this rise in women's overall activity rates has been accompanied by a corresponding decrease in activity among low-earning men, narrowing the gap in employment between genders (Bryson *et al.*, 2020). This trend is evident in the increasing proportion of women of prime working age in employment, which rose from 57% in 1975 to 80.6% in 2023 (Roantree & Vira, 2018; FRED, 2023). Changing family patterns, with women now less inclined to exit the labour market after the first childbirth, unlike previous generations, are cited as partly responsible for this shift (Roantree & Vira, 2018).

Unequal Distribution of Unpaid Work

Even with the increasing efforts of employers to provide more family-friendly conditions to facilitate a better balance between work and family life, the unequal distribution of unpaid work remains an ongoing challenge (Mosakova & Kizilova, 2021). The allocation of time to domestic tasks limits the available time for individuals to engage in paid work. This situation is recognized as the gender care gap (OECD, 2017). In the UK, women spend more than double the time doing unpaid work compared to men, except for transportation-related tasks where men are more involved. Despite being often overlooked, the value of this work would amount to an average of £260 per week if it were to be compensated (ONS, 2016).

While this disparity may be kind of expected and reasonable, it remains essential to note its impact, as it often translates into a greater overall time commitment to both paid and unpaid work, observed across most OECD nations, including the UK. Furthermore, there is evidence of a positive relationship between women's participation in the labour market and the extent to which their male partners share housework responsibilities. In countries with high female employment rates, women dedicate less time to unpaid work, shrinking the gap in unpaid work engagement to less than one hour per day in Scandinavian countries (OECD, 2017). In the UK setting, this gap remains below the threshold of two hours per day (OECD, 2024). This situation highlights the interplay between gender roles, unpaid work, and women's involvement in the workforce.

Perceptions and Attitudes towards Gender Roles

Prevailing stereotypes and societal expectations often constrain women to domestic roles. At the same time, men are generally positioned as the primary economic forces in the family. These stereotypes not only have the potential to be a barrier to women's active participation in the workforce but might also rob fathers of valuable time with their children (OECD, 2017). The Special Eurobarometer 465, a survey requested by the European Commission in 2017, explored how citizens feel about gender equality and tried to uncover their perceptions about gender roles. Among various questions, respondents were asked to provide their views on two key statements related to gender roles.

These statements were as follows:

- **Statement 1:** The most important role of a woman is to take care of her home and family.
- **Statement 2:** The most important role of a man is to earn money.

According to the results, Sweden and Denmark ranked the best among EU members, with the lowest agreement rates for both statements. When combining this information with the aforementioned statistics, a clear pattern emerges: *countries with the slightest acceptance of traditional gender roles also tend to demonstrate the smallest disparity in the distribution of unpaid work*. This association also remains valid for the UK; although it is still far from Scandinavian countries, the overall comparison with other European countries is quite favourable. In terms of the specific results of the Eurobarometer survey, the acceptance of these statements in the UK was 38% and 36% respectively, still below the EU average and among the top eight nations in this regard (European Commission, 2017).

3.2 Educational Attainment

Before the First World War, only a minority of women had qualifications, with merely one in twenty achieving tertiary education. Although the gap in educational attainment began to narrow for women born at the end of the Second World War, men were still more likely to attain tertiary education (Bryson *et al.*, 2020). However, this pattern has shifted over time. Since the 1970s, women have not only caught up with men but have even outpaced them in terms of both basic and tertiary qualifications. These continuously converging qualifications between men and women have led to the complete closure of the educational attainment gap in the UK by the 2000s, with women under the age of 55 now exhibiting higher education levels than their male counterparts (Scott, 2023). Notably, this trend is not limited to the UK; it resonates globally. Aliprantis *et al.* (2011) highlight that since the late 1970s in the US, men's progress in college attainment has stagnated, contrasting with the upward trajectory of women's educational accomplishments. Evidence from OECD members does not contradict this claim, indicating that women have exceeded men in college attainment in nearly all countries, with Japan and Turkey as the sole exceptions (Aliprantis *et al.*, 2011).

Over the last century, notable progress has been observed in narrowing the distinctions between men and women. Termed by Goldin (2014) as the **grand gender convergence**, these advancements extend beyond education and include labour force participation, hours of paid work, or domestic responsibilities. This term symbolizes the trend of women becoming more and more like men, particularly in work-related issues. After all, there does appear to be a shift in the dynamics of certain matters.

3.3 Changing Fertility Patterns

Fertility trends in the modern era, both globally and in the UK, are characterized by a decline. Unlike historical periods in which high mortality rates regulated population growth, reduced fertility rates now dictate the pace of demographic changes (Roser, 2017).

Following the Second World War, fertility rates increased significantly, reaching their peak at 2.94 in 1964. From there, they began to decline, stabilizing below 2 in 1980 and remaining at that level since. This shift was largely influenced by the advent of modern contraception in the early 1960s and the legalization of abortion in 1968, allowing women to postpone motherhood and have greater control over family planning (Scott, 2023). The change in reproductive decision-making was closely related to improved education, better employment opportunities, and increased women empowerment. Younger adults, especially those with higher qualifications, exhibited a growing tendency to postpone or entirely avoid parenthood, indicating a general change in attitudes and lifestyle choices (Bryson *et al.*, 2020). Despite this demographic transition, the UK and France have maintained one of the highest total fertility rates throughout the last quarter of the twentieth century (Kiernan, 1998). But in the current century, the situation is once again not favourable. In 2019, fertility rates in the UK had already fallen to some of the lowest levels ever recorded (Berrington *et al.*, 2021). As of 2023, the total fertility rate stands at 1.57, slightly above the European average, according to Statista (2023).

3.4 Segregation and Career Choices

The challenge of gender-based segregation in the labour market is apparent in both horizontal and vertical dimensions. Horizontal segregation refers to the uneven distribution of women and men across various industries and occupations (Carranza *et al.*, 2019). This suggests that certain jobs tend to be more commonly associated with either women or men. For instance, women often find employment in roles commonly referred to as **the five Cs** - cleaning, caring, clerical, cashiering, and catering. A common problem is that a significant proportion of these roles are classified as low-skilled, leading to a disproportionate representation of women in lower-paid jobs. This pattern persists despite women's academic success and generally higher educational attainment (Close Your Pay Gap, 2020). On the other hand, vertical segregation is associated with differences in career choices and progression, resulting in women predominantly occupying junior roles within organizations (Carranza *et al.*, 2019).

As acknowledged by Miller *et al.* (2004), occupational segregation stands out as one of the strongest influences on the career choices of young individuals. From a very early age, individuals tend to link certain jobs with either males or females, shaping their preferences accordingly. If a job is predominantly male-dominated, young children instinctively believe that a man should perform it. With age, they tend to adopt more liberal perspectives, especially girls (Miller *et al.*, 2004). Research conducted by the EOC on the attitudes of young people towards jobs in non-traditional sectors supports this observation. It reveals that girls are more likely to believe that traditionally male-dominated jobs could be performed equally well by both genders, with 80% of them expressing a willingness to consider pursuing a non-traditional career. Despite this greater openness, many adolescents follow in their parents' footsteps when making job choices, still closely mirroring the existing segregation patterns (Miller *et al.*, 2004; The House of Commons, Trade and Industry Committee, 2005).

Similarly, ingrained social norms associated with gender often influence our decision-making. This influence may represent a serious challenge, particularly for girls and women. The British Social Attitudes Survey revealed that although a minority of respondents expressed their beliefs about gender differences in math and computing abilities at school, a greater portion of them assumed the superiority of boys over girls (Huchet-Bodet *et al.*, 2019). Furthermore, as highlighted by the Longitudinal Survey of Young People in England, 60% of boys identified STEM subjects as their best, while only 33% of girls did so, despite actual results favouring girls (HM Government, 2019). This empirical evidence illustrates how perceptions and widely held beliefs can affect the mindset of individuals, even when these notions may not necessarily be accurate, as shown in this case. Martin (2011) argues that children acquire distinct gender-based expectations and predefined roles early in their lives. She highlights that what society deems appropriate and expected does not necessarily dictate behaviour. However, children may fear going against these norms, as evidenced by the findings of The Girls Attitudes Survey 2017. This survey revealed that half of the girls interviewed, aged 7 to 10, acknowledged the impact of stereotypes on various aspects of their lives, including appearance, choice of leisure activities, behaviour, and school participation. While the impact was weaker in the age category 11-21, a significant percentage, not lower than 43%, continued to perceive this influence (Girlguiding, 2017).

To demonstrate the degree of gender-based segregation present in the UK, we turn to empirical evidence. Charles (2003) pointed out that the UK and France exhibited the most evident vertical segregation among the countries studied. To quantify the level of horizontal segregation, Verashchagina *et al.* (2009) utilized the Karmel-MacLachlan (IP) index and identified a value of 25.3% in 2007, which means that to achieve gender balance in employment, 25.3% of all employed individuals would have to switch to different occupations. Despite still maintaining a relatively high level, the UK has experienced a considerable decrease since 1997, categorizing it as one of the rapidly de-segregating countries, as indicated by the change in the IP index (Verashchagina *et al.*, 2009). The examination of 2014 Eurostat data by Mavrikiou & Angelovska (2020) further supports the notion that the UK is one of the better-performing European countries in addressing horizontal segregation, consistent with the trends outlined earlier.

Although the UK performs relatively well compared to other nations, the data still reveal a notable presence of segregation, highlighting its influence on gender dynamics in the working environment, both within the country and globally.

Chapter 4

Literature review and hypotheses development

This chapter aims to provide an overview of the existing knowledge regarding gender discrimination in the workplace, especially focusing on the aspect of occupational and industrial segregation. Additionally, the GPG is presented as a metric for evaluating gender disparities in financial remuneration, along with its possible methods of calculation.

As emphasized in Section 3.4, gender segregation in the workplace is a significant and persistent phenomenon, showing some gradual improvements. This progress, however, is uneven. As observed by Yavorsky & Dill (2020), the ongoing process of desegregation in the workplace is mainly driven by women seeking to enter traditionally male-dominated occupations rather than vice versa. Moreover, the limited progress for men in workplace desegregation can be attributed to societal challenges, including concerns about perceived loss of masculinity or less favourable pay conditions. Unlike the natural progression observed among women, men's advancements may not occur organically and may require economic shocks, such as involuntary unemployment, forcing them to consider alternatives. Therefore, achieving a more balanced distribution through men's contribution is less likely to occur spontaneously without external pressures (Yavorsky & Dill, 2020).

Regarding pay, the so-called devaluation theory suggests that women in Western industrialized countries, such as the UK, are typically assigned lower value than men, involving all aspects associated with them, including their work. Consequently, occupations predominantly filled by women are linked to lower overall wages (Leuze & Strauß, 2016). While the past attributed this wage disparity to lower education levels or limited experience in the job market, these arguments are now losing relevance, especially considering the progress made by women, both in terms of education and active participation in the labour market (Miller, 2016). Connolly *et al.* (2016) further argue that the traditional concept of the sole male breadwinner in the UK is decreasing in impact. They point to a shift in family-working patterns towards a dual model, where both parents are actively engaged in full-time employment and share financial responsibilities. Although some women may still instead opt for more flexible roles, potentially leading to lower pay, this decision is often driven by the desire to prioritize family commitments. When women's lower earnings are not solely a matter of personal choice but are influenced by market conditions or discriminatory practices, we encounter issues of unequal treatment and valuation.

The GPG serves as a standard metric to assess the disparity between women's and men's wages, typically expressed as a percentage of men's earnings (WGEA, 2024a). As Anderson *et al.* (2001) emphasize, it is crucial to recognize the variability in calculating the GPG, as different methodologies can yield varying magnitudes of the gap and potentially reveal different trends. This measure can be reported based on either mean or median. The key distinction lies in how each statistic treats the distribution of wages. While the mean considers the entire distribution of wages, the median focuses solely on the wage of the middle earner, providing a basis for comparison. Given that men often dominate higher-paying roles, their wage distribution tends to be more skewed to the right, affecting the mean- but not median-measure. Consequently, the GPG calculated using the mean tends to be larger, as it accounts for all individuals in the distribution without distinction, while the median remains unaffected by extreme values. When prioritizing the average individual with characteristics that apply to the majority of workers, the median-based GPG emerges as a better measure. However, its reduced sensitivity can also be considered a limitation in certain situations, as it does not consider all observations and thus overlooks any progress occurring below or above the median earner's wage over time. For instance, if advancements are made in the lower-earner distribution, resulting in higher pay for women but still below the middle-earner level, such progress will remain unnoticed by this measure.

The magnitude of the gap is also influenced by the base from which the wage difference is computed. When calculated using weekly earnings, it becomes notably impacted by the number of hours worked, which generally tends to be lower for women. This aspect might be, moreover, cited as evidence to dismiss the issue of gender-based pay discrimination despite not necessarily being the driver. Hence, the hourly pay methodology is more commonly used, as it erases the effect of differences in weekly hours worked (Anderson *et al.*, 2001). This approach is also employed in the dataset under our examination, utilizing both mean- and median-based measures of the GPG.

Returning to the issue of prevalent undervaluation of women's work, Perales (2013) indeed finds a strong negative relationship between the feminization of occupations and wages in the UK context, thus supporting the idea of the devaluation theory. The fact that men are then discouraged from pursuing careers in these fields is not surprising, especially when male-dominated occupations tend to offer higher salaries compared to female-dominated ones, even when similar skill levels are required for the roles, as highlighted by Hegewisch & Mefferd (2021). In fact, men tend to earn more than women, even within female-dominated fields. This disparity in pay persists across various occupational categories, as highlighted by the 2017 American Community Survey, where no single field, whether dominated by men or women, showed women surpassing men in earnings. If these disparities were solely the result of divergent occupational preferences between genders, we would expect women to be able to achieve comparable pay when occupying the same roles as men. However, this is far from the reality we are encountering (Elsesser, 2019).

Considering the full scope of evidence, gender wage disparities can be observed not only between occupations but also within them, even when accounting for potential influencing factors. Despite earlier findings from Norway indicating minor wage disparities between men and women within the same job roles (Petersen *et al.*, 1997), research conducted by Penner *et al.* (2023) acknowledges that the within-job gender gaps are substantial across all 15 countries studied, contributing significantly to the overall gender pay differences. Although perceptions of its significance vary, both studies recognize the role of gender in shaping remuneration within occupations.

This emphasizes the reality that individuals are being paid unequally based on their gender, with several scholars acknowledging that these variations are often closely tied to gender composition within particular occupations. A study led by Blau & Kahn (2017), analyzing evidence from the United States, revealed that by 2010, while the significance of human capital variables in explaining the GPG diminished, the role of gender imbalance across occupations remained important, accounting for 32.9% of the GPG. Similarly, in the European context, Boll & Lagemann (2018) analyzed SES Eurostat data and recognized the connection between occupational segregation and the GPG. However, they noted variations in the strength of this relationship across countries, with an increasing effect observed in the UK and a decreasing effect in Italy. This suggests that while occupational segregation widens the gap in the UK, it reduces it in Italy. Research done by Bedaso (2024) also yielded related results, confirming the importance of female occupational segregation in explaining a portion of the GPG, particularly among workers at the lower and median wage levels. The impact of gender-based segregation has also been examined in the UK. According to Olsen *et al.* (2018), occupational segregation stood out as one of the top four most important factors influencing the GPG in the country during 2014/15. This link has been recognized in various global contexts, including Europe, the USA, and even the African continent, yet findings from Armenia present contradictory evidence (UN Women, 2020).

The issue of segregation extends beyond occupations to industries as well. The underlying distinction lies in the fact that while there is a defined set of industry categories, individual occupations are not restricted to a single industry and can span multiple sectors (Jones & Saulcy, 2009). Thus, while segregation within occupations and industries shares a common principle, they are slightly distinct concepts. As women tend to cluster in certain occupations, they also tend to cluster in specific industries. However, this segregation pattern varies across countries; an industry dominated by women in one country may not exhibit the same gender composition in another. Among European nations, women are often notably overrepresented in fields such as Education or Health and Social Work Activities (Boll & Lagemann, 2018). Conversely, in many rural countries in Latin America, South Asia, and North and Sub-Saharan Africa, women have historically constituted the primary workforce in agriculture. Moreover, their share in this sector has continued to grow in most of these regions from 1980 to 2010 (Bonny *et al.*, 2022). As demonstrated, the prevalence of industrial segregation varies across countries, and scholars have also presented differing opinions. Boll & Lagemann (2018) confirm the link between gender composition within industries and the resulting GPG, asserting that a notable portion of the gap stems from industrial segregation. This holds for all examined EU countries except for Malta and Luxembourg, where the relationship was either non-existent or negligible. Furthermore, the investigation by Olsen *et al.* (2018) in the UK revealed the importance of wage disparities stemming from industry sector affiliation in perpetuating the gap. The primary contributor identified was the disproportionate representation of men in two particular sectors - Manufacturing and Construction - collectively accounting for 29% of the gap.

4.1 Contribution of This Thesis

The study by Olsen *et al.* (2018), emphasizing earlier findings from the UK context, was conducted before the legislation foundational to this thesis was implemented. It also relied on older data specifically focusing on UK citizens. In contrast, our research plans to utilize newly available data tracking the performance of UK companies. This approach is also reflected in our choice of the dependent variable. While Olsen *et al.* (2018) and previous studies primarily used wages as the dependent variable, we plan to contribute to the field of GPG analysis by directly employing it as the dependent variable, enabled by the recent legislative changes. To quantify occupational segregation, Olsen *et al.* (2018) incorporated an additional variable indicating the percentage of male workers within each occupational category. However, in examining the industry-specific effect, Olsen *et al.* (2018) employed a set of dummy variables distinguishing between up to 8 industry sectors without directly quantifying industrial segregation. To address this gap, we propose utilizing the average percentage of female workers within each industry sector as a metric for industrial segregation. Additionally, we intend to contrast this approach with the use of an expanded set of dummy variables distinguishing between 16 industry categories based on SIC. Our objective is twofold: to investigate the dependence of the GPG on industry affiliation, with expected variations among the identified categories, and to determine whether there is evidence of an association between the average gender structure of industries and the resulting GPG.

Our two main hypotheses to answer are the following:

- Firstly, we believe that the GPG is dependent on industry affiliation, exhibiting variations among the 16 industry categories identified in our study.
- Secondly, we anticipate that there exists a significant relationship between the average gender structure of industries and the resulting GPG of companies operating in these industries, thus indicating the presence of industrial segregation.

To the best of our knowledge, the data framework we plan to employ has not been used yet for similar research purposes and objectives. Thus, we believe that our research has the potential to fill a gap in the literature and provide a new perspective to the current understanding of gender-related matters.

Chapter 5

Data and Variables

This section delves into the data and variables used in our analysis. Firstly, we explain the origins of the data and the transformations undertaken to prepare them, highlighting all the adjustments made beyond the original state of the dataset. Secondly, we take a closer look at our variables and their respective distributions.

5.1 Data Sources and Construction

Thanks to adopting the Equality Act 2010 Regulations 2017 and its subsequent adjustments to address Brexit, a considerable portion of companies in the UK have been obliged to annually disclose their GPG information since the year 2017/18. This information is publicly available on the GPG Service at GOV.UK, where we obtained data for all available firms reporting consistently from 2017/18 to 2022/23, constituting our primary dataset of use. This dataset contains several variables. The first category involves variables related to firm identification and characteristics. In contrast, the second category of variables is mandated by the legislation and focuses on structural aspects and GPG-related statistics. These required measures are self-reported by companies via the GPG Service, with companies taking legal responsibility for the accuracy of the data provided (Government Equalities Office, 2019).

To investigate the relationship between industry gender imbalance and the GPG, we employed a second dataset, *EMP13: Employment by industry*, retrieved from the Office for National Statistics (ONS). This dataset categorizes employment by gender and industry type, with 16 industry categories identified through the SIC system. Certain modifications were necessary to be able to work with this dataset directly throughout our research. Given that employment is reported as the actual count of employees, we modified the dataset by converting the count of female employees in each category into a percentage of the total employment count for that category. This adjustment allowed us to calculate the average proportion of women in each industry category across multiple years, with a specific focus on the period from 2017/18 to 2022/23.

Adjustments had to be made to our primary dataset as well. Prior to merging all reporting years, a variable indicating the year was manually assigned to facilitate differentiation between periods. Subsequently, to align with our research interests, we proceeded to identify the industry affiliation for each company listed in the dataset. For this purpose, we created a new variable *INDUSTRY*, based on the first two digits of the 5-digit SIC code. In case multiple SIC codes were associated with one company, the first one was used to classify it. If there was no SIC code or a non-existent one, possibly due to a typo, it was manually assigned according to the Companies House Register. To match each company with the right industry category, the Companies House list of SIC codes was used, with slight modifications made to its categories (Companies House, 2024).

These modifications specifically involved merging categories B, D, and E from the Companies House classification into a single category: Mining, energy and water supply. Similarly, the categories R, S, and T were grouped to form a single Other Services category. We made these adjustments to ensure compatibility with the *EMP13* dataset in the upcoming data merging, as industry categories B, D, E, R, S and T are not treated separately in *EMP13*. The resulting industrial framework, therefore, consists of 16 distinct categories, as shown in Table 5.1. The entire classification table, mapping out the specific SIC codes corresponding to each industry category, can be found in Appendix A.

In addition to this variable, we created another *PublicSector* variable that identifies whether the company belongs to the public sector. The value of this variable was derived from SIC codes as well. If the SIC code starts with number one and is separated by a comma, it indicates affiliation in the public sector.

Table 5.1: Modified List of Industry Categories based on SIC

Section	**Description**
A	Agriculture, forestry & fishing
B,D,E	Mining, energy and water supply
C	Manufacturing
F	Construction
G	Wholesale, retail & repair of motor vehicles
H	Transport & storage
I	Accommodation & food services
J	Information & communication
K	Financial & insurance activities
L	Real estate activities
M	Professional scientific & technical activities
N	Administrative & support services
O	Public admin & defence; social security
P	Education
Q	Human health & social work activities
R,S,T	Other services

As one of the last steps, we created another variable *ShareWomen*, which was linked for each company and year from *EMP13* dataset depending on the relevant industry type. In this phase, it was necessary to consider the absence of seasonal adjustment in *EMP13* dataset, which reports employment on a quarterly basis, in contrast to the annual reporting of the GPG. Thus, a value corresponding to the employment reported in the Oct-Dec quarter was assigned to each relevant year and industry to avoid seasonal effects. For example, the employment data from Oct-Dec 2017 was linked to each relevant industry for the 2017/18 reporting period to ensure clarity.

Finally, we resolved shortages that may have arisen from errors or typos in data input. In some cases, the information provided was either absent or unrealistic. Throughout the 6-year reporting period, we encountered inconsistencies in categorizing entities as public- or private-sector organizations. Following the same pattern, this anomaly applied almost exclusively to Academy Trust entities. All such instances were handled individually and resolved based on the latest information sourced from the Companies House Register. After all adjustments, we have obtained a subsample of 4908 companies, which have consistently reported throughout the entire period under examination, from the initiation of reporting duty in 2017/18 to 2022/23, hence constituting our final dataset.

5.2 Variable Description

5.2.1 Dependent Variable

To investigate our research questions, we use the reported annual GPG as the dependent variable. This metric, indicative of gender inequality in financial remuneration, can be reported in several forms. Within our dataset, two forms are available.

Mean gender pay gap for hourly pay (*DiffMeanHourlyPercent*) is the first reported form, representing the difference between average hourly pay for men and women as a percentage of average men's pay. If we denote $\bar{P}_m$ as the mean hourly pay for men, $\bar{P}_w$ as the mean hourly pay for women, then the mean-based GPG can be expressed using the following formula:

$$GPG_{\text{mean}} = \frac{(\bar{P}_m - \bar{P}_w)}{\bar{P}_m} \times 100 \tag{5.1}$$

Median gender pay gap for hourly pay (*DiffMedianHourlyPercent*) is the second reported form, representing the difference between median hourly pay for men and women as a percentage of median men's pay. Similarly, if we denote $\tilde{P}_m$ as the median hourly pay for men, $\tilde{P}_w$ as the median hourly pay for women, then the median-based GPG can be expressed using the following formula:

$$GPG_{\text{median}} = \frac{(\tilde{P}_m - \tilde{P}_w)}{\tilde{P}_m} \times 100 \tag{5.2}$$

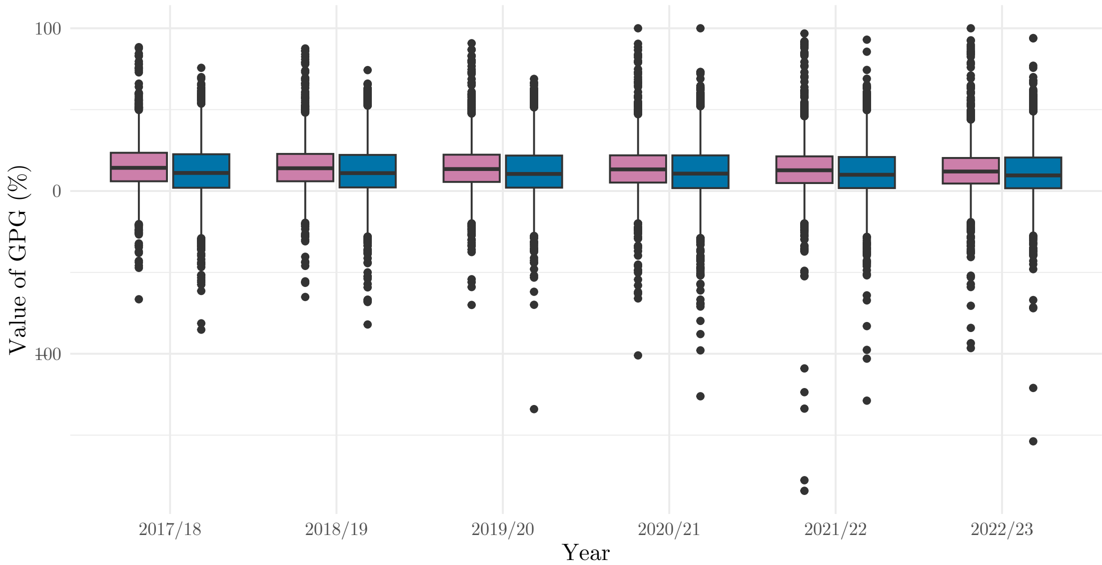
Both of these measures are calculated based on hourly wages; the difference lies in the type of wage used for comparison. It is also important to note that the obligation to include individuals' wages in this calculation applies only to full-pay relevant employees. Table 5.2 provides summary statistics of both GPGs computed using mean and median wages. Throughout the reporting period, we can observe a narrowing in the respective GPG in both cases. In addition, the average GPG based on median wages is consistently lower than that based on mean wages in the same period. This pattern corresponds with the standard practice, as the median-based measure is less affected by extreme values, typically resulting in lower rates.

Table 5.2: Summary Statistics for GPG by Year

Year	**DiffMeanHourlyPercent**				**DiffMedianHourlyPercent**			
	Mean	**St.Dev**	**Min**	**Max**	**Mean**	**St.Dev**	**Min**	**Max**
2017/18	15.0	13.4	-66.5	88.4	13.0	15.3	-85.2	75.7
2018/19	14.9	13.3	-65.1	87.6	12.9	15.2	-82.0	74.3
2019/20	14.4	13.1	-70.0	90.8	12.7	15.0	-134.0	68.9
2020/21	14.0	13.6	-101.0	100.0	12.7	15.8	-126.0	100.0
2021/22	13.3	14.3	-184.0	96.8	12.2	15.3	-129.0	93.0
2022/23	12.9	13.2	-96.5	100.0	11.9	14.8	-154.0	94.0

The variability of both measures is apparent, as evidenced by values at either end of the distribution: minimum or maximum. To illustrate this clearly, we create a boxplot for both GPG measures for each corresponding year. As depicted by Figure 5.1, many data points are identified as outliers, lying significantly above or below most of the data. This occurrence stems from the nature of the data. The GPG measure is designed in such a way that it can attain a wide range of values, thus resulting in a lot of statistics that significantly deviate from the norm. Even when these measures require attention, as some reach pretty great numbers, the most suspicious cases around the -100% and 100% thresholds were manually checked with the corresponding companies' statements, if available. As long as these two measures were consistent, we treated them as accurate. The legal responsibility still belongs to the respective company, and our ability to verify the data is limited to this extent. In our specific case, we choose to keep these measures in our analysis.

Figure 5.1: Boxplot of GPG across Years



Note: pink corresponds to mean-based GPG, blue to median-based GPG

5.2.2 Independent Variables

Independent variables cover individual company characteristics and gender composition across quartiles, reflecting wage distribution relative to gender. Additionally, they include the gender composition within industry categories to which the companies belong, aiming to highlight the potential impact of industrial segregation. Further elaboration on all independent variables is provided below.

Industry (*INDUSTRY*) is a categorical time-invariant variable, showing industry affiliation in one of the 16 industry categories (Table 5.1 presents the complete classification). To incorporate this categorical variable into the analysis, we transformed it into a set of 15 dummy variables. Construction was excluded to avoid the dummy variable trap and used as the reference category due to its status as the most male-dominated industry identified in our dataset. We further classified industries into three groups based on the prevailing gender of their workers, utilizing thresholds suggested by Torre & Jacobs (2021). An industry is classified as male-dominated if the proportion of female workers is below 33.3% and female-dominated if the proportion is 66.6% or higher. Industries falling between these thresholds are classified as gender-neutral. Gender dominance within each industry category and their overall representation in our dataset are shown in Table 5.3.

Table 5.3: Industry Distribution by Gender Dominance

Industry	**Dominance**	**n (%)**
Accommodation & food services	NEUTRAL	134 (2.7)
Administrative & support services	NEUTRAL	519 (10.6)
Agriculture, forestry & fishing	MALE	26 (0.5)
Construction	MALE	168 (3.4)
Education	FEMALE	725 (14.8)
Financial & insurance activities	NEUTRAL	245 (5)
Human health & social work activities	FEMALE	472 (9.6)
Information & communication	MALE	219 (4.5)
Manufacturing	MALE	693 (14.1)
Mining, energy and water supply	MALE	97 (2)
Other services	NEUTRAL	203 (4.1)
Professional, scientific & technical activities	NEUTRAL	327 (6.7)
Public admin & defence; social security	NEUTRAL	408 (8.3)
Real estate activities	NEUTRAL	62 (1.3)
Transport & storage	MALE	183 (3.7)
Wholesale, retail & repair of motor vehicles	MALE	427 (8.7)

As a result, only two sectors are female-dominated, namely Education and Human health & social work activities. Yet, almost a quarter of companies in our dataset are present in one of these, so they account for a large number of companies in total. The rest are either male-dominated or gender-neutral industries, with seven of each identified in our dataset.

Company size (*EmployerSize*) is a categorical variable as well, indicating into which of the total 7 size categories the company falls. We applied a similar procedure as before, converting this variable into a set of 6 dummy variables. For this variable, the smallest legally obliged group, consisting of 250-499 employees, was selected as the reference group and, therefore, omitted from the model. Unlike the previous one, this variable depends on time; therefore, its distribution is shown across the entire period. Yet, as observed in Table 5.4, it does not vary significantly over time.

Table 5.4: Employer Size Distribution by Year

Year	Employer Size Category						
	< 250	250-499	500-999	1000-4999	5000-19999	≥ 20,000	NP
2017/18	65	1867	1245	1282	292	47	110
	(1.32%)	(38.04%)	(25.37%)	(26.12%)	(5.95%)	(0.96%)	(2.24%)
2018/19	70	1783	1338	1341	329	42	5
	(1.43%)	(36.33%)	(27.26%)	(27.3%)	(6.70%)	(0.86%)	(0.10%)
2019/20	65	1705	1356	1378	326	48	30
	(1.32%)	(34.74%)	(27.63%)	(28.1%)	(6.64%)	(0.98%)	(0.61%)
2020/21	111	1700	1282	1335	317	44	119
	(2.26%)	(34.64%)	(26.12%)	(27.2%)	(6.46%)	(0.90%)	(2.42%)
2021/22	162	1684	1282	1315	327	42	96
	(3.30%)	(34.31%)	(26.12%)	(26.8%)	(6.66%)	(0.86%)	(1.96%)
2022/23	146	1652	1344	1372	351	42	1
	(2.97%)	(33.66%)	(27.38%)	(28%)	(7.15%)	(0.86%)	(0.02%)

Note: NP stands for Not Provided.

Public sector affiliation (*PublicSector*) is a binary variable equal to 1 if the company is present in the public sector and 0 otherwise. The effect of being part of the public sector is expected to be negatively correlated with the resulting GPG due to the anticipated higher compliance with equality principles and closer monitoring. Regarding the distribution, this feature remains constant over time, resulting in a stable count across each year of the reporting period. Among the 4908 companies, 797 declared their affiliation in the public sector, while 4111 reported their presence in the private sector. To provide a clearer illustration, these counts are converted into percentages, resulting in 16% and 84%, respectively. As this information is straightforward, we have omitted a corresponding table.

Female Representation in Lower Quartile (*FemaleLowerQuartile*) is a continuous numerical variable indicating the percentage of female employees among all employees in the lower quartile. The lower quartile explicitly refers to the lowest 25% earners within a company. The anticipated effect of this variable is positive, as a high proportion of females within this segment may contribute to higher resulting GPGs. However, the magnitude of this effect may vary depending on whether the base of the GPG is mean or median wage.

Female Representation in Top Quartile (*FemaleTopQuartile*) is a continuous numerical variable indicating the percentage of female employees among all employees in the top quartile, consisting of the highest 25% earners within a company. Thus, this measure sheds light on the gender distribution among the highest-paid employees. This variable is expected to be negatively correlated with the resulting GPG, as a higher percentage may indicate a better gender balance in leadership, mitigating wage disparities.

Female Representation in Industries (*ShareWomen*) is a continuous numerical variable that refers to the average gender composition in the industry category where the company operates. This variable is expressed in percentages, representing the percentage of women in each of the 16 industry categories defined by the SIC. The anticipated direction of the relationship is uncertain; our primary objective is to determine whether this relationship can be considered to exist at all, similarly to the case of occupations.

In terms of the distribution of these three variables, although they vary over time, their respective averages, minimum and maximum values, and standard deviations are pretty stable. Female representation in both the lowest and top quartile reaches a minimum of 0% and a maximum of 100% each year. The average representation in the top quartile stays consistently around 40%; in the lower quartile, it remains around 54%. Regarding the average female representation in industries, Construction shows the lowest share of female workers, with less than 15% throughout the entire period. The Human health & social work activities category is the direct opposite, with a female representation approaching 80%. Although this specific characteristic fluctuates over time, the changes are modest, typically within a range of $\pm$ 4% from the initial state. Detailed statistical information regarding the variable's evolution over time is available in Table B.1 in Appendix B.

Chapter 6

Methodology

Given our prepared data and variables, we intend to employ regression analysis to test our research questions. We first discuss the choice of models that are best suited to the specific nature of our data and make sure that all necessary assumptions for unbiasedness and valid statistical inference are addressed. Subsequently, we describe the theoretical modeling and execute our analysis using the R software.

6.1 Empirical Approach

We intend to investigate the dependence of GPG on industry affiliation and the possible link between industry gender imbalance and the resulting GPG. As outlined in Chapter 5, we have obtained a dataset tracking 4908 companies throughout the 6-year period ranging from 2017/18, from the beginning of the legislation, till 2022/23. Despite the panel data nature of our dataset, we have chosen to employ the OLS regression, treating each year separately. The justification for this step is the presence of time-invariant variables, which would be lost in traditional panel regression methods, such as Fixed Effects or First Differencing. Another panel regression method, Random Effects, could assist us in retaining the time-invariant characteristics within our analysis, as it does not assume that each independent variable changes over time for at least some i. However, the assumption of uncorrelatedness between unobserved effects and our chosen independent variables remains highly unreasonable (Wooldridge, 2013). Hence, we opt for the OLS method, allowing us to keep the time-invariant features in the analysis.

The main independent variables of interest are the industry dummies (*INDUSTRY*) and *ShareWomen*, representing the average proportion of women among all workers in each industry type each year. Due to the high multicollinearity observed when including these variables together in a single model, leading to substantial bias in the corresponding coefficients, we decided to use them separately. Firstly, we develop a model in which we replicate the usage of industry dummies, as Olsen *et al.* (2018) did, with some additional controls of our choice. As the legislation requires reporting two GPG measures and we dispose of both metrics, we alter the dependent variable for each model, resulting in double the number of final models. Thus, the first set of proposed models is as follows:

$$\begin{aligned} DiffMeanHourlyPercent_{i,t} = \beta_0 &+ \beta_1 INDUSTRY_{i,t} + \beta_2 PublicSector_{i,t} \\ &+ \beta_3 EmployerSize_{i,t} + \epsilon_{i,t} \end{aligned} \tag{6.1}$$

$$\begin{aligned} DiffMedianHourlyPercent_{i,t} = \beta_0 &+ \beta_1 INDUSTRY_{i,t} + \beta_2 PublicSector_{i,t} \\ &+ \beta_3 EmployerSize_{i,t} + \epsilon_{i,t} \end{aligned} \tag{6.2}$$

where *DiffMeanHourlyPercent* and *DiffMedianHourlyPercent* are the corresponding GPG measures of a selected company i in year t, with i=4908 and t=6. We include the index t to signal that we have a 6-year period, ranging from 2017/18 to 2022/23, during which each year is treated independently and corresponds to a distinct regression analysis. While recognizing the presence of both time-variant and time-invariant controls, the use of this index is solely intended to signify the separate execution of regressions for each year. *INDUSTRY* is the set of dummy variables indicating the industry category in which the selected company i operates. Further controls include the *PublicSector* affiliation and the company's size, indicated by *EmployerSize*. Epsilon (ϵ) represents the unobservable error.

To account for the gender structure within each company in the second set of models, considering potential individual variations within the industry categories, we further introduce two additional control variables - *FemaleLowerQuartile* and *FemaleTopQuartile*. Below, we present the second set of proposed models:

$$\begin{aligned} \mathit{DiffMeanHourlyPercent}_{i,t} = \beta_0 &+ \beta_1 \mathit{INDUSTRY}_{i,t} + \beta_2 \mathit{PublicSector}_{i,t} \\ &+ \beta_3 \text{EmployerSize}_{i,t} + \beta_4 \mathit{FemaleLowerQuartile}_{i,t} \\ &+ \beta_5 \mathit{FemaleTopQuartile}_{i,t} + \epsilon_{i,t} \end{aligned} \tag{6.3}$$

$$\begin{aligned} \mathit{DiffMedianHourlyPercent}_{i,t} = \beta_0 &+ \beta_1 \mathit{INDUSTRY}_{i,t} + \beta_2 \mathit{PublicSector}_{i,t} \\ &+ \beta_3 \mathit{EmployerSize}_{i,t} + \beta_4 \mathit{FemaleLowerQuartile}_{i,t} \\ &+ \beta_5 \mathit{FemaleTopQuartile}_{i,t} + \epsilon_{i,t} \end{aligned} \tag{6.4}$$

To connect our analysis with the average industrial gender composition, we substitute the industry dummies with the *ShareWomen* variable, which essentially reflects the industry type to some extent due to its design. However, it should precisely capture the average percentage of women out of the total worker base across industries. All other controls remain the same. The final third set of proposed models is then as follows:

$$\begin{aligned} \mathit{DiffMeanHourlyPercent}_{i,t} = \beta_0 &+ \beta_1 \mathit{ShareWomen}_{i,t} + \beta_2 \mathit{PublicSector}_{i,t} \\ &+ \beta_3 \text{EmployerSize}_{i,t} + \beta_4 \mathit{FemaleLowerQuartile}_{i,t} \\ &+ \beta_5 \mathit{FemaleTopQuartile}_{i,t} + \epsilon_{i,t} \end{aligned} \tag{6.5}$$

$$\begin{aligned} \mathit{DiffMedianHourlyPercent}_{i,t} = \beta_0 &+ \beta_1 \mathit{ShareWomen}_{i,t} + \beta_2 \mathit{PublicSector}_{i,t} \\ &+ \beta_3 \mathit{EmployerSize}_{i,t} + \beta_4 \mathit{FemaleLowerQuartile}_{i,t} \\ &+ \beta_5 \mathit{FemaleTopQuartile}_{i,t} + \epsilon_{i,t} \end{aligned} \tag{6.6}$$

6.2 Assumptions of OLS Regression

To use the model properly and interpret its results with confidence, we first need to ensure that all fundamental Multiple Linear Regression (MLR) assumptions are satisfied. The first five assumptions are collectively known as Gauss-Markov assumptions, and by checking them, we ensure that our estimates are not only unbiased and consistent but also efficient.

First of all, the model being estimated should be **linear in parameters** (Wooldridge, 2013, p. 83). The MLR model specifically assumes linearity between the dependent variable and each independent variable, as well as the collective linear relationship. If this does not hold for the model, it cannot be correctly specified and therefore confidently interpreted (Burton, 2021). To ensure that the linear relationship is indeed a suitable approximation for our data, we implement the residuals versus fits plot, commonly used as a diagnostic graph in linear regression (Harrell & Harrell, 2015). Our results show that residuals are approximately randomly allocated around the zero line without any systematic pattern. Thus, we can assume satisfaction of this assumption.

Additionally, **random sampling** is required (Wooldridge, 2013, p. 84). In our specific research, we can state that while the selection process was based on regulatory requirements, we ensured a degree of randomness by including all eligible companies with complete information available on GOV.UK. Companies meeting the criterion of 250 or more employees are, by legislation, obliged to provide annual data, while smaller companies may participate voluntarily. We also incorporated available data from these companies, ensuring our sample adequately represents the population of interest and meets the second assumption.

The third assumption concerns **no perfect collinearity**, meaning that none of the independent variables should be constant and there should be no exact linear relationship among them. This assumption allows for a certain level of correlation; only the case of perfect correlation among independent variables is ruled out (Wooldridge, 2013, p. 84). While these assumptions are satisfied by the choice of our variables, we conduct tests for multicollinearity, which refers to a near-perfect linear relationship and can destabilize the OLS estimators while inflating the standard errors. To assess whether multicollinearity poses a problem in our model, we utilize Adjusted Generalized Standard Error Inflation Factor (aGSIF). This metric accounts for the presence of categorical variables with different numbers of non-reference levels, allowing us to compare them directly (Fox & Monette, 1992). When employing aGSIF, we need to modify the auxiliary threshold levels, detecting potential multicollinearity accordingly. A commonly used rule of thumb for Variance Inflation Factor (VIF) is either 5 or 10. However, in our setting, we need to take the square root of these numbers, resulting in 2.2 or 3.2 as our comparative levels. Values exceeding these thresholds may indicate collinearity issues in our regression analysis (Nahhas, 2024). Since the values of aGSIF do not exceed 2 in any of our models, we consider the third assumption to be fulfilled.

The fourth assumption, which altogether with the previous three ensures the unbiasedness of OLS estimators, requires the **expected value of the error term *u* to be zero**, regardless of any values of the independent variables (Wooldridge, 2013, p. 86). As the intercept is kept in the model, this assumption is satisfied. In fact, the intercept balances the average discrepancies between the actual data and the model's predictions, resulting in the zero mean of errors. By inclusion, we lose one degree of freedom, but we ensure overall unbiasedness. Thus, it is desirable to keep it in the model, even if we are not particularly interested in the predicted value of dependent variables, while all independent are equal to zero (Nau, 2020).

Lastly, we verify one additional assumption regarding **homoskedasticity**. It explicitly states that the error u maintains a constant variance irrespective of any values of the independent variables (Wooldridge, 2013, p. 93). Although our regression coefficient estimators remain still unbiased when this assumption is violated, they are less efficient. In addition, they can hinder proper statistical inference due to overestimation of standard errors, resulting in wider confidence intervals and inflated p-values compared to cases where homoskedasticity holds (Hayes & Cai, 2007). To test it, we conduct a series of Breusch-Pagan diagnostic tests with the null hypothesis of homoskedasticity (Breusch & Pagan, 1979). Rejecting the null hypothesis with significantly low p-value (< 0.01) in each specific case indeed signals a violation of this assumption in our specified models. To correct this issue, we implement heteroskedasticity-robust standard errors in our estimations (White, 1980). For complete information on the verification process for testable assumptions, please consult the do-file.

Chapter 7

Discussion of Empirical Results

In this chapter, we first explore the potential relationships through data visualization. Subsequently, we present and discuss the results of our models, aiming to either confirm or reject the hypotheses outlined in Chapter 4. Each model proposed in the preceding section is estimated for each year within the 6-year time frame. We interpret these findings using the common p-value thresholds of 0.01 and 0.05.

7.1 Data Visualization and Exploration

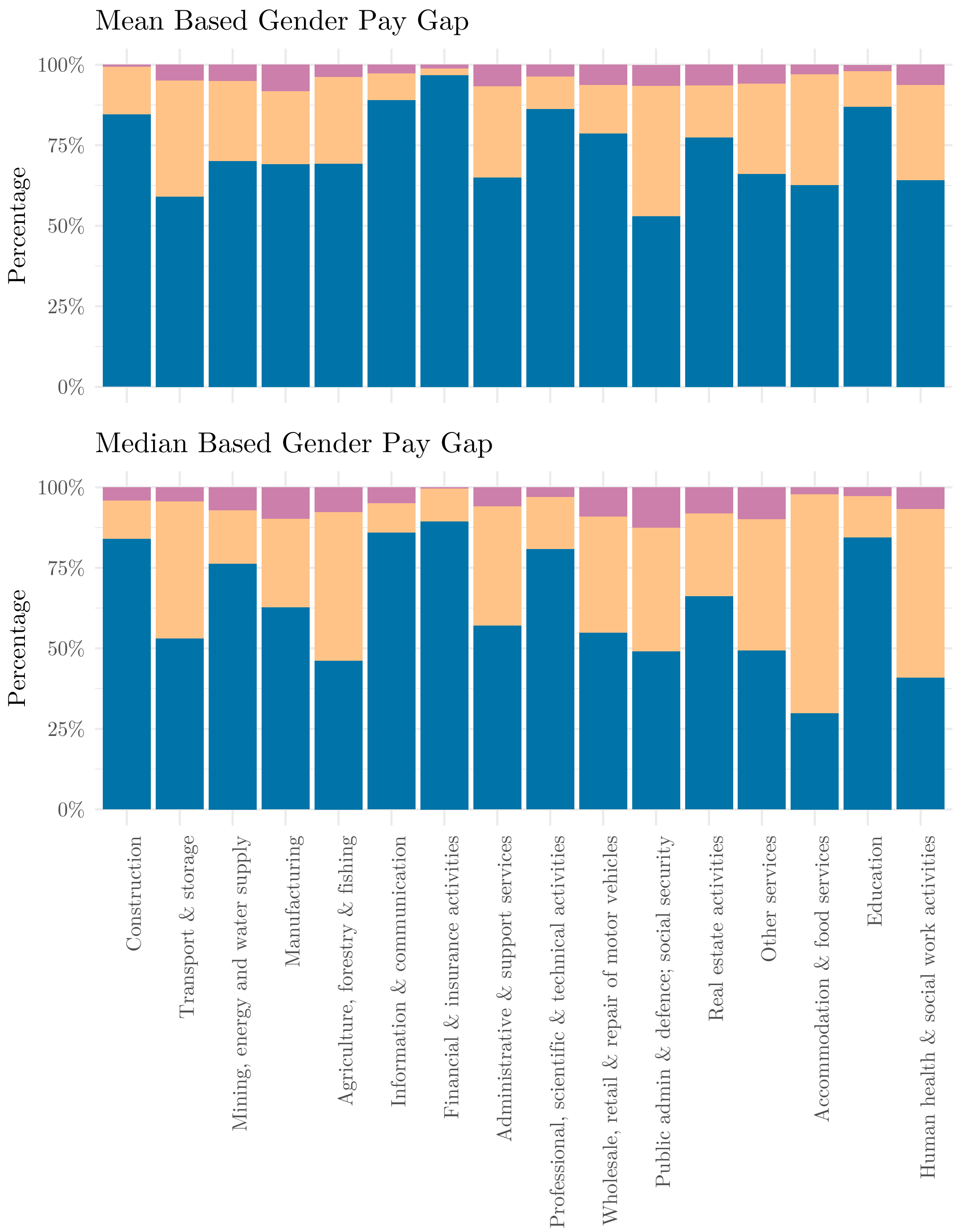
Before moving on to the methodology section, we provide some graphics to closely illustrate the interplay between industry types and the resulting GPGs. Figure 7.1 explores the GPG favourability across 16 industries identified in the final year of our analysis 2022/23. For this purpose, we classified each firm's GPG into one of three categories: gender-neutral if it fell within the range of -5% and 5%, in favour of men if it exceeded this range, and in favour of women if it fell below. This specific categorization was motivated by WGEA (2024b). This figure then specifically shows the percentage of firms favouring each gender or having a neutral GPG out of the total count of companies in each industry category. In this figure, both mean- and median-based GPG are depicted. To further account for the influence of gender diversity, constituting our second hypothesis, we rank the industries in ascending order from left to right based on the average share of women.

Firstly, we notice that in both cases, a significant portion of companies favours men, usually above 50%. This appears less dramatic when considering the median-based GPG, as it is less affected by high-income earners and, therefore, tends to fit more easily within the gender-neutral range. Although we may have expected to identify a visible trend, its presence might not be straightforward to determine because of the divergent performance among particular industries that are similar in terms of female distribution. For instance, we consider the contrast between Human health & social work activities and Education. Despite their similar average representation of women, companies in Education show a much higher preference for men than those in Human health & social work activities.

To further delve into this relationship, we designed a bubble plot, reflecting three distinct characteristics (see Figure 7.2). Along the horizontal x-axis, we depict the share of women, with industries situated further to the right demonstrating a higher average female representation. Vertically, on the y-axis, we depict the average GPG for each industry, computed as the average of all reporting firms within that industry category. Bubbles located higher up on the y-axis indicate poorer performance regarding the GPG. The size of each bubble corresponds to the scale of each industry type, specifically reflecting their prevalence in our dataset. This figure is based on the data collected in 2022/23 and employs the median-based GPG measure.

Upon looking at the figure, it is evident that certain sectors, specifically Construction, Financial & insurance activities, and Education, emerge as the worst-performing sectors. It is worth noting that, although there are significant differences in the average female representation across these sectors, yet their GPG results show surprisingly minimal differences. In contrast, many industries show positive outcomes in addressing the GPG, including those not predominantly female-dominated or gender-neutral, suggesting a complex mix of factors that influence the GPG beyond the workforce composition. Given the limited scope of industry categories and their variability in performance, identifying a clear trend might be challenging. As we aim for more precise and reliable conclusions, we will concentrate further on the regressions we proposed and their respective outcomes.

Figure 7.1: Distribution of GPG across Industry Categories (2022/23)



Note: Blue indicates a GPG favoring men pink denotes a GPG favoring women, and orange represents gender-neutral.

Figure 7.2: Share of Women vs. Average GPG by Industry Categories (2022/23)

7.2 Results from the First Set of Models

We start by presenting results derived from the first two regressions, with their corresponding findings detailed in Tables 7.2 and 7.3. Firstly, the low R-squared values indicate a limited predicting power in these models; therefore, we have to be cautious when interpreting these findings, despite their potential insights. To clarify, the reference group remains consistent across all regressions, with Construction as the reference category for industry and 250-499 for Employer Size.

Upon checking the results, it becomes evident that nearly all industry categories exhibit negative estimates throughout the reporting period, signalling lower GPGs compared to the reference category of Construction. This pattern holds when considering both measures of GPG, whether mean-based or median-based. The sole exception found was the Financial & insurance activities sector, having consistently higher GPG values of 6-8 percentage points throughout the period compared to the reference category. This effect is highly statistically significant, even at 1% level. This effect is, however, primarily observed when considering the mean-based GPG. In contrast, when using the median-based GPG in the second regression, the coefficients are mitigated and lose their significance to some extent. But still, this sector acts as the worst-performing. Conversely, sectors such as Human health & social work activities, Accommodation & food services or Public administration & defence; social security demonstrate the best performances with the most significant differences compared to Construction. These sectors consistently maintain favourable results for both measures of the GPG.

However, it should be noted that when analyzing industry estimates using the median-based GPG as the dependent variable, many estimates become lower, meaning more negative in our context, compared to the regression with mean-based GPG. This could plausibly be attributed to the presence of a skewed wage distribution. When there are few higher-income earners, their wages affect the mean-based GPG much more than the median-based. As a result, the mean-based GPG can be inflated by this attribute. This aspect can result in overestimating the GPG, resulting in smaller differences with the reference category compared to those using median wages. Considering the middle-income earner is often considered a better approach as it provides a more representative picture of the typical wage experienced by most workers without being influenced by disproportionately high wages barely earned by a few individuals. Consequently, when adopting this measure, many other industries exhibit more pronounced differences than the first model.

In most cases, the overall performance in time is that the resulting coefficients become higher over time, or more specifically, less negative. While fluctuations and exceptions exist, the overall trend observed across industry categories suggests a majority pattern. Based on this finding, we, however, cannot assess the absolute performance of each industry category. Every result is still only relative to the Construction category. Therefore, to provide a clearer perspective, Table 7.1 displays the evolution of the average GPG measures for Construction throughout the entire period. As can be observed, despite some fluctuations, both measures tend to decrease over time, with the resulting value in 2022/23 being lower than in 2017/18. As the performance of Construction improves over time in terms of the resulting GPG measures, the narrowing between Construction and the majority of other industries might stem from the fact that other industries in question have either progressed at a slower rate, remained relatively stagnant, or even performed slightly worse. However, determining the precise reasons behind these trends would require further investigation, which is not the primary aim of this study.

Table 7.1: Evolution of GPG in Construction

GPG (%)	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
Mean-based GPG	21.95	22.13	21.22	21.38	18.68	19.32
Median-based GPG	23.15	22.84	21.60	19.93	19.94	20.01

Yet, despite all these observations, we still have to keep in mind the lower predicting power of these first two models, which lies within the range of 0.1 to 0.2 for both models. The estimates and interpretation of results might change once we include controls for the specific redistribution of female workers within each company, which we address in the second set of models, discussing its corresponding results in the following section.

Table 7.2: Reg. Results: Industry Effects - mean

	Dependent variable:					
	DiffMeanHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
Accommodation & food services	-14.215*** (1.294)	-14.393*** (1.302)	-13.598*** (1.285)	-10.408*** (1.832)	-9.876*** (1.909)	-11.339*** (1.207)
Administrative & support services	-9.008*** (1.260)	-9.474*** (1.249)	-8.870*** (1.216)	-10.080*** (1.331)	-8.059*** (1.752)	-8.078*** (1.130)
Agriculture, forestry & fishing	-7.532** (3.089)	-8.064*** (3.102)	-8.858*** (2.660)	-9.268*** (2.812)	-3.360 (4.219)	-6.153** (2.658)
Education	-6.477*** (1.177)	-6.728*** (1.159)	-6.101*** (1.143)	-6.236*** (1.238)	-3.826** (1.618)	-4.710*** (1.036)
Financial & insurance activities	6.470*** (1.356)	6.269*** (1.353)	6.577*** (1.318)	5.957*** (1.402)	7.898*** (1.768)	5.646*** (1.199)
Human health & social work activities	-10.313*** (1.237)	-10.393*** (1.224)	-10.023*** (1.242)	-10.344*** (1.290)	-8.450*** (1.718)	-9.561*** (1.135)
Information & communication	-2.268* (1.322)	-2.499* (1.310)	-2.305* (1.326)	-3.678** (1.459)	-1.402 (1.751)	-2.988** (1.228)
Manufacturing	-8.373*** (1.181)	-9.204*** (1.164)	-8.966*** (1.153)	-8.875*** (1.260)	-8.212*** (1.649)	-9.211*** (1.080)
Mining, energy & water supply	-8.266*** (1.741)	-9.116*** (1.843)	-8.119*** (1.653)	-9.083*** (1.853)	-6.550*** (1.953)	-8.070*** (1.596)
Other services	-6.565*** (1.837)	-6.761*** (1.819)	-5.630*** (1.817)	-5.910*** (1.890)	-2.469 (2.252)	-5.025*** (1.737)
Professional, scientific & technical activities	-1.090 (1.316)	-1.975 (1.270)	-1.701 (1.246)	-2.475* (1.370)	-0.883 (1.712)	-2.886** (1.207)
Public admin & defence; social security	-14.704*** (1.315)	-15.449*** (1.300)	-15.455*** (1.283)	-15.494*** (1.353)	-13.269*** (1.722)	-14.968*** (1.170)
Real estate activities	-2.918 (2.289)	-3.219 (2.200)	-2.463 (2.344)	-3.955* (2.352)	-1.290 (2.644)	-2.962 (2.256)
Transport & storage	-12.228*** (1.469)	-11.016*** (1.450)	-11.098*** (1.416)	-12.085*** (1.492)	-9.788*** (1.839)	-9.785*** (1.303)
Wholesale, retail & repair of motor vehicles	-7.331*** (1.259)	-6.940*** (1.254)	-6.376*** (1.230)	-8.088*** (1.340)	-4.215** (1.704)	-5.571*** (1.131)
Public Sector	1.168* (0.653)	1.314** (0.644)	1.488** (0.639)	0.777 (0.612)	0.610 (0.630)	0.860 (0.621)
1000 to 4999	-0.510 (0.478)	-0.229 (0.469)	-0.237 (0.449)	0.111 (0.481)	-0.144 (0.530)	0.919** (0.464)
20,000 or more	-0.565 (1.349)	-2.066 (1.300)	-2.007** (0.917)	-2.029* (1.074)	-0.193 (1.182)	-0.550 (0.950)
500 to 999	-0.242 (0.473)	-0.294 (0.463)	0.028 (0.466)	-0.332 (0.492)	0.172 (0.489)	0.345 (0.478)
5000 to 19,999	1.311* (0.729)	1.121 (0.690)	1.287* (0.697)	1.480** (0.698)	1.840*** (0.675)	2.606*** (0.703)
Less than 250	-1.015 (2.087)	-0.176 (1.732)	-0.456 (2.126)	-4.228*** (1.444)	-3.361** (1.640)	-0.775 (1.417)
Not Provided	-1.762* (1.068)	-6.499 (5.778)	1.244 (2.647)	-0.764 (1.209)	1.609 (1.398)	-14.998*** (0.702)
Constant	22.145*** (1.115)	22.214*** (1.104)	21.225*** (1.087)	21.470*** (1.187)	18.679*** (1.565)	18.875*** (0.990)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.124	0.124	0.127	0.115	0.105	0.110
Adjusted R^2	0.120	0.120	0.123	0.111	0.101	0.106
Residual Std. Error (df = 4885)	12.541	12.460	12.308	12.866	13.596	12.509
F Statistic (df = 22; 4885)	31.368***	31.539***	32.317***	28.731***	26.183***	27.398***

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

Table 7.3: Reg. Results: Industry Effects - median

	Dependent variable:					
	DiffMedianHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
Accommodation & food services	-19.491*** (1.390)	-18.988*** (1.396)	-17.540*** (1.476)	-12.333*** (2.308)	-15.176*** (1.969)	-15.338*** (1.662)
Administrative & support services	-11.782*** (1.358)	-12.171*** (1.351)	-11.013*** (1.365)	-9.102*** (1.768)	-9.830*** (1.646)	-9.088*** (1.648)
Agriculture, forestry & fishing	-11.858*** (3.315)	-13.919*** (2.367)	-14.647*** (2.145)	-11.631*** (2.752)	-12.190*** (2.635)	-12.512*** (2.759)
Education	-1.073 (1.385)	-0.344 (1.377)	1.239 (1.416)	2.941 (1.804)	3.015* (1.667)	2.427 (1.679)
Financial & insurance activities	1.566 (1.478)	1.785 (1.466)	2.242 (1.479)	3.631* (1.855)	3.508** (1.743)	2.420 (1.735)
Human health & social work activities	-17.552*** (1.312)	-16.650*** (1.295)	-15.680*** (1.328)	-14.426*** (1.735)	-13.877*** (1.611)	-13.941*** (1.599)
Information & communication	-4.548*** (1.446)	-5.234*** (1.516)	-3.856*** (1.451)	-2.759 (1.851)	-3.772** (1.759)	-4.752*** (1.811)
Manufacturing	-12.362*** (1.283)	-12.324*** (1.268)	-11.588*** (1.311)	-9.483*** (1.735)	-11.126*** (1.577)	-11.509*** (1.588)
Mining, energy & water supply	-9.005*** (1.898)	-9.619*** (1.836)	-8.400*** (1.824)	-6.350*** (2.175)	-7.513*** (1.995)	-6.694*** (1.976)
Other services	-16.490*** (1.514)	-16.084*** (1.508)	-14.899*** (1.484)	-13.164*** (2.013)	-12.522*** (1.824)	-13.830*** (1.734)
Professional, scientific & technical activities	-5.511*** (1.430)	-5.166*** (1.397)	-4.436*** (1.407)	-3.127* (1.839)	-4.033** (1.672)	-4.581*** (1.666)
Public admin & defence; social security	-13.386*** (1.506)	-13.066*** (1.516)	-12.567*** (1.520)	-11.127*** (1.884)	-10.845*** (1.759)	-11.780*** (1.752)
Real estate activities	-9.147*** (2.348)	-8.362*** (2.286)	-7.446*** (2.245)	-6.685*** (2.594)	-5.767** (2.540)	-8.363*** (2.713)
Transport & storage	-13.288*** (1.587)	-12.591*** (1.613)	-11.641*** (1.598)	-10.258*** (1.979)	-10.646*** (1.848)	-10.782*** (1.813)
Wholesale, retail & repair of motor vehicles	-14.900*** (1.361)	-14.544*** (1.337)	-13.177*** (1.354)	-11.604*** (1.782)	-11.507*** (1.633)	-11.509*** (1.620)
Public Sector	-2.046** (0.804)	-2.656*** (0.809)	-2.640*** (0.783)	-3.078*** (0.795)	-4.027*** (0.788)	-3.358*** (0.777)
1000 to 4999	-0.197 (0.529)	0.593 (0.514)	0.854* (0.501)	1.400*** (0.531)	1.077** (0.523)	1.130** (0.494)
20,000 or more	-1.187 (1.163)	-1.392 (1.152)	-0.348 (1.053)	-0.813 (1.274)	0.866 (1.346)	-0.985 (1.300)
500 to 999	-0.055 (0.529)	0.426 (0.518)	0.587 (0.517)	0.363 (0.563)	0.292 (0.528)	0.044 (0.519)
5000 to 19,999	1.483* (0.770)	1.623** (0.788)	1.734** (0.728)	2.637*** (0.769)	2.883*** (0.734)	2.323*** (0.669)
Less than 250	3.283 (2.073)	0.365 (2.232)	−0.987 (2.174)	−2.005 (1.722)	−2.596* (1.451)	−0.819 (1.354)
Not Provided	-0.868 (1.301)	-8.936 (8.966)	1.715 (2.712)	1.095 (1.554)	3.439** (1.748)	−15.276*** (0.698)
Constant	23.140*** (1.208)	22.525*** (1.201)	21.183*** (1.235)	19.412*** (1.655)	19.542*** (1.516)	19.664*** (1.522)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.171	0.171	0.179	0.152	0.169	0.169
Adjusted R^2	0.167	0.167	0.175	0.149	0.166	0.166
Residual Std. Error (df = 4885)	13.954	13.887	13.595	14.583	13.994	13.507
F Statistic (df = 22; 4885)	45.783***	45.871***	48.346***	39.937***	45.240***	45.317***

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

7.3 Results from the Second Set of Models

After accounting for the redistribution of female workers within each company, the coefficients related to the industry categories became less negative, as evident in Tables 7.4 and 7.5. When individual firm circumstances are considered, some variance previously attributed solely to industry type is now explained. Including these variables improved the overall fit of the data, as indicated by the elevated R-squared. Regarding these two newly added variables, their effects are as anticipated. A higher concentration of female workers among the lowest-paid workers increases the GPG, while the category of highest-paid workers has the exact opposite effect. This holds for both choices of the dependent variable.

Firstly, we discuss the regression employing the mean-based GPG as the dependent variable. Its associated results can be found in Table 7.4. When we return our attention to the industry categories, Accommodation & food services and Public administration & defence; social security are still the best-performing sectors, as was observed previously. However, the Human health & social work activities sector is not anymore, after the inclusion of further controls, but continues to perform pretty well. Financial & insurance activities is still the worst-performing industry even in this third model, being the single sector consistently exhibiting positive GPGs in relation to the reference category. Other controls are not significant at all or only in some years of the reporting period. As previously emphasized, the median-based GPG is often considered a more telling measure of the real situation within the company, although it comes with limitations. We conduct regressions with both measures employed as dependent variables. Still, based on our theoretical consideration and practical evidence of higher R-squared values, we assume a slightly greater predictive power of the regression utilizing median-based GPG. Thus, we place greater emphasis on this regression, and the main findings are derived from it.

The newly added controls exhibit the expected signs, as demonstrated in Table 7.5. Specifically, increasing the proportion of women by 10 percentage points within the lower quartile is associated with an approximately 6 percentage point increase in the median-based GPG. A similar effect is observed for the top quartile but in the opposite direction. Both effects are statistically significant even at 1%.

We can make a new observation regarding the public sector: despite its insignificance when considering the mean-based GPG, its effect becomes significant in the case of median-based GPG, with the resulting negative sign. This aligns with the findings of Olsen *et al.* (2018), where the presence in the public sector is cited as one of the protective factors against the GPG. In terms of company size, no effect is consistently significant, except for the lower GPG observed in the largest companies, those with 20,000 or more employees, compared to the reference category of companies with 250-499 employees.

When turning our attention back to industry categories, our primary interest in these regressions, we can notice changes. First of all, we observe an overall rise in the significance of industrial effects, with nearly all industry-specific effects showing statistical significance at the 1% level throughout the reporting period. Second of all, Financial & insurance activities are no longer the worst-performing industry, replaced by Education, identified as the single industry having a greater GPG relative to Construction. This effect is statistically significant at 5% in the first year and at 1% in the subsequent years. Surprisingly, Education, one of the only two female-dominated industries in our dataset with the second-largest average proportion of female workers, consistently performs the worst across all years. All other industries, even when diverging in magnitudes, share the collective better performance relative to Construction, including those identified as male-dominated. As noted earlier, Construction maintains the lowest proportion of female workers, leading to a potential conclusion that this factor influences its comparatively poorer performance. However, Education, in contrast, is female-dominated, with a significantly higher average proportion of women workers, approximately 60% greater, and still performs worse. Hence, the potential link between the average gender composition within industries and the GPG will be of interest in the following section and cannot be stated at this point. Nevertheless, thanks to the evidence presented, we can support our first hypothesis that the GPG is dependent on industry affiliation. The industry-specific effect is statistically significant by most industry types even at 1% level, providing us with credible evidence to reach this conclusion. Whether this can be somehow attributed to the average gender structure within industries is the subject of our second hypothesis, which will be tested in our next section.

Table 7.4: Reg. Results: Industry and Quartile Effects - mean

	Dependent variable:					
	DiffMeanHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
Accommodation & food services	−9.045*** (0.998)	−8.743*** (1.049)	−6.779*** (1.036)	−6.028*** (1.461)	−5.384*** (1.634)	−5.886*** (1.088)
Administrative & support services	−5.528*** (0.881)	−5.947*** (0.974)	−5.095*** (0.841)	−5.975*** (0.924)	−4.196*** (1.458)	−5.090*** (0.880)
Agriculture, forestry & fishing	−5.486** (2.231)	−6.368*** (2.348)	−4.825** (2.167)	−7.958*** (2.940)	−1.053 (4.320)	−5.257** (2.435)
Education	−4.938*** (0.965)	−5.502*** (1.044)	−4.093*** (0.946)	−3.562*** (0.989)	−1.740 (1.377)	−3.228*** (0.951)
Financial & insurance activities	1.878** (0.921)	1.628 (1.006)	2.798*** (0.897)	2.176** (0.977)	4.097*** (1.411)	2.422*** (0.889)
Human health & social work activities	−2.436** (1.091)	−3.070*** (1.170)	−2.289** (1.111)	−1.970* (1.127)	−0.775 (1.533)	−2.409** (1.134)
Information & communication	−1.635* (0.950)	−1.838* (1.013)	−1.506 (0.942)	−2.658** (1.047)	−0.132 (1.452)	−1.681* (0.918)
Manufacturing	−4.854*** (0.804)	−5.693*** (0.874)	−5.210*** (0.763)	−5.499*** (0.842)	−4.504*** (1.395)	−5.216*** (0.797)
Mining, energy and water supply	−5.264*** (1.203)	−6.166*** (1.460)	−5.412*** (1.116)	−6.185*** (1.101)	−4.204*** (1.486)	−5.854*** (1.017)
Other services	−2.269 (1.664)	−3.153* (1.685)	−0.935 (1.643)	−1.799 (1.660)	1.606 (2.087)	−0.688 (1.670)
Professional, scientific & technical activities	−1.371 (0.917)	−1.995** (0.964)	−1.353 (0.851)	−2.084** (0.922)	−0.299 (1.384)	−1.930** (0.906)
Public admin & defence; social security	−9.055*** (1.011)	−9.479*** (1.082)	−8.556*** (0.986)	−7.395*** (1.034)	−5.518*** (1.469)	−7.623*** (1.005)
Real estate activities	−2.653 (1.746)	−3.753** (1.720)	−1.742 (1.751)	−2.547 (1.782)	−0.525 (2.126)	−1.891 (1.786)
Transport & storage	−6.792*** (0.954)	−6.345*** (1.075)	−5.785*** (0.916)	−6.643*** (0.975)	−4.613*** (1.545)	−4.700*** (0.949)
Wholesale, retail & repair of motor vehicles	−2.213** (0.888)	−2.547*** (0.965)	−1.460* (0.860)	−3.401*** (0.929)	0.315 (1.428)	−1.935** (0.843)
Public Sector	0.071 (0.508)	0.330 (0.510)	0.162 (0.515)	−0.819 (0.507)	−0.498 (0.550)	−0.148 (0.540)
1000 to 4999	−0.139 (0.364)	−0.152 (0.361)	−0.030 (0.346)	−0.124 (0.370)	−0.570 (0.445)	0.651* (0.377)
20,000 or more	−1.324 (1.133)	−3.016** (1.172)	−3.301*** (0.903)	−2.414*** (0.859)	−2.233** (1.001)	−1.851* (1.025)
500 to 999	0.244 (0.350)	0.176 (0.362)	0.171 (0.370)	−0.260 (0.377)	0.354 (0.397)	0.466 (0.393)
5000 to 19,999	0.543 (0.491)	0.217 (0.487)	0.193 (0.497)	0.347 (0.519)	0.684 (0.533)	1.501*** (0.558)
Less than 250	−1.954 (1.547)	−2.059 (1.344)	−1.563 (1.545)	−2.995** (1.243)	−2.318* (1.328)	−1.315 (1.182)
Not Provided	−2.010** (0.803)	−4.509* (2.309)	−0.585 (1.743)	0.097 (1.000)	1.568 (1.179)	−3.607*** (0.749)
FemaleLowerQuartile	0.575*** (0.013)	0.551*** (0.014)	0.540*** (0.014)	0.539*** (0.015)	0.527*** (0.016)	0.517*** (0.016)
FemaleTopQuartile	−0.541*** (0.014)	−0.515*** (0.015)	−0.515*** (0.015)	−0.529*** (0.014)	−0.513*** (0.016)	−0.488*** (0.016)
Constant	8.741*** (0.862)	9.555*** (0.929)	8.946*** (0.804)	9.677*** (0.908)	7.410*** (1.500)	7.481*** (0.847)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.511	0.479	0.472	0.472	0.400	0.409
Adjusted R^2	0.509	0.477	0.470	0.469	0.397	0.406
Residual Std. Error (df = 4883)	9.367	9.612	9.571	9.938	11.138	10.196
F Statistic (df = 24; 4883)	212.960***	187.154***	182.130***	181.803***	135.600***	140.735***

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

Table 7.5: Reg. Results: Industry and Quartile Effects - median

	Dependent variable:					
	DiffMedianHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
Accommodation & food services	−12.532*** (1.034)	−11.146*** (1.055)	−8.722*** (1.160)	−7.042*** (1.971)	−10.030*** (1.640)	−8.165*** (1.444)
Administrative & support services	−7.076*** (0.920)	−7.177*** (0.946)	−6.002*** (0.912)	−4.211*** (1.320)	−5.409*** (1.286)	−5.009*** (1.337)
Agriculture, forestry & fishing	−9.330*** (2.751)	−11.683*** (1.897)	−9.825*** (1.794)	−10.060*** (3.377)	−9.558*** (2.795)	−11.305*** (2.628)
Education	2.345** (1.082)	2.908*** (1.076)	5.056*** (1.091)	6.491*** (1.419)	5.453*** (1.351)	5.703*** (1.363)
Financial & insurance activities	−3.066*** (0.952)	−3.067*** (0.945)	−1.557* (0.923)	−0.492 (1.308)	−0.790 (1.286)	−0.659 (1.309)
Human health & social work activities	−6.567*** (1.135)	−5.827*** (1.098)	−5.013*** (1.111)	−4.293*** (1.468)	−5.072*** (1.390)	−3.950*** (1.388)
Information & communication	−3.431*** (0.978)	−4.017*** (1.048)	−2.577*** (0.962)	−1.480 (1.335)	−2.311* (1.368)	−2.891** (1.421)
Manufacturing	−8.186*** (0.851)	−7.971*** (0.859)	−7.123*** (0.857)	−5.590*** (1.296)	−6.903*** (1.244)	−6.852*** (1.305)
Mining, energy and water supply	−5.453*** (1.282)	−6.007*** (1.268)	−5.217*** (1.176)	−3.025* (1.582)	−4.841*** (1.371)	−4.122*** (1.410)
Other services	−10.667*** (1.069)	−10.802*** (1.062)	−8.674*** (1.025)	−8.224*** (1.612)	−7.856*** (1.411)	−8.049*** (1.414)
Professional, scientific & technical activities	−5.229*** (0.966)	−4.517*** (0.958)	−3.480*** (0.925)	−2.495* (1.350)	−3.344*** (1.272)	−2.898** (1.294)
Public admin & defence; social security	−5.662*** (1.133)	−4.599*** (1.118)	−3.486*** (1.098)	−1.520 (1.468)	−1.981 (1.418)	−2.208 (1.444)
Real estate activities	−8.146*** (1.391)	−8.252*** (1.397)	−5.994*** (1.332)	−4.877*** (1.686)	−4.871*** (1.698)	−6.476*** (1.859)
Transport & storage	−7.073*** (1.052)	−7.035*** (1.170)	−5.532*** (1.035)	−4.068*** (1.456)	−4.763*** (1.472)	−5.067*** (1.478)
Wholesale, retail & repair of motor vehicles	−8.459*** (0.917)	−8.686*** (0.906)	−7.000*** (0.881)	−6.097*** (1.321)	−6.333*** (1.275)	−6.860*** (1.293)
Public Sector	−3.515*** (0.660)	−4.054*** (0.644)	−4.350*** (0.646)	−4.960*** (0.659)	−5.295*** (0.688)	−4.732*** (0.680)
1000 to 4999	0.272 (0.377)	0.718* (0.369)	1.134*** (0.368)	1.145*** (0.397)	0.595 (0.401)	0.873** (0.387)
20,000 or more	−1.905** (0.867)	−2.448** (1.045)	−1.757* (1.025)	−1.239 (0.933)	−1.448 (1.165)	−2.316* (1.289)
500 to 999	0.514 (0.389)	0.977** (0.388)	0.764* (0.401)	0.445 (0.440)	0.499 (0.413)	0.184 (0.415)
5000 to 19,999	0.660 (0.498)	0.622 (0.537)	0.528 (0.492)	1.359** (0.532)	1.572*** (0.579)	1.095** (0.502)
Less than 250	2.301 (1.748)	−1.876 (1.508)	−2.172 (1.380)	−0.595 (1.433)	−1.408 (1.066)	−1.434 (1.024)
Not Provided	−1.039 (0.951)	−6.150 (4.184)	−0.226 (1.638)	2.091 (1.322)	3.393** (1.402)	−3.210*** (0.737)
FemaleLowerQuartile	0.649*** (0.015)	0.645*** (0.016)	0.612*** (0.016)	0.610*** (0.019)	0.599*** (0.018)	0.577*** (0.018)
FemaleTopQuartile	−0.645*** (0.016)	−0.640*** (0.016)	−0.616*** (0.016)	−0.610*** (0.019)	−0.584*** (0.017)	−0.579*** (0.018)
Constant	8.446*** (0.907)	8.210*** (0.914)	7.684*** (0.901)	6.207*** (1.380)	6.756*** (1.352)	7.480*** (1.414)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.563	0.557	0.535	0.499	0.503	0.481
Adjusted R^2	0.561	0.555	0.532	0.496	0.501	0.478
Residual Std. Error (df = 4883)	10.131	10.153	10.236	11.215	10.824	10.679
F Statistic (df = 24; 4883)	262.296***	256.010***	233.823***	202.581***	206.063***	188.584***
Note:						*p<0.1; **p<0.05; ***p<0.01

7.4 Results from the Third Set of Models

The last section aims to test the second hypothesis. In the final two models, we use the same set of control variables; we simply substitute the set of industry dummies with the *ShareWomen* variable as a measure of industrial segregation. In line with our previous approach, we interpret the results using both available measures of GPG, with their corresponding results detailed in Tables 7.6 and 7.7. Again, the model incorporating median-based GPG exhibits greater R-squared, indicating greater overall power. The effect of quartile controls is consistent with previous regressions, and the effect of public sector affiliation is again identified as a protective factor. All these effects are statistically significant even at 1% level throughout the entire period in both models. The effect of the last set of controls, *EmployerSize*, remains similar as well, with the largest companies exhibiting lower GPG compared to those with 250-499 employees. This effect appears consistently observable and significant only when employing the median-based GPG as the dependent variable, following the pattern from our previous set of models.

However, our primary topic of interest is the *ShareWomen* variable. When considering the mean-based GPG, *ShareWomen* variable is statistically significant at least at 5% except for the first two years. The significance is even higher when employing the median-based GPG as the dependent variable, specifically at 1%, indicating a very high level of confidence in the relationship. Based on our data, the overall positive sign suggests that industries with a higher average proportion of female workers tend to have larger GPGs. Notably, the estimated coefficient of this variable nearly doubles during the reporting period in both regressions. Despite the *ShareWomen* variable showing relatively steady development, as indicated in Table B.1, its influence increases over time. However, since the initial estimates are small, the final estimates do not become notably larger. When looking at the estimated coefficient in the year 2022/23, using the median-based GPG as the dependent variable, we find that a 10 percentage point increase in the average proportion of women within a specific industry corresponds to roughly a 1 percentage point increase in the resulting GPG of a firm stating affiliation within that industry, on average. While the high statistical significance proves the reliability of this effect, its practical significance is deemed modest due to the small estimated impact. Even in the final year 2022/23, the effect remains quite negligible despite being the greatest among all identified.

As outlined in Chapter 5, we utilized a classification method to connect the average gender composition within industries with the GPG by categorizing them into three groups based on their gender dominance (see Table 5.3). To ensure we do not overlook this classification, we ran two additional regressions in which we replaced the *ShareWomen* variable with the *occupation* variable reflecting gender dominance. Female dominance served as the reference category. The results of these additional regressions revealed patterns similar to those observed in previous models. Notably, GPGs are lower in both male-dominated and gender-neutral industries compared to female-dominated ones. Again, this effect is more pronounced when considering median-based GPG as the dependent variable but remains observable and statistically significant in both cases. Such findings further support our previous result that industries with a greater average concentration of women also tend to exhibit larger GPGs. Since this step involves additional regressions, detailed results can be found in Appendix B for reference.

Based on this evidence, our second hypothesis finds support as well, suggesting the existence of a significant relationship between the average gender structure of industries and the resulting GPG of companies operating within those sectors. Yet, the magnitude of the coefficient is relatively minor. These results indicate that industrial segregation may not be as influential in driving the GPG, or at least its effect is not observed as highly influential in our specific case when examining it from the perspective of individual companies. While gender disparities indeed vary based on industrial affiliation, as indicated by the significantly low p-values observed in the preceding sections, the average gender composition within industries, as a measure of industrial segregation, does not appear to have a notable overall real-world influence. Plausibly, the extent of the GPG might depend more on the type of work individuals are engaged in, regardless of the specific industry type they are employed in. As previously emphasized, while occupations as well as industries are clearly defined, the presence of occupations across multiple or all industry categories can complicate the interpretation of their effects, as evidenced by our analysis.

Table 7.6: Reg. Results: Share of Women and Quartile Effects - mean

	Dependent variable:					
	DiffMeanHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
ShareWomen	0.019	0.021	0.033**	0.036***	0.057***	0.038**
	(0.013)	(0.013)	(0.013)	(0.014)	(0.017)	(0.015)
FemaleLowerQuartile	0.598***	0.576***	0.567***	0.566***	0.553***	0.543***
	(0.012)	(0.014)	(0.013)	(0.014)	(0.015)	(0.015)
FemaleTopQuartile	−0.568***	−0.547***	−0.548***	−0.553***	−0.546***	−0.519***
	(0.013)	(0.014)	(0.014)	(0.014)	(0.016)	(0.015)
Public Sector	−3.196***	−2.894***	−3.184***	−3.008***	−2.987***	−2.897***
	(0.349)	(0.331)	(0.327)	(0.339)	(0.359)	(0.342)
1000 to 4999	0.174	0.176	0.300	0.107	−0.297	0.896**
	(0.370)	(0.364)	(0.349)	(0.377)	(0.445)	(0.376)
20,000 or more	−1.166	−2.751**	−2.874***	−2.422***	−1.811*	−1.761*
	(1.270)	(1.322)	(0.962)	(0.916)	(0.976)	(1.030)
500 to 999	0.183	0.159	0.235	−0.184	0.455	0.638
	(0.358)	(0.369)	(0.377)	(0.383)	(0.404)	(0.397)
5000 to 19,999	1.250**	0.835	0.779	0.707	1.186**	1.955***
	(0.548)	(0.528)	(0.532)	(0.551)	(0.545)	(0.570)
Less than 250	−1.302	−1.978	−1.424	−3.069**	−1.942	−1.079
	(1.508)	(1.357)	(1.548)	(1.221)	(1.302)	(1.166)
Not Provided	−1.890**	−2.763*	−1.140	0.269	1.578	−4.971***
	(0.788)	(1.487)	(1.878)	(1.004)	(1.179)	(0.600)
Constant	4.024***	4.365***	4.053***	3.772***	2.882***	2.351***
	(0.511)	(0.530)	(0.497)	(0.514)	(0.581)	(0.568)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.478	0.446	0.440	0.446	0.374	0.383
Adjusted R^2	0.477	0.445	0.439	0.445	0.372	0.382
Residual Std. Error (df = 4897)	9.664	9.897	9.847	10.163	11.363	10.399
F Statistic (df = 10; 4897)	449.147***	394.633***	384.523***	394.538***	292.167***	304.413***

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

Table 7.7: Reg. Results: Share of Women and Quartile Effects - median

	Dependent variable:					
	DiffMedianHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
ShareWomen	0.050***	0.059***	0.072***	0.080***	0.067***	0.099***
	(0.013)	(0.012)	(0.013)	(0.015)	(0.014)	(0.013)
FemaleLowerQuartile	0.705***	0.700***	0.674***	0.659***	0.656***	0.635***
	(0.014)	(0.015)	(0.015)	(0.018)	(0.018)	(0.017)
FemaleTopQuartile	−0.680***	−0.674***	−0.655***	−0.641***	−0.615***	−0.619***
	(0.015)	(0.015)	(0.015)	(0.019)	(0.017)	(0.017)
Public Sector	−2.192***	−2.180***	−2.786***	−3.140***	−3.082***	−2.897***
	(0.423)	(0.400)	(0.397)	(0.418)	(0.427)	(0.414)
1000 to 4999	−0.695*	−0.204	0.326	0.468	0.018	0.513
	(0.394)	(0.388)	(0.393)	(0.418)	(0.419)	(0.408)
20,000 or more	−4.216***	−4.491***	−3.684***	−3.159***	−3.268***	−4.192***
	(1.021)	(1.272)	(1.167)	(0.991)	(1.266)	(1.362)
500 to 999	0.178	0.675	0.624	0.421	0.605	0.378
	(0.414)	(0.416)	(0.430)	(0.467)	(0.444)	(0.444)
5000 to 19,999	−1.335***	−1.384***	−1.412***	−0.601	−0.427	−0.820
	(0.494)	(0.513)	(0.477)	(0.512)	(0.568)	(0.505)
Less than 250	1.937	−3.029*	−2.934**	−1.830	−2.548**	−1.869*
	(1.837)	(1.623)	(1.344)	(1.465)	(1.064)	(1.057)
Not Provided	−1.194	−7.161*	−1.935	2.721*	3.584**	−3.026***
	(1.027)	(3.924)	(1.888)	(1.443)	(1.525)	(0.558)
Constant	−0.651	−1.126*	−1.132*	−1.443**	−1.636**	−1.976***
	(0.616)	(0.603)	(0.612)	(0.687)	(0.642)	(0.646)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.509	0.503	0.476	0.445	0.445	0.424
Adjusted R^2	0.508	0.502	0.475	0.444	0.444	0.423
Residual Std. Error (df = 4897)	10.725	10.742	10.844	11.782	11.421	11.232
F Statistic (df = 10; 4897)	507.743***	495.474***	445.289***	393.268***	393.091***	360.868***

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

Chapter 8

Conclusion

This thesis aimed to explore the dynamics of the Gender Pay Gap, an issue drawing increasing attention nowadays with the current greater emphasis on workplace equality. This area is complex and involves a range of interrelated aspects. While occupational segregation is frequently cited and studied, this work shifted its focus to industrial segregation, particularly examining how the GPG depends on industry affiliation and the average proportion of women in these industries as a measure of segregation.

The legal framework in the UK has recently been updated, obliging employers with a workforce of 250 or more to regularly report their GPG-related statistics. This regulation supplied us with the most up-to-date information on company performance regarding gender pay. It served as our primary data source, providing most of the variables implemented in our research. These variables include two measures of the GPG - based on either mean or median wages - along with quartile distributions of employees, industry affiliation, and other firm-identifying information. When combined with our second dataset, providing details on average female representation within industries, we reached a total sample of 4908 companies reporting consistently throughout the entire 6-year period from 2017/18 to 2022/23, constituting our main entities studied.

To investigate our research questions, we proposed several models using both measures of the GPG as the dependent variable to capture the potential effects identified by each. Further, we employed the set of industry dummies or the average proportion of female workers as the main independent variables. Controls regarding the company size, public or private sector affiliation, and quartile distribution of employees were introduced to isolate these effects. Our approach utilized OLS estimation, performed separately for each year within the reporting period. The choice of OLS was justified mainly by our aim to retain time-invariant features in our analysis. To ensure the accuracy and reliability of the results, all models were adjusted to account for the presence of heteroskedasticity.

The initial finding confirmed the strong dependence of GPG on industry affiliation, with the majority of industries exhibiting smaller GPGs compared to Construction, selected as the reference category. Considering the median-based GPG, these differences were even more evident. Generally, this pattern was identified when employing both GPG measures as dependent variables, with this effect frequently being statistically significant even at 1% level. Regarding the performance over time, we observed that the differences in relation to the reference category tended to narrow.

Secondly, we indeed found a significant relationship between the average gender composition of industries, used as the industrial segregation indicator, and the resulting GPG of companies within these sectors. This effect emerged as significant for both GPG measures, except for the first two years when employing the mean-based GPG. However, regression using the median-based GPG had consistently revealed statistically significant results, confirming this effect at even 1% level, indicating a very high level of confidence in these findings. The overall positive sign suggests that a higher average concentration of women within industries is associated with larger resulting GPGs. This relationship mirrors the findings of Schneider *et al.* (2022), who identified a similar trend within the occupational context. Despite the high statistical significance of our identified effect, the estimated magnitude of the coefficient is relatively minor each year, resulting in limited practical significance and real-world implications.

8.1 Limitations and Future Research

While this work provides valuable insights, it does have its limitations, which became apparent as we delved into the analysis. Firstly, the legislation providing a framework for this thesis mandates a simplified reporting format, plausibly to aid all employers in calculating required metrics and reducing errors. However, this simplified approach to such a naturally complex issue has its drawbacks. Despite the requirement to include only full-pay relevant employees in the calculations, it fails to account for factors like level of experience or education. Although controlling for these factors could improve the overall data quality, it would greatly complicate the calculation process for individual companies.

Secondly, the legislation requires only companies exceeding a certain threshold of employee count to report. But there is an ongoing discussion about lowering this threshold to 150 and eventually to 50 employees (UK Parliament, 2016). If these changes are to be implemented, the scope of the research could be extended by including additional data from smaller companies, thereby enriching the study with insights from a broader range of businesses. Another potential direction for further research could involve considering alternative methodologies, such as the Hausman-Taylor model, which accounts for both the panel data structure and time-invariant features. However, given its advanced nature, we have opted for less complex methodologies that we are more confident in.

Despite limitations stemming from the design of GPG reporting, this research still makes valuable contributions to understanding gender inequality in remuneration and provides a solid foundation for further investigation in this field. Continued investigation into this issue is indeed called for, as gender pay disparity remains a prevalent phenomenon that is unlikely to disappear without efforts to address it.

Bibliography

Aliprantis, D., T. Dunne, & K. D. Fee (2011): "The Growing Difference in College Attainment between Women and Men." *Economic Commentary* **(2011-21)**.

Anderson, T., J. Forth, H. Metcalf, & S. Kirby (2001): *The Gender Pay Gap: Final Report to the DfEE.* Women and Equality Unit, Cabinet Office.

Atkins, S. (1986): "The Sex Discrimination Act 1975: The End of a Decade." *Feminist Review* **(24)**: pp. 57–70.

Bedaso, F. J. (2024): "Occupational Segregation and the Gender Wage Gap: Evidence from Ethiopia." *Technical report*, GLO Discussion Paper.

Berrington, A., J. Ellison, B. Kuang, S. Vasireddy, & H. Kulu (2021): "Recent trends in UK fertility and potential impacts of COVID-19." .

Blau, F. D. & L. M. Kahn (2017): "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of economic literature* **55(3)**: pp. 789–865.

Boll, C. & A. Lagemann (2018): *Gender pay gap in EU countries based on SES (2014).* Publications Office, European Commission - Directorate-General for Justice.

Bonny, B. P., A. Ajith, & S. Lokesh (2022): "Gender in Agriculture: A Development Perspective." In "Engendering Agricultural Development," pp. 1–14. CRC Press.

Böök, B., S. Burri, L. Senden, & A. Timmer (2021): *A comparative analysis of gender equality law in Europe 2020 - The 27 EU Member States, Albania,* *Iceland, Liechtenstein, Montenegro, North Macedonia, Norway, Serbia, Turkey and the United Kingdom compared*. Publications Office.

Brett Taylor, R. & A. L. Wilson (2023): "Legislating for a Post-Brexit Scotland: Scottish Parliamentary Scrutiny of UK Statutory Instruments on Retained EU Law." *Edinburgh Law Review* **27(1)**: pp. 34–63.

Breusch, T. S. & A. R. Pagan (1979): "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica: Journal of the econometric society* pp. 1287–1294.

Bryson, A., H. Joshi, B. Wielgoszewska, & D. Wilkinson (2020): "A Short History of the Gender Wage Gap in Britain." *Oxford Review of Economic Policy* **36(4)**: pp. 836–854.

Burton, A. L. (2021): "OLS (Linear) regression." *The Encyclopedia of Research Methods in Criminology and Criminal Justice* **2**: pp. 509–514.

Carranza, E., S. Das, & A. Kotikula (2019): *Gender-based Employment Segregation: Understanding Causes and Policy Interventions*. World Bank.

Charles, M. (2003): "Deciphering Sex Segregation: Vertical and Horizontal Inequalities in Ten National Labor Markets." *Acta sociologica* **46(4)**: pp. 267–287.

Close Your Pay Gap (2020): Women's jobs, men's jobs? Job segregation, and what it means for the Gender Pay Gap. [online]. [cit. 2024-02-29]. Retrieved from: `https://www.closeyourpaygap.org.uk/files/briefing-5-segregation-what-means-2020.pdf`.

Companies House (2024): Nature of business: Standard Industrial Classification (SIC) codes. [online]. [cit. 2024-03-20]. Retrieved from: `https://resources.companieshouse.gov.uk/sic/`.

Connolly, S., M. Aldrich, M. O'Brien, S. Speight, & E. Poole (2016): "Britain's slow movement to a gender egalitarian equilibrium: parents and employment in the UK 2001–13." *Work, employment and society* **30(5)**: pp. 838–857.

Department for Business and Trade (2023): "The Equality Act 2010 (Amendment) Regulations 2023." *Explanatory Memorandum 2023 No. 1424*, Department for Business and Trade, Government of the United Kingdom.

Dex, S., H. Sutherland, & H. Joshi (2000): "Effects of Minimum Wages on the Gender Pay Gap." *National Institute Economic Review* **173(1)**: pp. 80–88.

Dickens, L. (1995): "Road Blocks on the Route to Equality: The Failure of Sex Discrimination Legislation in Britain." *Melbourne University Law Review* **18**: p. 277.

Dickens, R. & A. Manning (2004): "Has the National Minimum Wage Reduced UK Wage Inequality?" *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **167(4)**: pp. 613–626.

Elsesser, K. (2019): The Gender Pay Gap And The Career Choice Myth. Forbes. [online]. [cit. 2024-03-15]. Retrieved from: `https://www.forbes.com/sites/kimelsesser/2019/04/01/the-gender-pay-gap-and-the-career-choice-myth/`.

European Commission (2017): *Special Eurobarometer 465: Gender Equality 2017.* Directorate-General for Justice and Consumers.

Fagan, C. & J. Rubery (2018): "Advancing Gender Equality through European Employment Policy: The Impact of the UK's EU Membership and the Risks of Brexit." *Social Policy and Society* **17(2)**: pp. 297–317.

Fell, E. V. & M. Dyban (2017): "Against Discrimination: Equality Act 2010 (UK)." *The European Proceedings of Social & Behavioural Sciences (EpSBS). Vol. 19: Lifelong Wellbeing in the World (WELLSO 2016).-Nicosia, 2017.* **192016**: pp. 188–194.

Fox, J. & G. Monette (1992): "Generalized Collinearity Diagnostics." *Journal of the American Statistical Association* **87(417)**: pp. 178–183.

FRED (2023): Employment Rate: Aged 25-54: Females for United Kingdom. [online]. [cit. 2024-04-15]. Retrieved from: `https://fred.stlouisfed.org/series/LREM25FEGBA156S`.

Girlguiding (2017): *Girls' Attitudes Survey 2017*. London: Girlguiding.

Goldin, C. (2014): "A Grand Gender Convergence: Its Last Chapter." *American economic review* **104(4)**: pp. 1091–1119.

Government Equalities Office (2019): Mandatory Gender Pay Gap Reporting: Summary of reported data for 2018/19. [online]. [cit. 2024-03-20]. Retrieved from: https://assets.publishing.service.gov.uk/media/5d35b836ed915d0d0e7a828c/Mandatory_Gender_Pay_Gap_Reporting_Summary_of_reported_data_for_2018-19.pdf.

Guerrina, R. & A. Masselot (2018): "Walking into the footprint of EU law: Unpacking the gendered consequences of Brexit." *Social Policy and Society* **17(2)**: pp. 319–330.

Harrell, Jr, F. E. & F. E. Harrell (2015): "General Aspects of Fitting Regression Models." *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis* pp. 13–44.

Hayes, A. F. & L. Cai (2007): "Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation." *Behavior research methods* **39**: pp. 709–722.

Hegewisch, A. & E. Mefferd (2021): "The Gender Wage Gap by Occupation, Race, and Ethnicity 2020." *Institute for Women's Policy Research* .

Hlavac, M. (2022): *Stargazer: Well-Formatted Regression and Summary Statistic Tables*. R package version 5.2.3.

HM Government (2019): *The case for change: How economic gender inequalities develop across the life course*. London: Government Equalities Office.

Huchet-Bodet, A., M. Albakri, & S. Neil (2019): *Attitudes to equalities: the British Social Attitudes Survey 2017*. London: Government Equalities Office. Reference: GEO-RB-006.

Jones, S. D. & S. Saulcy (2009): A Closer Look at Occupational Projections for Wyoming, 2006-2016. [online]. [cit. 2024-03-18]. Retrieved from: `https://doe.state.wy.us/lmi/occ_proj_06_16/text.htm`.

Joshi, H. & S. Owen (1987): "How long is a piece of elastic? The measurement of female activity rates in British censuses, 1951-1981." *Cambridge Journal of Economics* **11(1)**: pp. 55–74.

Kiernan, K. E. (1998): "Parenthood and Family Life in the United Kingdom." *Review of Population and Social Policy* **7**: pp. 63–81.

Leuze, K. & S. Strauss (2016): "Why do occupations dominated by women pay less? How 'female-typical' work tasks and working-time arrangements affect the gender wage gap among higher education graduates." *Work, employment and society* **30(5)**: pp. 802–820.

Martin, B. (2011): *Children at Play: Learning Gender in the Early Years*. ERIC.

Mavrikiou, P. M. & J. Angelovska (2020): "The impact of sex segregation by economic activity on the gender pay gap across europe." *UTMS Journal of Economics* **11(1)**.

Maříková, H. & R. M. Volejníčková (2022): *Dobrá praxe rovného odměňování - Velká Británie*. Ministerstvo práce a sociálních věcí České republiky.

Miller, C. C. (2016): "As women take over a male-dominated field, the pay drops." *The New York Times* **18**.

Miller, L., F. Neathey, E. Pollard, & D. Hill (2004): *Occupational Segregation, Gender Gaps and Skill Gaps*. Citeseer.

Mosakova, E. & K. Kizilova (2021): "Labor market in the UK in digital era: The gender dimension." *RUDN Journal of Sociology* **21**: pp. 512–519.

Nahhas, R. W. (2024): Introduction to Regression Methods for Public Health Using R: Collinearity. [online]. [cit. 2024-04-08]. Retrieved from: `https://www.bookdown.org/rwnahhas/RMPH/mlr-collinearity.html`.

Nau, R. (2020): Linear regression models: Additional notes on regression analysis. [online]. [cit. 2024-04-08]. Retrieved from: https://people.duke.edu/~rnau/regnotes.htm.

OECD (2017): *The Pursuit of Gender Equality: An Uphill Battle.* OECD Publishing.

OECD (2024): Employment: Time spent in paid and unpaid work, by sex. Organisation for Economic Co-operation and Development.[online]. [cit. 2024-02-25]. Retrieved from: https://stats.oecd.org/index.aspx?queryid=54757.

Olsen, W., V. Gash, S. Kim, & M. Zhang (2018): *The Gender Pay Gap in the UK: Evidence from the UKHLS.* Government Equalities Office Working Papers Series. United Kingdom: Department for Education. Reference: DFE-RR804.

ONS (2016): Women shoulder the responsibility of 'unpaid work'. [online]. [cit. 2024-02-25]. Retrieved from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/womenshouldertheresponsibilityofunpaidwork/2016-11-10.

Penner, A. M., T. Petersen, A. S. Hermansen, A. Rainey, I. Boza, M. M. Elvira, O. Godechot, M. Hällsten, L. F. Henriksen, F. Hou, A. K. Mrčela, J. King, N. Kodama, T. Kristal, A. Křížková, Z. Lippényi, S. M. Melzer, E. Mun, P. Apascaritei, D. Avent-Holt, N. Bandelj, G. Hajdu, J. Jung, A. Poje, H. Sabanci, M. Safi, M. Soener, D. Tomaskovic-Devey, & Z. Tufail (2023): "Within-job gender pay inequality in 15 countries." *Nature Human Behaviour* **7**.

Perales, F. (2013): "Occupational sex-segregation, specialized human capital and wages: Evidence from Britain." *Work, Employment and Society* **27**.

Petersen, T., V. Snartland, L. E. Becken, & K. M. Olsen (1997): "Within-Job Wage Discrimination and the Gender Wage Gap: The Case of Norway." *European Sociological Review* **13**.

Roantree, B. & K. Vira (2018): *The rise and rise of women's employment in the UK.* The Institute for Fiscal Studies.

ROSER, M. (2017): Fertility Rate: How does the number of children vary across the world and over time? What is driving the rapid global change? Our World in Data. [online]. [cit. 2024-02-26]. Retrieved from: `https://ourworldindata.org/fertility-rate`.

SANDERS, A. & J. FLAVELL (2023): "The direction of gender equality policy in Britain post-Brexit: towards a masculinised Westminster model." *Journal of European Public Policy* pp. 1–23.

SCHNEIDER, S., K. RENTZSCH, & A. SCHÜTZ (2022): "The gender pay gap is smaller in occupations with a higher ratio of men: Evidence from a national panel study." *Plos one* **17(7)**.

SCOTT, P. (2023): "From "Pin Money" to Careers: Britain's Late Move to Equal Pay, Its Consequences, and Broader Bmplications." *Enterprise & Society* pp. 1–26.

STATISTA (2023): Total fertility rate in Europe in 2023, by country. [online]. [cit. 2024-02-27]. Retrieved from: `https://www.statista.com/statistics/612074/fertility-rates-in-european-countries/`.

THE HOUSE OF COMMONS, TRADE AND INDUSTRY COMMITTEE (2005): *Jobs for the girls: The effect of occupational segregation on the gender pay gap*. London: The Stationery Office Limited.

TORRE, M. & J. A. JACOBS (2021): "The Gender Mobility Paradox: Gender Segregation and Women's Mobility Across Gender-Type Boundaries, 1970–2018." *Gender & Society* **35(6)**: pp. 853–883.

UK PARLIAMENT (2016): Making GPG reporting regulations work. [online]. [cit. 2024-04-18]. Retrieved from: `https://publications.parliament.uk/pa/cm201516/cmselect/cmwomeq/584/58411.htm`.

UN WOMEN (2020): "Analysis of the gender pay gap and gender inequality in the labour market in armenia." .

University of Kent (2020): Sex Discrimination Act 1975.[online]. [cit. 2024-02-20]. Retrieved from: https://womenslegallandmarks.com/2017/08/08/sex-discrimination-act-1975/.

Verashchagina, A., F. Bettio, E. Commission, S. A. Directorate-General for Employment, & Inclusion (2009): *Gender segregation in the labour market - Root causes, implications and policy responses in the EU*. Publications Office.

WGEA (2024a): What is the gender pay gap? [online]. [cit. 2024-03-14]. Retrieved from: https://www.wgea.gov.au/the-gender-pay-gap.

WGEA (2024b): Employer Gender Pay Gaps Snapshot. [online]. [cit. 2024-04-15]. Retrieved from: https://www.wgea.gov.au/publications/employer-gender-pay-gaps-snapshot.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica: journal of the Econometric Society* pp. 817–838.

Wooldridge, J. (2013): *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning.

Yavorsky, J. E. & J. Dill (2020): "Unemployment and men's entrance into female-dominated jobs." *Social Science Research* **85**.

Zabalza, A. & Z. Tzannatos (1985): "The Effect of Britain's Anti-Discriminatory Legislation on Relative Pay and Employment." *The Economic Journal* **95(379)**: pp. 679–699.

Appendix A

Appendix A

Table A.1: Full List of SIC Codes in each Industry Category

Section	Industry	SIC Codes involved
A	Agriculture, forestry & fishing	01110, 01120, 01130, 01140, 01150, 01160, 01190, 01210, 01220, 01230, 01240, 01250, 01260, 01270, 01280, 01290, 01300, 01410, 01420, 01430, 01440, 01450, 01460, 01470, 01490, 01500, 01610, 01621, 01629, 01630, 01640, 01700, 02100, 02200, 02300, 02400, 03110, 03120, 03210, 03220
B,D,E	Mining, energy and water supply	05101, 05102, 05200, 06100, 06200, 07100, 07210, 07290, 08110, 08120, 08910, 08920, 08930, 08990, 09100, 09900, 35110, 35120, 35130, 35140, 35210, 35220, 35230, 35300, 36000, 37000, 38110, 38120, 38210, 38220, 38310, 38320, 39000

Continued on next page

Table A.1 – continued from previous page

Section	Industry	SIC Codes involved
C	Manufacturing	10110, 10120, 10130, 10200, 10310, 10320, 10390, 10410, 10420, 10511, 10512, 10519, 10520, 10611, 10612, 10620, 10710, 10720, 10730, 10810, 10821, 10822, 10831, 10832, 10840, 10850, 10860, 10890, 10910, 10920, 11010, 11020, 11030, 11040, 11050, 11060, 11070, 12000, 13100, 13200, 13300, 13910, 13921, 13922, 13923, 13931, 13939, 13940, 13950, 13960, 13990, 14110, 14120, 14131, 14132, 14141, 14142, 14190, 14200, 14310, 14390, 15110, 15120, 15200, 16100, 16210, 16220, 16230, 16240, 16290, 17110, 17120, 17211, 17219, 17220, 17230, 17240, 17290, 18110, 18121, 18129, 18130, 18140, 18201, 18202, 18203, 19100, 19201, 19209, 20110, 20120, 20130, 20140, 20150, 20160, 20170, 20200, 20301, 20302, 20411, 20412, 20420, 20510, 20520, 20530, 20590, 20600, 21100, 21200, 22110, 22190, 22210, 22220, 22230, 22290, 23110, 23120, 23130, 23140, 23190, 23200, 23310, 23320, 23410, 23420, 23430, 23440, 23490, 23510, 23520, 23610, 23620, 23630, 23640, 23650, 23690, 23700, 23910, 23990, 24100, 24200, 24310, 24320, 24330, 24340, 24410, 24420, 24430, 24440, 24450, 24460, 24510, 24520, 24530, 24540, 25110, 25120, 25210, 25290, 25300, 25400, 25500, 25610, 25620, 25710, 25720, 25730, 25910, 25920, 25930, 25940, 25990, 26110, 26120, 26200, 26301, 26309, 26400, 26511, 26512, 26513, 26514, 26520, 26600, 26701, 26702, 26800, 27110, 27120, 27200, 27310, 27320, 27330, 27400, 27510, 27520, 27900, 28110, 28120, 28131, 28132, 28140, 28150, 28210, 28220, 28230, 28240, 28250, 28290, 28301, 28302, 28410, 28490, 28910, 28921, 28922, 28923, 28930, 28940, 28950, 28960, 28990, 29100, 29201, 29202, 29203, 29310, 29320, 30110, 30120, 30200, 30300, 30400, 30910, 30920, 30990, 31010, 31020, 31030, 31090, 32110, 32120, 32130, 32200, 32300, 32401, 32409, 32500, 32910, 32990, 33110, 33120, 33130, 33140, 33150, 33160, 33170, 33190, 33200

Continued on next page

Table A.1 – continued from previous page

Section	Industry	SIC Codes involved
F	Construction	41100, 41201, 41202, 42110, 42120, 42130, 42210, 42220, 42910, 42990, 43110, 43120, 43130, 43210, 43220, 43290, 43310, 43320, 43330, 43341, 43342, 43390, 43910, 43991, 43999
G	Wholesale, retail & repair of motor vehicles	45111, 45112, 45190, 45200, 45310, 45320, 45400, 46110, 46120, 46130, 46140, 46150, 46160, 46170, 46180, 46190, 46210, 46220, 46230, 46240, 46310, 46320, 46330, 46341, 46342, 46350, 46360, 46370, 46380, 46390, 46410, 46420, 46431, 46439, 46440, 46450, 46460, 46470, 46480, 46491, 46499, 46510, 46520, 46610, 46620, 46630, 46640, 46650, 46660, 46690, 46711, 46719, 46720, 46730, 46740, 46750, 46760, 46770, 46900, 47110, 47190, 47210, 47220, 47230, 47240, 47250, 47260, 47290, 47300, 47410, 47421, 47429, 47430, 47510, 47520, 47530, 47540, 47591, 47599, 47610, 47620, 47630, 47640, 47650, 47710, 47721, 47722, 47730, 47741, 47749, 47750, 47760, 47770, 47781, 47782, 47789, 47791, 47799, 47810, 47820, 47890, 47910, 47990
H	Transport & storage	49100, 49200, 49311, 49319, 49320, 49390, 49410, 49420, 49500, 50100, 50200, 50300, 50400, 51101, 51102, 51210, 51220, 52101, 52102, 52103, 52211, 52212, 52213, 52219, 52220, 52230, 52241, 52242, 52243, 52290, 53100, 53201, 53202
I	Accommodation & food services	55100, 55201, 55202, 55209, 55300, 55900, 56101, 56102, 56103, 56210, 56290, 56301, 56302
J	Information & communication	58110, 58120, 58130, 58141, 58142, 58190, 58210, 58290, 59111, 59112, 59113, 59120, 59131, 59132, 59133, 59140, 59200, 60100, 60200, 61100, 61200, 61300, 61900, 62011, 62012, 62020, 62030, 62090, 63110, 63120, 63910, 63990
K	Financial & insurance activities	64110, 64191, 64192, 64201, 64202, 64203, 64204, 64205, 64209, 64301, 64302, 64303, 64304, 64305, 64306, 64910, 64921, 64922, 64929, 64991, 64992, 64999, 65110, 65120, 65201, 65202, 65300, 66110, 66120, 66190, 66210, 66220, 66290, 66300
L	Real estate activities	68100, 68201, 68202, 68209, 68310, 68320

Continued on next page

Table A.1 – continued from previous page

Section	Industry	SIC Codes involved
M	Professional scientific & technical activities	69101, 69102, 69109, 69201, 69202, 69203, 70100, 70210, 70221, 70229, 71111, 71112, 71121, 71122, 71129, 71200, 72110, 72190, 72200, 73110, 73120, 73200, 74100, 74201, 74202, 74203, 74209, 74300, 74901, 74902, 74909, 74990, 75000
N	Administrative & support services	77110, 77120, 77210, 77220, 77291, 77299, 77310, 77320, 77330, 77341, 77342, 77351, 77352, 77390, 77400, 78101, 78109, 78200, 78300, 79110, 79120, 79901, 79909, 80100, 80200, 80300, 81100, 81210, 81221, 81222, 81223, 81229, 81291, 81299, 81300, 82110, 82190, 82200, 82301, 82302, 82911, 82912, 82920, 82990
O	Public admin & defence; social security	84110, 84120, 84130, 84210, 84220, 84230, 84240, 84250, 8430
P	Education	85100, 85200, 85310, 85320, 85410, 85421, 85422, 85510, 85520, 85530, 85590, 85600
Q	Human health & social work activities	86101, 86102, 86210, 86220, 86230, 86900, 87100, 87200, 87300, 87900, 88100, 88910, 88990
R,S,T	Other services	90010, 90020, 90030, 90040, 91011, 91012, 91020, 91030, 91040, 92000, 93110, 93120, 93130, 93191, 93199, 93210, 93290, 94110, 94120, 94200, 94910, 94920, 94990, 95110, 95120, 95210, 95220, 95230, 95240, 95250, 95290, 96010, 96020, 96030, 96040, 96090, 97000, 98000, 98100, 98200

Note: The table is based on the Companies House (2024) list, rewritten for clarity.

Appendix B

Appendix B

Table B.1: Share of Women Classified by Year and Industry

Industry	**Year**					
	2017	**2018**	**2019**	**2020**	**2021**	**2022**
Accommodation & food services	53.4	52.6	55.2	55.2	55.1	54.0
Administrative & support services	44.1	44.0	46.2	46.2	44.3	43.7
Agriculture, forestry & fishing	30.0	27.7	25.9	26.4	27.1	28.0
Construction	13.1	12.5	12.3	13.3	14.5	14.8
Education	72.0	72.5	71.5	72.8	69.9	71.7
Financial & insurance activities	42.7	44.3	43.0	42.8	42.7	43.2
Human health & social work activities	78.3	78.4	77.5	75.9	76.5	75.6
Information & communication	29.4	27.1	29.8	30.9	31.1	30.3
Manufacturing	24.9	24.7	25.1	27.1	27.2	26.2
Mining, energy and water supply	22.2	21.2	24.2	22.5	24.3	26.1
Other services	53.7	54.7	53.4	54.1	55.2	53.0
Professional, scientific & technical activities	43.4	44.5	44.0	43.0	44.3	43.9
Public admin & defence; social security	52.6	50.8	52.7	52.6	51.3	52.1
Real estate activities	50.6	55.7	55.4	52.9	50.9	50.5
Transport & storage	21.1	19.9	18.9	19.8	24.0	23.4
Wholesale, retail & repair of motor vehicles	46.5	46.1	46.9	46.5	45.6	47.0

Note: The Year mentioned is sourced from the *EMP:13* dataset and, in our study, aligns with the turn of the years, as depicted in other tables.

Table B.2: Reg. Results: Gender Dominance and Quartile Effects - mean

	Dependent variable:					
	DiffMeanHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
as.factor(occupation)MALE	-1.749^{***}	-1.658^{***}	-2.413^{***}	-2.783^{***}	-3.092^{***}	-2.475^{***}
	(0.572)	(0.596)	(0.595)	(0.598)	(0.662)	(0.635)
as.factor(occupation)NEUTRAL	-0.974^{**}	-0.829^{**}	-0.867^{**}	-1.552^{***}	-0.972^{**}	-1.131^{***}
	(0.392)	(0.398)	(0.414)	(0.404)	(0.449)	(0.409)
FemaleLowerQuartile	0.595^{***}	0.573^{***}	0.563^{***}	0.562^{***}	0.549^{***}	0.539^{***}
	(0.012)	(0.014)	(0.013)	(0.014)	(0.016)	(0.015)
FemaleTopQuartile	-0.573^{***}	-0.550^{***}	-0.551^{***}	-0.558^{***}	-0.545^{***}	-0.521^{***}
	(0.012)	(0.013)	(0.013)	(0.013)	(0.015)	(0.015)
Public Sector	-3.254^{***}	-2.946^{***}	-3.236^{***}	-3.054^{***}	-2.977^{***}	-2.908^{***}
	(0.337)	(0.323)	(0.317)	(0.327)	(0.351)	(0.328)
1000 to 4999	0.216	0.197	0.305	0.140	−0.338	0.905^{**}
	(0.366)	(0.361)	(0.347)	(0.376)	(0.448)	(0.375)
20,000 or more	−1.097	-2.742^{**}	-2.948^{***}	-2.336^{**}	-2.102^{**}	-1.809^{*}
	(1.275)	(1.330)	(0.976)	(0.931)	(1.001)	(1.042)
500 to 999	0.205	0.167	0.247	−0.172	0.446	0.640
	(0.357)	(0.367)	(0.376)	(0.383)	(0.402)	(0.397)
5000 to 19,999	1.300^{**}	0.872^{*}	0.800	0.740	1.198^{**}	1.968^{***}
	(0.547)	(0.528)	(0.534)	(0.548)	(0.550)	(0.570)
Less than 250	−1.276	−1.958	−1.465	-3.026^{**}	−2.018	−1.070
	(1.502)	(1.352)	(1.535)	(1.219)	(1.308)	(1.168)
Not Provided	-1.839^{**}	-2.756^{*}	−1.099	0.187	1.594	-5.441^{***}
	(0.789)	(1.503)	(1.861)	(1.006)	(1.189)	(0.649)
Constant	6.238^{***}	6.470^{***}	7.038^{***}	7.500^{***}	7.067^{***}	5.699^{***}
	(0.802)	(0.839)	(0.844)	(0.901)	(0.980)	(0.964)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.479	0.447	0.441	0.448	0.375	0.384
Adjusted R^2	0.478	0.446	0.440	0.447	0.374	0.383
Residual Std. Error (df = 4896)	9.658	9.892	9.835	10.150	11.350	10.390
F Statistic (df = 11; 4896)	409.464^{***}	359.544^{***}	351.652^{***}	360.881^{***}	267.316^{***}	278.032^{***}

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

Table B.3: Reg. Results: Gender Dominance and Quartile Effects - median

	Dependent variable:					
	DiffMedianHourlyPercent					
	2017/18	2018/19	2019/20	2020/21	2021/22	2022/23
	(1)	(2)	(3)	(4)	(5)	(6)
as.factor(occupation)MALE	−5.567***	−5.954***	−6.737***	−6.443***	−6.411***	−7.050***
	(0.637)	(0.630)	(0.645)	(0.691)	(0.659)	(0.621)
as.factor(occupation)NEUTRAL	−6.298***	−6.465***	−6.667***	−6.406***	−6.116***	−6.608***
	(0.494)	(0.472)	(0.510)	(0.525)	(0.522)	(0.496)
FemaleLowerQuartile	0.693***	0.688***	0.659***	0.647***	0.639***	0.621***
	(0.014)	(0.015)	(0.015)	(0.019)	(0.018)	(0.017)
FemaleTopQuartile	−0.705***	−0.697***	−0.676***	−0.660***	−0.639***	−0.634***
	(0.015)	(0.015)	(0.015)	(0.018)	(0.016)	(0.016)
Public Sector	−2.351***	−2.368***	−2.918***	−3.131***	−3.253***	−2.863***
	(0.408)	(0.393)	(0.382)	(0.402)	(0.418)	(0.402)
1000 to 4999	−0.211	0.258	0.725*	0.803*	0.361	0.824**
	(0.385)	(0.379)	(0.382)	(0.410)	(0.416)	(0.400)
20,000 or more	−2.458**	−2.829**	−2.167*	−1.895*	−1.891	−2.808**
	(1.003)	(1.239)	(1.143)	(0.977)	(1.237)	(1.346)
500 to 999	0.409	0.867**	0.749*	0.475	0.706	0.414
	(0.405)	(0.407)	(0.420)	(0.460)	(0.436)	(0.438)
5000 to 19,999	−0.887*	−0.903*	−1.025**	−0.316	−0.114	−0.563
	(0.492)	(0.512)	(0.476)	(0.509)	(0.567)	(0.496)
Less than 250	2.658	−2.141	−2.161	−1.113	−1.880*	−1.537
	(1.815)	(1.567)	(1.332)	(1.453)	(1.067)	(1.050)
Not Provided	−0.828	−6.354	−1.169	2.377*	3.325**	−2.665***
	(1.019)	(4.144)	(1.784)	(1.431)	(1.474)	(0.578)
Constant	7.639***	7.733***	8.869***	8.516***	7.921***	9.032***
	(0.843)	(0.828)	(0.884)	(0.976)	(0.955)	(0.910)
Observations	4,908	4,908	4,908	4,908	4,908	4,908
R^2	0.526	0.520	0.494	0.459	0.460	0.440
Adjusted R^2	0.525	0.519	0.493	0.458	0.459	0.439
Residual Std. Error (df = 4896)	10.541	10.556	10.658	11.636	11.271	11.079
F Statistic (df = 11; 4896)	493.652***	482.383***	434.883***	377.954***	378.945***	349.691***

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