

Educational Segregation and the Gender Wage Gap

The Direct and Indirect Contribution of Field of Study Among Tertiary-Educated Workers in Belgium

Author: Zoé Grajek

Thesis Director: Vincent Vandenberghe

Thesis Joint Director: Anna Bernard

Academic Year 2025-2026

In order to obtain the Joint Degree

Master 120 en Sciences économiques, Orientation générale, Finalité spécialisée (UCL/UNamur)

and

Mestrado em Economia (UCP)

Educational Segregation and the Gender Wage Gap: The Direct and Indirect Contribution of Field of Study Among Tertiary-Educated Workers in Belgium

Zoé Lucie T Grajek

September 2026

Supervisors: Professors Anna Bernard & Vincent Vandenberghe

Abstract

Belgium ranks among the European countries with the smallest gender wage gaps, yet a puzzle remains: previous evidence suggests that men and women working in the same firms and jobs are generally paid in line with their productivity. If pay-setting itself were the only channel at play, this would leave little room for an aggregate wage gap. This thesis argues that part of the answer lies in sorting: men and women tend to work in different jobs, and it examines how much of this sorting can be attributed to differences in fields of study. Using the *Vacature SalarisEnquête* (2008–2012) and the Gelbach (2016) decomposition, the analysis estimates the direct contribution of field of study to the gender wage gap among tertiary-educated workers in Belgium, as well as its indirect contribution through occupation, sector and working time. The raw gender wage gap of 20.1% falls to 8.3% after accounting for demographic characteristics, working time, occupation, sector, and field of study, leaving a residual that none of these factors explain further. Among the reduction achieved, field of study accounts for 25.2% of the explained reduction on its own, rising to 40.0% once its indirect effects through occupation and sector are included, close to working time (41.9%). These findings suggest that what women and men study is related to where they work, and ultimately, how much they earn. This suggests that part of the gender wage gap may take shape long before entering the labour market.

Keywords: gender wage gap; field of study; educational segregation; occupational sorting; Gelbach decomposition; Belgium

Acknowledgments

Looking back on the past year, I could not have imagined how much this thesis would ask of me – not only academically, but also emotionally. There were moments of real doubt, evenings when progress felt impossibly slow, and others when everything suddenly fell into place. Writing this thesis taught me as much about patience and perseverance as it did about the gender wage gap.

I would first like to thank my supervisor at UCLouvain, Vincent Vandenberghe, for his support from the very beginning of this project. He introduced me to the topic, the *Vacature SalarisEnquête* dataset and the Gelbach method early on, which allowed me to start working on the project almost right away. Our brainstorming meetings throughout the year were also helpful in shaping my ideas and moving the project forward. His guidance throughout made this project feel possible from the start.

I am equally grateful to my UCP supervisor, Anna Bernard, for her valuable and constructive feedback throughout this work.

My thanks also go to the Economic School of Louvain (ESL), where I completed my second year of master's studies and developed my interest in labour economics. I am equally grateful to Universidade Católica Portuguesa (UCP) in Lisbon, where I completed my first year of master's studies, which gave me strong academic rigour that has served me well throughout this thesis.

To my family, thank you for your unconditional support through every high and low of this process, and for never doubting that I would get here, even on the days when I doubted it myself. To Kot Erasmus, my student association this year, thank you for the countless evenings working side by side at the library and for making this process feel less lonely.

And finally, to Manon Delsemme, my closest friend, thank you for the many hours we spent working together, and for the many times you reminded me I was making more progress than I thought.

Any remaining errors in this thesis are, of course, entirely my own.

Use of Artificial Intelligence

Artificial Intelligence (AI) tools were used as supporting tools during the preparation and writing of this thesis. Their use was limited to specific aspects of the writing and research process and remained under my supervision. No text was directly copy-pasted from AI-generated outputs; any suggestions provided by these tools were reviewed, adapted, and integrated where appropriate.

ChatGPT (OpenAI) was mainly used to improve English phrasing and grammar, as English is not my native language. Claude (Anthropic) was used to help organise ideas and improve the structure of certain sections, as well as to assist with coding in RStudio to obtain the desired outputs. Consensus was used as a tool to identify potentially relevant academic literature and support the construction of the literature review.

All information and sources suggested by AI tools were independently reviewed and verified. Therefore, I remain fully responsible for the content, arguments, and results presented in this thesis.

Table of Contents

1.	Introduction	1
2.	Literature Review	3
2.1.	Strand 1: Labour-Market Determinants	5
2.1.1.	Gender Wage Gap Definition.....	5
2.1.2.	Evolution and Cross-Country Patterns	6
2.1.3.	Tertiary-Educated Workers and the Belgian Case	6
2.1.4.	Working Time and Human Capital	7
2.1.5.	Norms and Discrimination	8
2.1.6.	Occupational Sorting.....	10
2.2.	Strand 2: Pre-Labour-Market Determinants.....	10
2.2.1.	Field of Study and later Wage Inequality.....	11
2.2.2.	Sorting versus Devaluation, applied to field of study	11
2.2.3.	Gendered Norms and the choice of Field of Study	12
2.2.4.	The Belgian Gap & the Contribution of this Thesis.....	12
3.	Conceptual Framework and Empirical Strategy.....	14
3.1.	Conceptual Framework.....	14
3.2.	Research and Sub-Research Questions.....	14
3.3.	Empirical Strategy	15
3.3.1.	The Gelbach Decomposition	16
3.3.2.	Gelbach Relative to Alternative Approaches	18
3.4.	Extending the Strategy: Direct and Indirect Contribution	20
3.5.	Statistical Inference: Bootstraps.....	21
4.	Data description.....	22
4.1.	Description of the data	22
4.2.	Main variables and indicators	23
4.3.	Missing values	25
5.	Descriptive Statistics	26
5.1.	Sample and Key Control Variables	27
5.2.	Three Key Patterns.....	28
5.2.1.	The raw Gender Wage Gap.....	29
5.2.2.	Representativeness (ρ): Gender Segregation across Fields of Study	31
5.2.3.	Wage Premium (Γ) across Fields of Study.....	33
5.3.	Field of Study and the Residual Gap	35
6.	Main Results.....	36
6.1.	Model and Sample	37
6.2.	Overview: Aggregate Contributions	38

6.3.	Field of Study.....	41
6.4.	Occupation.....	45
6.5.	Sector	45
6.6.	Cross-Dimension Synthesis	46
6.7.	Robustness Check	48
7.	Limitations	49
7.1.	Data quality limitations already addressed	49
7.2.	Fundamental Limitations	50
8.	Discussion	53
9.	Next Steps	56
10.	Conclusion.....	58
11.	Bibliography.....	59
12.	Appendices	68

List of Tables

Text

Table 1: Variables and indicators used in the wage equation	23
Table 2: Gender differences in degree level, working time and experience	27
Table 3: The gender wage gap, overall, by year, and by degree	29
Table 4: Correlations between field feminization, mean wage, and within-field gap	35
Table 5: Raw and residual gender wage gap, baseline vs. full model.....	38
Table 6a: Raw, dimension-by-dimension contribution to the gender wage gap	39
Table 6b: Final grouped contribution to the gender wage gap, net of field of study	39
Table 7: Bootstrap estimates, standard errors, and 95% confidence intervals	41
Table 8: Direct contribution of each field of study to the gender wage gap	42
Table 9: Correlation between gender representation and wage premium, by dimension	44
Table 10: Direct and indirect contribution of working time, field of study, occupation, and sector to the gender wage gap	46
Table 11: Robustness to the 2012 compositional skew.....	48
Table 12: Robustness to the exclusion of the public sector	49

Appendices

Table A1: Top 10 female- and male-dominated occupations	68
Table A2: Sample size by field of study, broken down by Bachelor and Master degrees.....	68
Table A3: Sample construction and missing data	70
Table A4: Missing values on field of study (dipl), by gender.....	70
Table A5: Missing values on occupation (occh), by gender	70
Table A6: Top 10 most feminized and most masculinized fields	71
Table A7: Wage comparison across feminized and masculinized fields, by matched degree level	72
Table A8: Dissimilarity index by field of study, occupation, and sector.....	73
Table A9: Robustness check — age vs. experience as control for career stage	73
Table A10: Missing values by variable, tertiary sample (2008–2012)	73
Table A11: Prevalence of a second recorded diploma and consistency with the field-of-study variable.....	74

Table A12a: Omitted (reference) category for each categorical control.....	74
Table A12b: Invariance check — releveling the sector reference category	74
Table A13: Fields most extreme in gender representation or wage premium (values centered on the sample average).....	75
Table A14: Direct contribution of each occupation to the gender wage gap	76
Table A15: Direct and Indirect contribution of each sector to the gender wage gap.....	79

List of Figures

Text

Figure 1: Gender wage gap, from raw to residual.....	30
Figure 2: Share of women by field of study	32
Figure 3: Mean FTE-equivalent monthly wage by field of study	34

Appendices

Figure A1: Distribution of gross monthly wages by gender and degree level	71
Figure A2: Gender representation (ρ_k) versus wage premium (γ_k), by field of study	75
Figure A3: Gender representation (ρ_k) versus wage premium (γ_k), by occupation	78

1. Introduction

Two students start university side by side, equally curious and equally ambitious. Ten years later, one earns considerably more than the other. The difference may have less to do with how hard they worked than with a decision they made at the very beginning of their student life: what they chose to study.

Women have made remarkable progress in higher education over the past decades, and in many countries, they now graduate from university at higher rates than men (Bar-Haim et al., 2018; OECD, 2021). Yet this progress has not translated into equal pay. In 2023, women in the EU earned, on average, 12% less per hour than men (European Commission, 2024). Despite decades of policy efforts, this gap remains. Part of the explanation may lie not only in what happens in the workplace, but also in choices made much earlier: the degree students choose, the occupations it leads them to, and the sectors in which they eventually work.

Belgium offers an interesting case. By European standards, it has one of the smallest gender wage gaps, with an unadjusted gap of 9.1% (Boll & Lagemann, 2019; De Poli & Maier, 2024). Yet Vandenberghe (2011) and Martins and Vandenberghe (2010), using matched employer-employee data, find that men and women in the same job are paid broadly in line with their productivity. If the aggregate gap does not mainly result from women being paid less for the same work, where does it come from? This thesis argues that part of the answer lies in where men and women end up working, and that these differences may start before individuals enter the labour market.

A large body of research, as discussed in the literature review, has examined how occupation, sector, and working time contribute to the gender wage gap once individuals enter the labour market (Bardasi & Gornick, 2008; Blau & Kahn, 2017; Block, 2023). More recent research has also looked further back at the fields of study men and women choose before entering the labour market – such as Engineering vs Nursing or Economics vs Psychology – and how these choices are linked to later wages (Rozhkova et al., 2024; Sloane et al., 2021). However, no Belgian study has directly estimated how much of the gender wage gap can be linked to these educational choices, or separated the direct contribution of field of study from its indirect contribution through occupational and sectoral sorting. This is precisely the gap this thesis addresses: **to what extent does educational segregation among tertiary-educated workers**

contribute directly and indirectly to the gender wage gap in Belgium between 2008 and 2012?

To answer this question, this thesis uses the *Vacature SalarisEnquête*, combining its 2008, 2010, and 2012 waves into a sample of over 100,000 tertiary-educated respondents. The analysis uses Gelbach (2016) decomposition, which provides an exact attribution of the gender wage gap to different observable factors without depending on the order in which correlated variables are introduced. The framework is extended to capture both the direct contribution of field of study on wages – that is, the wage premium or penalty associated with a given field – and its indirect contribution through working time, and occupational and sectoral sorting. The latter effect reflects the extent to which differences in occupations and sectors are partly linked to fields of study that men and women chose earlier in their education.

The results tell a clear story. The raw gender wage gap among tertiary-educated workers stands at 20.1%, falling to 8.3% once demographic characteristics, working time, occupation, sector, and field of study are jointly accounted for. Field of study alone accounts for 25.2% of this reduction, compared with 48.4% for working time. However, when its indirect contribution through occupation, sector and working time is included, its total contribution rises to 40.0%, close to working time (41.9%) and well above occupation (13.9%) and sector (4.1%). These results suggest that field of study is an important contributor to both occupational and sectoral sorting.

These findings have an important policy implication: if part of the gender wage gap is already shaped by field-of-study choices before labour-market entry, labour-market policies alone, such as pay transparency and anti-discrimination measures, can only address part of the problem. This leaves room for earlier intervention, such as exposure to role models or targeted encouragement toward specific fields of study (Li, 2018; Porter & Serra, 2020). This implication is further discussed in Section 8.

This thesis is organised as follows. Section 2 reviews the literature on labour-market and pre-labour-market determinants of the gender wage gap. Section 3 presents the conceptual framework and Gelbach decomposition used for the data analysis. Section 4 describes the data, while Section 5 presents the descriptive statistics. Section 6 reports the main results of the decomposition and a series of robustness checks. Section 7 discusses the limitations, and Section 8 concludes by considering what these findings may imply for addressing gender differences in educational choices. Finally, Section 9 outlines potential next steps and avenues for further research.

2. Literature Review

Belgium is consistently identified as a country with relatively small conditional gender wage gaps in international comparisons (Section 2.1.3). These estimates – wage differences between men and women after controlling for observable characteristics or endowments – are often interpreted as reflecting potential wage discrimination. In this respect, Belgium performs relatively well. Evidence from the Belgian private sector reinforces this conclusion. Using matched employer–employee data, Martins and Vandenberghe (2010) and Vandenberghe (2011) find little evidence of gender wage discrimination within firms: gender differences in wages are no larger than corresponding differences in productivity. In other words, women appear to be paid broadly in line with their relative productivity. However, a small conditional or unexplained wage gap does not imply the absence of a raw gender wage gap. Men and women may receive similar wages for comparable work while still earning substantially different amounts overall because they are unevenly distributed across firms, sectors, occupations, hierarchical positions, and working-time arrangements. This distinction also highlights a limitation of the conventional gender-wage-gap literature. By focusing on wage differences conditional on observed characteristics, it concentrates mainly on mechanisms operating within the labour market, once education level, occupation, sector, and possibly also working time have been determined. Yet part of the raw gender wage gap may originate before labour-market entry and deserves closer investigation. As stated in the introduction, educational choices are particularly relevant. Although gender differences in overall educational attainment have narrowed or reversed, women and men continue to make markedly different choices regarding field of study i.e. the pre-labour-market mechanism at the core of this thesis. These choices, reflecting differences in preferences and occupational aspirations, can shape subsequent sorting across occupations, sectors, and working-time arrangements, and hence later earnings. For anyone interested in the gender earnings gap, the relevant question is therefore broader than “Are equally productive women and men paid differently for comparable work?”. It is also “Why do women and men enter the labour market with different characteristics and subsequently sort into different jobs and occupations?” Understanding the raw gender wage gap requires considering both pre-labour-market mechanisms – here, particularly field-of-study choices – and the labour-market mechanisms traditionally emphasised in the gender wage discrimination literature.

The literature provides two broad explanations. The first view this unequal distribution – commonly referred to as occupational segregation and described by some as a “primary driver

of the gender wage gap” (Block, 2023) – as a **labour-market** phenomenon. It focuses on factors that operate once individuals have entered employment, such as hours worked, human capital accumulated through work, employer discrimination, and the sorting of workers across occupations and sectors (Borrowman & Klasen, 2020; Buligescu et al., 2020). The second, more recent strand focuses on **pre-labour-market** factors that influence individuals before they enter the labour market. In particular, it considers the role of fields of study in higher education, which can shape the types of occupations graduates enter later on (Francesconi & Parey, 2018; Galos & Kulic, 2022; Petrenko et al., 2026). Section 2.1. will review the first strand, while Section 2.2. will develop the second, which is the main focus of this thesis.

One important distinction needs to be made before reviewing this literature, as it will be relevant throughout the section. When men and women are distributed differently across a category – such as field of study, occupation, or sector – this can affect the gender wage gap through two different mechanisms. The first is a **sorting effect**, i.e. when men and women are concentrated in categories that have different wage levels (Blau & Kahn, 2017; Sloane et al., 2021). The second is a **devaluation effect**: the wage associated with a category may decrease as it becomes more feminized (Busch, 2018; Leuze & Strauss, 2016). The empirical strategy used in this thesis, presented in Section 3.3., focuses on the first mechanism. It estimates how much of the gender wage gap can be attributed to the different representation of men and women across fields of study, occupation, and sectors, while keeping the wage returns to these categories constant.

This choice has two separate implications that should not be confused. First, the analysis does not examine whether men and women receive different wages within the same category. In other words, it does not test whether men and women are paid differently when they have the same field of study, occupation, or sector. This assumption is discussed in more detail in Section 3.2.2. However, preliminary descriptive evidence on gender wage differences across fields of study will be presented in the Descriptive Statistics (Section 5). Second, the analysis does not directly test the devaluation mechanism mentioned above. Testing this mechanism would require examining whether the wage associated with a given category decreases as the proportion of women in that category increases. This would need a different empirical approach, such as tracking changes in a category’s wage level as its gender composition changes over time. Since the data used in this thesis cover a 4 years-period (2008-2012), they do not allow for such a test. This is therefore acknowledged as a limitation in Section 7.

2.1. Strand 1: Labour-Market Determinants

This first strand of the literature analyses the gender wage gap primarily as a labour-market phenomenon, shaped by factors that operate once individuals are already in employment – from working time and human capital to occupational sorting and discrimination. Before discussing these factors, it is useful to briefly explain how the gender wage gap is defined and measured.

2.1.1. Gender Wage Gap Definition

The gender wage gap is usually defined as the difference in average wages between men and women, expressed as a percentage of men’s wages. This basic – or **unadjusted** – gap compares earnings of all male and all female workers without controlling for any characteristics (Harman, 2023; Kolesnikova, 2011), and is typically computed as:

$$U_{GPG} = \frac{\text{Average hourly wage of men} - \text{Average hourly wage of women}}{\text{Average hourly wage of men}} * 100 \quad [1]$$

or, equivalently, estimated econometrically as:

$$\ln(W_t) = \beta_0 + \beta_1 \text{Women}_i + \varepsilon_i \quad [2]$$

where: W_t is the average hourly wage, β_1 is the estimated coefficient representing the unadjusted gender pay gap, Women_i is a dummy variable equal to 1 for women and 0 for men, and ε_i is the error term (Harman, 2023, building on Blau & Kahn, 2003).

Because this raw indicator mixes pure gender differences with the effect of observable characteristics, such as hours worked e.g., it is an imperfect measure of “unequal pay for equal work” (Harman, 2023; Strawinski et al., 2018). Many studies therefore also compute an **adjusted** gender wage gap, which holds these observable characteristics – productivity and labour supply – constant through regressions or decomposition methods. This is more often interpreted as capturing unaccounted wage differences that may point at wage discrimination, rather than pure wage discrimination itself (Edelsztein & Waisbrot, 2023; Goldin, 2014). Several theoretical models have been developed to explain and compute this discrimination component, including taste-based and statistical discrimination models (Baraku & Busetta, 2024; Becker, 1971), discussed further below.

In practice, both measures serve different purposes: the unadjusted gap summarises overall inequality and allows simple cross-country comparisons, while the adjusted gap approximates how far labour markets are from equal pay for the equal work. The specific decomposition adopted in this thesis is presented in Section 3.1.1.

2.1.2. Evolution and Cross-Country Patterns

Over the past decades, the gender wage gap has narrowed across most developed countries, although at different rates. It remains considerably larger in the US and UK (around 30%, unadjusted) than in continental Europe (10-15%), while Belgium stands out within the EU with one of the smallest unadjusted wage gaps (Boll & Lagemann, 2019; Olivetti & Petrongolo, 2016). However, these overall patterns can vary substantially across groups of workers within the same country. The next section therefore focuses on explaining the choice to consider tertiary-educated workers in Belgium.

2.1.3. Tertiary-Educated Workers and the Belgian Case

❖ Tertiary-Educated workers

Although women now attain tertiary education at similar or higher rates than men in many countries (Bar-Haim et al., 2018; OECD, 2021), a substantial gender wage gap remains among graduates. The gap can appear shortly after entering the labour market and often increases with work experience (Francesconi & Parey, 2018; Rozhkova et al., 2024; Zając et al., 2024). For example, in Germany, a wage gap of 5-10 log points appears within 12-18 months of graduation, despite similar working hours. In this specific case, field of study is identified as the most important factor explaining this early gap (Francesconi & Parey, 2018). Similar patterns have been found in Poland (Zając et al., 2024), Russia (Rozhkova et al., 2024), and North America (Fan, 2019), suggesting that field of study and occupational segregation¹ are important factors, alongside differences in educational attainment.

The gender wage gap has also been found to increase among higher earners in several countries, a pattern often referred to as the “**glass ceiling**” effect (Bonaccolto-Töpfer et al., 2023; Christofides et al., 2013). This thesis examines whether a similar pattern exists in Belgium, where the gap seems to be slightly larger among Master’s graduates than among Bachelor’s graduates (Section 5.2.).

❖ The Belgian Case

As mentioned above, Belgium is consistently among the European countries with the smallest gender wage gaps, although estimates vary depending on the data sources and years. Recent studies report an unadjusted gap of around 9.1% and an adjusted gap below 8% (De Poli & Maier, 2024), while other estimates place the gap below 5% depending on the measures used (Boll & Lagemann, 2019; Christofides et al., 2013). Despite this relatively small overall gap,

¹ **Occupational segregation** refers to the uneven distribution of men and women across occupations, with women concentrated in certain jobs and men in others.

meaningful wage differences remain. As discussed above, however, the sources of this gap are still not well documented for tertiary-educated workers specifically. This justifies why this thesis focuses on this population and on the Belgian context.

2.1.4. Working Time and Human Capital

Beyond the wage gaps documented among tertiary-educated workers and across the Belgian wage distribution, two further labour-market factors are consistently identified in the literature as key contributors of the gender wage gap: differences in working time, and differences in the accumulation of human capital over the course of a career.

❖ Working time and career interruption

Globally, women work fewer paid hours on average and are more likely than men to work part-time. Moreover, part-time work is associated with lower wages, even after accounting for observable characteristics (Bardasi & Gornick, 2008; Matteazzi et al., 2018). This pattern is particularly strong in Belgium. In 2011, part-time employment accounted for 18.8% of total employment, 2.2 percentage points above the OECD average. This was mainly driven by women, whose part-time employment rate was 25.4 percentage points higher than men's (Garnero et al., 2014). Belgian data also show that the lower wages of part-time workers cannot be fully explained by differences in education, occupation, or sector. The remaining difference is often interpreted as a possible part-time wage penalty² (Jepsen et al., 2005). Since women are much more likely to work part-time, largely because of the unequal distribution of domestic responsibilities, they are disproportionately affected. Therefore, they face both the disadvantage of working fewer hours and, in many cases, a lower hourly wage due to this part-time wage penalty (Deschacht, 2017; Garnero et al., 2014).

Beyond working time, parenthood is another contributor to the gender wage gap, mainly through career interruptions and reduced labour supply following the birth of a first child (Gangl & Ziefle, 2009; Kleven, Landais & Leite-Mariante, 2023). In Belgium, mothers earn 32% less than fathers up to eight years after their first child, with larger penalties in sectors with more atypical working schedules (Fontenay, Murphy & Tojerow, 2021).

² **Part-time wage penalty** refers to the empirical finding that part-time workers earn lower **hourly** wages than comparable full-time workers, even if they are equally productive at the same job.

❖ Human Capital (HK) Accumulation

Human Capital has traditionally been measured in the gender wage gap literature using years of schooling (Oaxaca, 1973; discussed further in Section 3.3.1.). However, this measure has become less informative as women have reached, and in many countries surpassed, men in educational attainment, while the gender wage gap persists (Bar-Haim et al., 2018).

In Belgium, shorter job tenure³ among women is one of the factors that most consistently contributes to the remaining wage gap, according to the literature. This is consistent with career interruptions reducing women's opportunities to accumulate firm-specific experience (Tverdostup & Paas, 2022). However, even when women have similar or higher educational qualifications than men, they do not always receive the same returns from their Human Capital, *ceteris paribus*. Using Belgian matched employer-employee data, Kampelmann et al. (2018) found that highly educated women are systematically “underpaid” relative to their measured productivity. Education attainment alone, therefore, does not eliminate the gender wage gap. This motivates looking beyond the level of education and considering a more qualitative dimension of the pre-labour-market: field of study, which is introduced and examined in Section 2.2. of this literature review.

2.1.5. Norms and Discrimination

The labour-market factors discussed so far describe *what* men and women bring to the labour market in terms of measurable, observable characteristics. Two further mechanisms help explain the gender wage gap. First, social norms and preferences shift the focus further upstream, toward *why* men and women make different choices and develop different characteristics in the first place. This is directly relevant to this thesis, as similar factors may also influence the choice of field of study examined in Section 2.2. Then discrimination, by contrast, provides a different explanation: part of the wage gap may remain even after accounting for observable characteristics, reflecting differences in how employers treat women and men with similar characteristics.

❖ Social Norms and Preferences

Gender norms, particularly those related to family roles, are closely linked to wage inequality. In the US, for example, it has been shown that labour market areas with more egalitarian gender norms tend to have smaller gender wage gaps, largely because men and women in the same occupation receive more similar wages (Scarborough & Moeder, 2022). Gender stereotypes can also influence the types of occupations women and men choose. The literature shows that

³ **Job tenure** refers to the number of years an individual has been employed by their current employer.

women are more likely to enter “communal” occupations, such as care and support jobs, which generally pay less, than “agentic” occupations, such as leadership and achievement-oriented jobs (Arceo-Gómez et al., 2022). This pattern is discussed further in the section on occupational sorting below (Section 2.1.6.). In Belgium, gender norms and stereotypes are estimated to explain a substantial share of the gender wage gap, around 40%, while motherhood- and family-related norms are found to account for much of the remaining difference (Ciminelli et al., 2021; Scarborough & Moeder, 2022).

Crucially, these norms may influence individuals before they enter the labour market. Research suggests that young people can anticipate their future family roles well before entering employment, and that these expectations may shape their choices. Cech (2016) finds that women’s expectations about future family responsibilities – not only their actual family circumstances – are associated with choosing occupations that are more compatible with anticipated caregiving responsibilities, even among women who do not yet have children. Similarly, a longitudinal study on students planning to pursue careers in law shows that men and women already have different expectations about combining work and family at the beginning of their careers. Women are more likely to anticipate that family responsibilities will play a central role in shaping their future careers (Warner & Clair, 2026).

If these expectations already influence career and occupational choices before labour-market entry, they may also, unconsciously, affect an even earlier decision: the choice of field of study, made several years before individuals enter an occupation or start a family. This possibility, largely absent from the labour-market-focused literature reviewed so far, is examined explicitly in Section 2.2.

❖ Discrimination

It is important to note that an unexplained wage gap often remains after controlling for education, experience, occupation, sector and working hours. Some studies interpret this residual as evidence of discrimination, either **taste-based**, when employers have a preference against employing women (Becker, 1971), or **statistical**, when employers use a gender proxy for expected productivity under limited information (Baraku, 2024; Barron et al., 2024). However, an unexplained gap may also result from unobserved characteristics, measurement errors, or model specification rather than discrimination itself (Blau & Kahn, 2017; Goldin, 2014). Distinguishing discrimination from these alternative explanations requires a specific empirical strategy, which is beyond the scope of this thesis. The focus here therefore remains on the observable, pre-labour-market factor of field of study, as well as some labour-market factors.

Evidence for Belgium is mixed. As mentioned earlier, matched employer-employee data found no detectable within-firm discrimination in the private sector overall (Martins & Vandenberghe, 2010; Vandenberghe, 2011), which is consistent with this thesis's focus on pre-labour-market determinants, examined in Strand 2 below. However, evidence from audit firms⁴ suggests that discrimination may still occur in specific segments, where women earn significantly less than men despite generating higher fee premiums (Hardies, Lennox & Li, 2020).

2.1.6. Occupational Sorting

This leads to analysing the concept of occupational segregation, i.e. the uneven distribution of men and women across occupations. It has been described as “a primary driver of the gender wage gap” (Block, 2023), as occupations with more women tend to pay less in general (Borrowman & Klasen, 2020). Cross-country and US evidence attributes around 1/3 of the overall gap to occupation alone, rising to 40-50% of women's earnings losses for some groups in the US (Alonso-Villar & Del Río, 2023; Buligescu et al., 2020), although these estimates vary across countries and wage levels (Brynin & Perales, 2016; Christofides et al., 2013).

The extent of occupational sorting is substantial: some occupations are almost entirely female or male (Table A1, Appendix 1; Ruijter et al., 2003), and moving from an entirely male to an entirely female occupation is associated with earnings losses of up to 13% (Murphy & Oesch, 2016). The dissimilarity index, measuring the share of workers who would need to change occupation to achieve gender balance, was around 0.30 in the EU in 2010 (Mavrikiou, 2020).

Belgium provides an interesting case: despite relatively high occupational segregation, it has one of the smallest gender pay gaps in Europe, suggesting that the contribution of segregation to wage inequality depends not only on where men and women work, but also on how these occupations are valued (Mavrikiou, 2020).

Therefore, occupational segregation is both an outcome of gender differences in employment and one of the way in which these differences are reflected in wages. Its origins may also lie before labour-market entry, particularly in gendered choices of fields of study. This leads to the next section on field-of-study segregation.

2.2. Strand 2: Pre-Labour-Market Determinants

As the previous section suggested, occupational segregation may partly originate prior to labour-market entry. Its roots may lie earlier in individuals' educational choices, particularly

⁴ **Audit firms** refer to companies that provide independent auditing and assurance services, primarily by examining the financial statements of other organizations.

their choice of field of study in higher education. This second, more limited strand of the literature examines this pre-labour-market channel and forms the central focus of this thesis.

2.2.1. Field of Study and later Wage Inequality

Field of study is a key link between education and later wage inequality. Men and women remain unevenly distributed across fields, which contributes to subsequent occupational segregation and pay inequality (Galos & Kulic, 2022; Petrenko et al., 2026). In general, evidence shows that women tend to concentrate in fields associated with lower earnings and subsequently enter lower-paying occupations, even after accounting for field of study (Sloane et al., 2021). For instance, Aerospace Engineering, one of the highest-earning US majors, is 88% male, while Early Childhood Education, one of the lowest-earning, is 97% female (Sloane et al., 2021). Similar patterns are observed across countries, with women underrepresented in high-earning STEM⁵ fields and overrepresented in arts, humanities, health, and education (Rozhkova et al., 2024; Zajac et al., 2024).

The contribution of field of study can be considerable, although it varies across contexts: among Russian graduates, it explains 32% of starting wage differences between men and women, with educational factors more broadly accounting for 53% of the explained entry gap (Rozhkova et al., 2024). Importantly, its contribution is not only direct. Because fields of study strongly influence occupational sorting, part of its contribution to the gender wage gap operates indirectly through the occupations graduates enter (Petrenko et al., 2026). Reducing educational segregation could therefore be expected to narrow the gender wage gap, although it would not eliminate it entirely (Zheng & Weeden, 2023). Distinguishing between these direct and indirect contributions is the empirical objective of this thesis, developed further in Section 3.4.

2.2.2. Sorting versus Devaluation, applied to field of study

As discussed above, gender segregation across fields of study can affect the gender wage gap through two mechanisms. The first is **sorting**, whereby men and women are concentrated in fields with different wage levels. The second is **devaluation**, implying that the pay associated with a field decreases as its share of women increases (Busch, 2018; Leuze & Strauss, 2016). In the context of field of study, the sorting mechanism suggests that women are more likely to enter fields such as education and care that are already relatively low-paying (Perales, 2013; Sloane et al., 2021). The devaluation mechanism, by contrast, suggests that the increasing share

⁵ **STEM** stands for Science, Technology, Engineering and Mathematics, and refers to academic fields and professions related to these disciplines.

of women in a field may itself reduce its associated pay, linking it to the discrimination mechanisms discussed above (Section 2.1.5.). However, evidence of such devaluation in fields of study remains mixed and appears to depend on the context (Busch, 2018; Leuze & Strauss, 2016). Moreover, as mentioned above, the dataset used in this thesis does not allow for an empirical test of the devaluation mechanism, and the focus therefore shifts to the sorting mechanism.

This distinction is important for interpreting the results of this thesis. The empirical strategy presented in the next section focuses on the **sorting** mechanism: it estimates how much of the gender wage gap is associated with differences in the distribution of men and women across fields of study, taking the wage returns associated with these fields into account. It therefore does not test whether the wage return associated with a given field vary with its degree of feminization, which would correspond to the devaluation mechanism discussed above.

2.2.3. Gendered Norms and the choice of Field of Study

As introduced above (Section 2.1.5.), anticipated family responsibilities can shape career choices well before labour-market entry (Cech, 2016; Warner & Clair, 2026). This influence may extend to the choice of field of study itself. Adolescents' career aspirations strongly influence the fields they later choose, meaning that educational segregation can begin years before entering the labour market (Morgan, 2013), potentially reflecting similar expectations about future family roles.

Field-of-study choices also respond differently to new information by gender, often reinforcing existing patterns. In a German field experiment, for example, information on major-specific earnings encouraged male students to apply to higher-income majors, but had no similar effect on female students (Finger et al., 2020), suggesting that gender differences in field choices are not simply driven by limited information about returns. Family and social expectations also play an important role: in Beijing, parental expectations strongly influenced subject choices linked to university majors, with boys more often directed toward STEM and girls toward arts and humanities, on average (Li, 2025).

2.2.4. The Belgian Gap & the Contribution of this Thesis

Taken together, these findings suggest that field-of-study choices are not simply individual decisions but are shaped by gendered expectations and can have lasting consequences for later labour-market outcomes.

Pre-labour-market choices appear to play an important role in shaping later gender wage inequality, influencing educational pathways, occupational opportunities, and earning potential before individuals enter employment. These choices are themselves partly shaped by gendered norms and expectations, as discussed above. Evidence on this mechanism is available across several countries, including Germany, Finland, Russia, Poland, South Korea, Chile, and Norway (Choi et al., 2023; Galos & Kulic, 2022; Hägglund, 2024; Parada-Contzen & Jara, 2025; Rozhkova et al., 2024; Seehuus & Strømme, 2024; Zajac et al., 2024). However, to the best of my knowledge, no Belgian study has directly estimated the contribution of field of study to the gender wage gap. The closest evidence comes from cross-national research showing a large wage penalty for overeducated women compared with men. Yet, this does not address the role of gendered field-of-study choices in Belgium (Jagannathan & Camasso, 2025). Similarly, Kampelmann et al. (2018) use Belgian matched employer-employee data to show that highly educated women receive lower wage returns than men, consistent with a glass ceiling. Their analysis focuses on the level of education rather than its field, leaving the role of field-of-study segregation in Belgian wage inequality unexplored.

This gap is particularly relevant because field-of-study effects vary across countries, fields, and career stages. Early-career wage gaps differ considerably across fields in Europe, with female wage growth penalties especially pronounced in Economics, Business, and Law (Sánchez-Mangas, 2021), while even within STEM fields, the gap ranges from over 25% in mathematics to below 3% in chemical and Earth sciences in Poland (Zajac et al., 2024). The size and timing of effects identified in other countries cannot simply be assumed to apply to Belgium (Hägglund, 2024; Petrenko et al., 2026; Zajac et al., 2024).

Therefore, the main contribution of this thesis is to provide the first Belgian estimate of the contribution of field of study to the gender wage gap, distinguishing between its direct effect and the indirect effects operating through occupation, sector and working time. It also contributes to the occupational sorting literature reviewed above (Section 2.1.6.) by examining whether occupational and sectoral segregation can partly be traced back to field-of-study sorting, rather than treating them as independent labour-market processes. The decomposition method used for this purpose, presented in Section 3.3.1., is invariant to the order in which covariates⁶ are considered, unlike standard sequential regression approaches.

⁶ In this case, a **covariate** refers to an explanatory variable included in a regression to account for a factor, other than gender, that may also influence wages.

3. Conceptual Framework and Empirical Strategy

3.1. Conceptual Framework

The literature review has established that occupational and sectoral segregation may themselves originate earlier, through gendered choices of field of study. This relationship can operate through two distinct mechanisms: sorting and devaluation (Section 2.2.2.). As mentioned above, this thesis focuses on the sorting mechanism. Before discussing how the mechanisms are measured empirically (Section 3.2), this section sets out the conceptual logic underlying the empirical strategy, without yet specifying its functional form.

The central object of interest is field of study. Its contribution to the gender wage gap can operate **directly** through the wage premium associated with a given field, and **indirectly**, through the occupations and sectors that graduates enter. Two graduates with different fields of study may therefore earn different wages because their fields have different associated wage premiums and because they lead to occupations with different wage levels.

This distinction matters both conceptually and empirically. Among US college graduates, accounting jointly for field of study and occupational sorting explains approximately 60% of the raw gender earnings gap (Sloane et al., 2021). This illustrates how part of what appears to be an occupational effect may originate in earlier educational choices.

To capture this indirect channel, this thesis applies the same logic, to occupation, sector, and working time as to field of study, decomposing each into its contribution to the gender wage gap. However, occupation, sector and working time are not the main object of interest. They are examined to assess how much of their contribution can be traced back to field-of-study segregation rather than from independent labour-market processes. Field of study's overall contribution therefore combines its direct effect with the part of the occupation, sector and working time effect that can be attributed to field of study segregation.

3.2. Research and Sub-Research Questions

Building on this conceptual framework, this thesis addresses the following Research Question:

“To what extent does educational segregation — i.e., gender differences in fields of study — among tertiary-educated workers contribute directly and indirectly to the gender wage gap in Belgium between 2008 and 2012?”

This question brings together the main elements developed throughout the thesis. First, the focus on tertiary-educated workers defines the population of interest. Individuals with similar

levels of education provide a better basis for examining whether wage differences are related to field of study rather than to differences in educational attainment itself (Section 2.1.3.). Second, educational segregation, i.e. the unequal distribution of men and women across fields of study, is the main mechanism of interest, motivated by its role in wage differences through sorting (Section 3.1.) and by the lack of Belgian evidence quantifying its contribution to the gender wage gap (Section 2.2.4.). Finally, as discussed in Section 3.1., field of study may be associated with differences in wages both directly and indirectly, through the occupation and sectors graduates enter. Distinguishing between these channels requires a decomposition method that provides order-independent estimates of each covariate's contribution, unlike sequential regression approaches. This motivates the use of the Gelbach (2016) decomposition, presented below in Section 3.3.

These elements translate into three sub-research questions, examined for the same population and context as the main research question – tertiary-educated workers in Belgium between 2008 and 2012 – and addressed in turn in Sections 5 and 6.2.:

1. “What are the respective contributions of working time, field of study, occupation, and sector to the gender wage gap, and how large is the unexplained residual that remains once these factors are accounted for?”
2. “What is the direct contribution of field of study to the gender wage gap, and which fields act as the main drivers of this contribution?”
3. “How does gender segregation across fields of study (pre-labour-market) compare with segregation across occupations, sectors and working time (labour-market), both in terms of its extent and how much of their contributions to the gender wage gap can be traced back to field-of-study segregation?”

3.3. Empirical Strategy

Having established the conceptual framework, this section presents the empirical strategy used to measure these effects. This thesis implements a simple decomposition of the raw gender wage gap in order to assess how much field of study accounts for this gap, relative to other observable factors – occupation, sector, and working time. This section presents the Gelbach (2016) decomposition, the method used for this purpose, explains why it is particularly well suited to this objective, and situates it relative to alternative decomposition approaches.

3.3.1. The Gelbach Decomposition

This thesis addresses the limitation of the incremental regression method by adopting the decomposition proposed by Gelbach, which provides a unique and order-invariant way of attributing the change in a coefficient of interest – here, the gender wage gap – to a set of covariates.

The method starts from two nested wage regressions⁷. The **baseline model** relates log wages to gender only, where Y represents the logarithm of individual gross monthly wages:

$$Y = \alpha^b + \beta^b \cdot fem + \varepsilon^b \quad [3]$$

Here, β^b is the raw wage gap between women and men. The **full model** adds the covariates of interest, namely field of study, occupation, sector, and working time, jointly denoted Z :

$$Y = \alpha^f + \beta^f \cdot fem + \mathbf{Z}\mathbf{F} + \varepsilon^f \quad [4]$$

Here, β^f is the adjusted wage gap, net of all the covariates $\mathbf{Z}$, while $\mathbf{F}$ represents the Shadow Price of each element of Z – that is, how much each characteristic or control variable (e.g. a given field of study, occupation, or sector) is associated with higher or lower wages, independently of who holds it. The central insight of Gelbach is that the gap between β^b and β^f can be written as an exact function of a single **ancillary regression**, which regresses each covariate in Z on the gender indicator:

$$\mathbf{Z} = \mu + \boldsymbol{\rho} \cdot fem + \boldsymbol{\nu} \quad [5]$$

Here, $\boldsymbol{\rho}$ captures how much women and men differ, on average, in that covariate. For example, it captures how over- or under-represented women are in a given field of study, relative to men. When Z is a categorical variable, represented by a set of dummies – such as field of study, occupation, or sector – $\boldsymbol{\rho} \neq 0$ reflects gender segregation across categories, i.e. the extent to which men and women are unevenly distributed across fields of study, occupations or sector. Using the standard omitted-variable bias result, which shows that omitting a relevant covariate can affect the estimated coefficient of the remaining variable, Gelbach shows that:

$$\hat{\beta}^b = \hat{\beta}^f + \hat{\boldsymbol{\delta}}, \quad \hat{\boldsymbol{\delta}} \equiv \hat{\boldsymbol{\rho}}\hat{\mathbf{F}} \quad [6]$$

This single equation is the core of the method, and it clarifies exactly why a covariate matters for the gap: it contributes to $\boldsymbol{\delta}$ only if two conditions hold jointly. First, women and men must

⁷ Two regressions are **nested** when one includes all the variables of the other, plus additional covariates.

differ in that covariate ($\rho \neq 0$) – a representation effect. Second, that covariate must carry a wage premium ($\Gamma \neq 0$) – a price effect. If either condition fails – women and men are equally represented in a category, or that a category pays the same regardless of who occupies it – the covariate cannot contribute to the gap δ , however important it may seem on its own. Gelbach therefore provides a natural way of linking segregation along each dimension directly to its contribution to the raw gender wage gap.

Consider two hypothetical fields of study. Field A is gender-balanced ($\rho_A \approx 0$) but offers a large wage premium (Γ_A large). Despite its high return, it contributes little to the gender wage gap because men and women benefit from it equally. Conversely, field B is strongly male dominated (ρ_B large in absolute value) but offers no wage premium ($\Gamma_B \approx 0$). It also contributes little to the gap because the field offers no wage advantage. A field therefore contributes meaningfully to the gender wage gap only when both unequal gender representation and non-zero wage premium are present.

When Z is a categorical variable with many categories, such as field of study, this logic extends naturally to each category k : ρ_k measures women's over- or under-representation in field k , and γ_k – a single element of the vector Γ , corresponding specifically to category k – is the wage premium of field k relative to an omitted reference category. Their product:

$$\delta_k = \rho_k * \gamma_k \quad [7]$$

is field k 's direct contribution to the gender wage gap. The contribution of all fields sums exactly to the total contribution of the “field of study” variable, i.e. $\delta = \sum_k \delta_k$.

The sign of δ_k is read directly from ρ_k and γ_k . For instance, a field widens the gender wage gap when women are under-represented in it and it carries a positive wage premium⁸, or when they are over-represented in it and it carries a negative premium. Conversely, a field narrows the gap when women are over-represented in field that pays more than the reference category or under-represented in one that pays less. For instance, a strongly male-dominated field with a large positive premium ($\rho_k < 0$, γ_k large and positive) produces a large negative δ_k , widening the gap. Conversely, a strongly female-dominated field with a smaller, but still positive, premium ($\rho_k > 0$, γ_k modest but positive) produces a positive δ_k , narrowing the gap instead.

⁸ A **positive wage premium** means that individuals in that category earn systematically more than those in the omitted reference category, ceteris paribus.

Applied to field of study more broadly, this framework directly operationalises the sorting mechanism introduced in Section 2.3.1.: ρ_k captures the sorting dimension of segregation – how unevenly women and men are distributed across fields – while γ_k captures the wage premium associated with each field, independently of who holds it. This decomposition does not directly test the devaluation mechanism. Instead, γ_k captures existing differences between fields, regardless of their causes. The framework cannot determine whether female-dominated fields have lower wage premiums because they are feminized. This limit, already mentioned above, is discussed in Section 7.

It is important to mention that this thesis assumes a single wage premium (γ_k) for each field of study, occupation, and sector, common to men and women, rather than estimating separate premia by gender. This choice also avoids the reference-category issue discussed above: using separate wage premia for men and women within each category would require choosing one wage structure as the benchmark for the decomposition – that of men, women, or a pooled sample. Since the choice of benchmark can affect the results, this would introduce the same index-number problem found in traditional Oaxaca-Blinder decomposition. This evidence is suggestive rather than a formal test: the within-field gender wage gap is not significantly related to the share of women in the field ($r = 0.079$, $p = 0.6284$, weighted; Table 4, Section 5). This indicates that relaxing this assumption would likely add limited explanatory value. This is broadly consistent with the absence of a clear devaluation pattern across fields, and with previous evidence of limited within-firm gender discrimination in the Belgian private sector (Martins & Vandenberghe, 2010; Vandenberghe, 2011). The remaining unexplained heterogeneity is discussed briefly in Section 5 and as a direction for future research in Section 9.

An additional advantage of this framework is that it can also be used to estimate the indirect contribution of the field-of-study variable. Field-of-study choices may affect wages not only directly, but also indirectly by influencing the occupations and sectors in which individuals later work. By accounting for these links, the analysis can distinguish between the direct contribution of field of study and its indirect contribution through occupational and sectoral sorting. This extension is developed further in Section 3.4.

3.3.2. Gelbach Relative to Alternative Approaches

The most relevant comparison for evaluating Gelbach's advantages is with sequential regression approaches. Rather than treating observable characteristics as a whole, these

approaches estimate the contribution of specific groups of variables, such as field of study, occupation, or working time. A common approach, based on incremental regression⁹, adds these groups of variables ($Z_1, Z_2, Z_3, \dots$) sequentially and uses the resulting changes in R^2 to estimate their contributions to the explained gender wage gap. However, because these variables are often correlated with each other, the results depend on the order in which they are added – a limitation explicitly identified by Gelbach (2016) as the central motivation for his method. Moreover, Naszodi (2021) shows that decomposition results can be highly sensitive to the choice of decomposition scheme, and that a path-independent specification is preferable when the sequence of changes in the underlying factors is unknown. For instance, if field of study is added before occupation, it may capture part of the explanatory power that would otherwise be attributed to occupation, and vice versa. Gelbach avoids this problem by providing an exact decomposition that does not depend on the order of the variables.

Traditional Oaxaca-Blinder decompositions (Blinder, 1973; Oaxaca, 1973), by contrast, involve the familiar index-number/reference-coefficient problem: the explained and unexplained components depend on the reference wage structure used to value differences in characteristics – men’s returns, women’s returns, or a pooled wage structure – and different choices can produce different explained/unexplained decompositions.

A related question is whether Gelbach’s aggregate contributions depend on the reference category chosen for a categorical variable such as field of study, occupation, or sector. They do not. Changing the reference category only changes the way the regression coefficients are defined: although the individual coefficients Γ change, the predicted value $Z\hat{\Gamma}$ for each individual remains the same. As a result, the aggregate contribution $\delta_Z = \sum_k p_k * \gamma_k$ is unchanged. However, the contribution of individual categories can change because each γ_k is measured relative to the chosen reference category. This category therefore affects the category-level results in Section 6.2., but not the overall contribution of the dimension. This was also verified for the Sector variable (Tables A12a & A12b, Appendix 12).

Finally, Gelbach should be understood as a decomposition method rather than a causal identification strategy. It shows how much of the gender wage gap is statistically associated with each dimension, given the observed wage structure, but it does not establish that field of study, occupation, or sector causally determine wages.

⁹ **Incremental regression** refers to the practice of adding covariates to a regression model one at a time, rather than all at once.

3.4. Extending the Strategy: Direct and Indirect Contribution

As discussed in Section 3.1., field of study can affect the gender wage gap through more than its direct effect. It can also influence wages indirectly by shaping the occupations – and, to a lesser extent, sector – in which graduates work. As shown in the same section, this channel can be substantial (Sloane et al., 2021), suggesting that part of what appears to be an occupational effect may actually originate from differences in field-of-study choices.

Graduates from different fields may therefore earn different wages for two reasons: their fields may have different associated wage premiums, or they may lead to occupations with different wage levels. To capture the full contribution of field of study, it is therefore necessary to identify the part of occupational sorting that can be traced back to field-of-study segregation.

To do so, the ancillary regression for each occupation k is augmented with field of study, denoted Z^{field} , as an additional control:

$$occupation_k = \tilde{\alpha}_k + \tilde{\rho}_k \cdot fem + Z^{field} \pi_k + \tilde{u}_k \quad [8]$$

Here $\tilde{\rho}_k$ measures the remaining gender difference in representation in occupation k after accounting for field of study. Repeating the aggregation gives:

$$\tilde{\delta}^{occ} = \sum_k \tilde{\rho}_k \gamma_k \quad [9]$$

Intuitively, the ancillary regression asks whether occupational gender imbalances remain after accounting for field of study. Consider for instance two occupations, doctors and teachers, with estimated wage premia of $\gamma_{doctor} = 10$ and $\gamma_{teacher} = 5$. Although these premia are already estimated conditional on field of study in the main regression, men and women may still differ in how they are distributed across these occupations because of their earlier field-of-study choices. For example, medical graduates may be more likely to become doctors, while literature graduates are more likely to become teachers. As a result, the heterogeneity term¹⁰ can differ between men and women because of field-of-study sorting, even when the occupation-specific wage premia are held fixed. Adding field of study as a control to the ancillary regression removes the part of the occupational imbalance linked to field-of-study choice. The remaining difference, $\rho_k - \tilde{\rho}_k$, therefore captures the share of occupational gender imbalance that can be attributed to field of study. Multiplying this difference by the relevant wage premium and

¹⁰ The **heterogeneity term** for a given dimension is the estimated value $Z\hat{\mathbf{f}}$ from equation [4], evaluated using each individual's own category. For occupation, it takes the value $\hat{\Gamma}_{doctor}$ for doctors, $\hat{\Gamma}_{teacher}$ for teachers, and so on. In other words, each individual is assigned the estimated wage premium to the occupation they actually hold.

aggregating across categories, as in equation [9], gives the corresponding share of occupational sorting's contribution to the wage gap. The same approach is applied to sector.

Overall, field of study's total contribution is thus the sum of its direct effect and its indirect effects through occupation and sector. Comparing these components shows whether field of study matters mainly because it shapes where men and women work – through their over- or under-representation across occupation and sector – or because fields themselves are associated with different wage premiums.

3.5. Statistical Inference: Bootstrap

Beyond estimating each δ_k , it is also important to determine whether these contributions are statistically different from zero. This is not straightforward at first because δ_k is not a single estimated coefficient, but the product of two parameters: γ_k , estimated from the full wage regression [4], and ρ_k , estimated from the ancillary regression [5]. Since these parameters come from two models estimated on the same sample, the standard error of δ_k depends not only on their individual uncertainty but also on their covariance. Standard regression outputs do not provide this covariance directly. The same issue arises for the grouped direct and indirect contributions discussed in Section 3.4. above, which are based on differences between several estimated quantities.

To address this issue, standard errors and significance levels for all δ_k , as well as their aggregated and grouped versions, are obtained using a nonparametric bootstrap (Efron & Tibshirani, 1993). The sample is repeatedly resampled with replacement, and the entire estimation procedure – including the baseline and full wage regressions, and the calculation of the direct and indirect contributions – is repeated for each resample. This produces a bootstrap distribution for each contribution, from which standard errors, confidence intervals, and p-values are calculated. Because $\hat{\rho}_k$ and $\hat{\gamma}_k$ are re-estimated together in each replication, the bootstrap automatically accounts for their covariance. This approach is used throughout Section 6 to assess the statistical significance of the estimated contributions, including the grouped decomposition presented in Table 6b.

4. Data description

4.1. Description of the data

Building on the conceptual framework, research and sub-research questions, and the empirical strategy of the thesis, this section presents the dataset used for the empirical analysis, including its source, coverage, and suitability for answering the research question.

The analysis uses the *Vacature SalarisEnquête* (Vacature Salary Survey), one of Belgium's longest-running and largest wage surveys. Conducted annually since the early 2000s by *Vacature*, a Dutch-language recruitment magazine and job platform, the survey collects information on salaries (monthly gross wage), fringe benefits¹¹, working time, and career intentions, as well as individual characteristics such as educational attainment and field of study, age, gender, seniority, occupation, sector and a range of employer and benefit characteristics. Unlike administrative payroll data¹², it is based on a voluntary online questionnaire completed directly by employees. Each wave typically gathers several thousand respondents, making it one of the largest available sources of individual wages in Belgium, even though the exact sample size varies across editions.

This analysis draws on the 2008, 2010 and 2012 waves of the survey, pooled into a single harmonised dataset containing 158,579 observations¹³. Pooling these three waves offers two main advantages. First, it substantially increases sample size and statistical power for the field-of-study decomposition, which requires sufficiently large cell sizes across a large number of categories (40 fields of study). Second, because educational segregation and the wage structure across fields are not expected to shift considerably over a four-year window, pooling these waves does not compromise the interpretability of the results, while allowing year fixed effects to absorb any residual time variation.

As with any survey-based source, the sampling design has implications for interpretation. Participants are not selected through a probability sample of the Belgian workforce but are recruited voluntarily through the *Vacature* website, newsletters, media campaigns, and partner organisations. This recruitment strategy implies potential coverage and self-selection bias, since

¹¹ **Fringe benefits** refer to non-wage forms of compensation provided by an employer, such as a company car, meal vouchers, group or hospitalization insurance, or additional paid leave.

¹² **Administrative payroll** records refer to wage data collected directly from official employer declarations or administrative sources, rather than self-reported by individuals. Unlike the survey data used in this thesis, they are not subject to recall error or survey selection bias.

¹³ The data were provided by my UCLouvain thesis supervisor, **Vincent Vandenberghe**, who granted access to a harmonized dataset covering the 2008, 2010 and 2012 waves.

online opt-in surveys tend to over-represent respondents who are better educated, more digitally engaged, and more inclined to participate in salary-related questionnaires than the average worker (Smyk et al., 2018). This limitation is particularly relevant for studies seeking to generalise to the entire Belgian workforce.

However, this concern is somewhat reduced here, as the analysis focuses on tertiary-educated workers, who are more likely to be reached by this type of online survey (Smyk et al., 2018). In 2025, more than half of Belgians aged 25-34 held a tertiary degree (52.7%, Statbel, 2026), and tertiary education has, for the first time, become the most commonly held qualification among the population aged 15 and above (37.5%, compared with 37.3% for upper-secondary education) (Statbel, 2026). This suggests that the digitally reachable, employed respondents in the *Vacature* survey are more representative of the tertiary-educated population than of the general workforce.

Finally, as mentioned above, the dataset does not include sampling weights. All descriptive statistics and regression results presented in this thesis are therefore unweighted and reflect the composition of survey respondents rather than a weighted estimate of the Belgian tertiary-educated workforce – a limitation revisited in Section 7.

4.2. Main variables and indicators

Using this dataset, this section describes the variables entering the wage equation, organised into three groups: the dependent variable, the variable of interest, and the control variables.

Table 1: Variables and indicators used in the wage equation

/	Indicator	Description	Construction / Source
Dependent variable	<i>gwage</i>	Gross monthly wage	Self-reported, raw survey variable
	<i>lngwage</i>	Log of gross monthly wage	$\log(\text{gwage})$; used in the Gelbach decomposition
	<i>gwagew_eur</i>	Gross monthly wage adjusted for working time, in euros	$(\text{gwage} / \text{wkpc}) * 100$; used for descriptive statistics only
Variable of Interest	<i>dipl</i>	Field of study (highest tertiary diploma)	40 categories, recoded from text labels to numeric codes via the codebook
Control variables	<i>wkpc</i>	Working time (% of full-time equivalent = FTE)	Self-reported, raw survey variable
	<i>occh</i>	Occupation, harmonised across waves	Recoded from text labels via the codebook
	<i>secth</i>	Sector of activity, harmonised across waves	Recoded from text labels via the codebook

/	Indicator	Description	Construction / Source
	<i>age</i>	Age of respondent	Self-reported; entered as a factor in the wage regression
	<i>yr</i>	Survey year (2008 / 2010 / 2012)	Entered as a factor (year fixed effects)
Sample restriction	<i>educat</i>	Highest educational attainment	Restricted to codes 5–6 (Bachelor / Master) to define the work sample
	<i>fem</i>	Gender indicator (1 = woman, 0 = man)	Self-reported, raw survey variable

*Notes: The working sample (work) restricts the pooled 2008, 2010 and 2012 waves to respondents with codes 5–6 on *educat* (Bachelor or Master), consistent with the tertiary-educated population targeted by the research question (Section 2.1.3.).*

❖ Dependent Variable

As summarised in the Table 1 above, the dependent variable is the individual’s gross monthly wage. Two versions of this variable will be used depending on the purpose of the analysis: the log of gross monthly wage (*lngwage*) will be used in the Gelbach decomposition, consistent with the Mincerian wage equation framework (Mincer & Ofek, 1980), as the log specification allows coefficients to be interpreted in proportional terms. By contrast, the working-time-adjusted wage in euros (*gwagew_eur*) will be used exclusively for the descriptive statistics presented in Section 5, where a euro-denominated measure is more intuitive than a log-transformed one. Because gross wages are recorded in nominal terms, some variation across survey years reflects overall wage growth rather than changes in the gender wage gap. This issue is addressed by including survey year (*yr*) as a factor in all wage regressions, with year fixed effects capturing differences in average wage levels across waves.

❖ Variable of Interest

The main variable of interest is field of study (*dipl*), which identifies each respondent’s highest tertiary diploma across 40 categories. The full list of fields of study – with “b-” denoting a Bachelor’s degree and “m-” a Master’s degree – is presented in Table A2 (Appendix 2). Of the 158,577 raw observations, the analysis retains 106,520 respondents with a Bachelor’s or Master’s degree, as lower education levels are not classified by field of study in the survey (missing values for *dipl* are addressed in Section 4.3. below).

As introduced in the conceptual framework, field of study is the covariate Z whose contribution to the gender wage gap is isolated through the Gelbach decomposition. Its choice follows directly from the Research Question: rather than treating education as a single continuous or ordinal measure, *dipl* captures qualitative differences in what men and women study, allowing

the analysis to focus on educational segregation rather than on differences in educational attainment, i.e. highest education level.

❖ Control Variables

Finally, the wage equation will further include working time (*wkpc*), occupation (*occh*), sector of activity (*secth*), and standard demographic controls – age (*age*) and survey year (*yr*). Working time captures the part-time wage penalty (Section 2.1.4.), while sector controls for broad differences in pay levels across industries. Occupation plays a somewhat different role: it is also expected to be a channel through which field of study indirectly affects the gender wage gap, as men and women who choose different fields of study subsequently sort into different occupations. This direct–indirect distinction, formalised in Section 3.4., will be revisited when interpreting the results. As with any categorical variable included in a linear regression, one category is omitted for each of *dipl*, *occh*, and *secth* to serve as the reference against which all other coefficients are interpreted. The omitted categories are **Secretarial studies/Languages** (for field of study), **Administrative staff** (for occupation), and **Agriculture & Food Industry** (for sector), and all reported coefficients on these variables should be read as wage differentials relative to these respective reference groups. As explained in Section 3.3.2., the choice of reference category affects only the distribution of the contribution across individual categories, not the overall Gelbach contribution of the variable.

4.3. Missing values

Before presenting the descriptive statistics, it is useful to examine the extent and sources of missing data, since excluding observations with missing values reduces the estimation sample from 106,520 to 48,785 observations (Table A3 in Appendix 3). Missing values are highly concentrated in a small number of variables (Table A10, Appendix 10): occupation (*occh*) is missing for 49.1% of the tertiary sample, by far the main source of attrition. By comparison, field of study (*dipl*) is missing for 8.4% of respondents, gross wage (*gwage*) for 1.2%, and working time (*wkpc*) and sector (*secth*) are almost fully observed. Overall, these numbers indicate that occupation, rather than field of study, is the main reason for the reduction in sample size.

Given the central role of field of study in this thesis, missing values for this variable were further examined by gender. Table A4 (Appendix 4) shows that missing values for *dipl* are slightly more common among women than men (9.48% versus 7.52%). Although this difference is statistically significant ($\chi^2 = 130.04$, $p < 0.001$), its magnitude is small. This suggests that missingness on field of study is only weakly related to gender, although a limited degree of

selection bias cannot be completely excluded. This issue will be acknowledged as a limitation of the analysis (Section 7).

Because occupation accounts for most of the sample attrition, the same check is needed for *occh*. Table A5 (Appendix 4) shows a considerably larger gender gap than for field of study: missing occupation data are more common among men than women (51.90% vs. 45.42%; $\chi^2 = 439.01$, $p < 0.001$). This suggests that occupation missingness is not random with respect to gender and could affect the estimated occupation contribution. However, the robustness checks in Section 6.7. show that the main results remain unchanged when restricting the analysis to 2012, the year with the most complete occupation data, suggesting that missing occupation data do not drive the main conclusions.

Finally, it is also worth addressing the treatment of outliers. No additional trimming¹⁴ or winsorizing¹⁵ was applied to the wage or working-time variables beyond the missing-data restrictions described above. A diagnostic check found no implausible values: working time remains within the expected 0-100% range, and the highest reported wages (>€15,000/month, 0.1% of the sample) are all associated with full-time work. This avoids introducing an additional layer of sample selection. A small number of unusually high wages remain in the field-of-study analysis, particularly in Medicine, and are addressed when interpreting the results rather than removed from the sample.

5. Descriptive Statistics

Before using these data to estimate the contribution of educational segregation through the Gelbach decomposition, it is useful to first examine the main patterns of the analysed sample. This section first describes its key characteristics and the main differences between men and women in education, working time, and labour-market experience (Section 5.1.). It then examines three patterns central to this thesis: the overall gender wage gap and how it changes as controls are added; the uneven representation of men and women across fields of study; and differences in wage levels across fields (Section 5.2.). Finally, it considers whether wage differences across fields mainly occur between or within fields, and whether the segregation observed in fields also extends to occupations (Section 5.3.)

¹⁴ **Trimming** refers to removing observations beyond a chosen threshold (e.g. below the 1st or above the 99th percentile).

¹⁵ **Winsorizing** refers to capping such observations at that threshold rather than removing them. Both are common ways of limiting the influence of extreme values on estimated means or regression coefficients.

These descriptive results will provide the necessary background for the Gelbach decomposition developed in the next section, which isolates the specific contribution of field of study from these correlated factors.

5.1. Sample and Key Control Variables

To start with, this section presents the working sample and main gender differences in observable characteristics – degree level, working time, and labour-market experience – that need to be accounted for before any wage differential can be attributed to field of study.

As presented in the Table 2 below, the working sample further restricts the 106,520 tertiary-educated respondents introduced in Section 4.2. to those with non-missing wage, working time and age – 105,164 respondents (60,200 men and 44,964 women). Missingness on these three variables remains limited overall (Section 4.3.), consistent with the small gap between this sample and the 106,520 tertiary-educated respondents from which it is drawn.

Table 2: Gender differences in degree level, working time and experience

Indicator	Men	Women
N	60,200	44,964
Master's degree (%)	53.4	45.8
Mean age (years)	36.3	34
Mean experience (years)	12.2	10.5
Mean working time (% FTE)	99.4	95.4
Working part-time (%)	2	14.8
Experience/age ratio – Bachelor	0.367	0.346
Experience/age ratio – Master	0.315	0.263

It is visible that men are slightly more likely than women to hold a master's degree (53.4% versus 45.8%). Although this difference in educational level is relatively modest, it may still contribute to wage differences independently of field of study. As explained in Section 6.1., this difference is already captured by the field-of-study variable (*dipl*), as each field-of-study category is defined separately for Bachelor's and Master's degrees.

Consistently with the literature, working-time patterns differ much more sharply. Average working time is only slightly lower among women (95.4% versus 99.4%), yet women are substantially more likely to work part-time (14.8% compared with 2.0% of men). This

asymmetry is in line with the part-time wage penalty documented for Belgium in Section 2.1.4., and motivates the use of the FTE-equivalent wage measure (*gwagew_eur*) throughout the rest of this Descriptive Analysis. Consequently, field-of-study wage comparisons will not mechanically be confounded by differences in hours worked. It is also worth noting that working time is heavily concentrated near full-time, a pattern not fully representative of real-world working arrangements, where part-time work remains more common, particularly among women. This is discussed further in the limitations (Section 7).

Moreover, women accumulate less labour-market experience on average (10.5 years against 12.2 years), consistent with their slightly younger mean age in this dataset. The last two lines of Table 2 show that this gap, however, is not uniform across degree levels. Relative to age, the experience difference is small among Bachelor's graduates but substantially larger among Master's graduates. This pattern aligns with the motherhood penalty discussed in Section 2.1.4. (Working time and career interruption), suggesting that career interruptions are more pronounced – and more costly in wage terms – among highly educated women. Therefore, age and survey-year fixed effects will be included in the regressions of the Section 6.

Although the analysis is not restricted to the private sector, the sample is predominantly composed of private-sector workers, with only 8.7% of respondents reporting work in the “Government/Public sector” category. Belgian public-sector wages have historically followed standardised salary scales based on seniority and diploma level (Weber, 2015), rather than field of study. As a result, the estimated wage premia may be somewhat lower than those in a purely private-sector sample, which, if anything, would lead this thesis to underestimate rather than overestimate field of study's true contribution.

Taken together, these covariates justify controlling for the education level (*educat*), working time (*wkpc*), and age and survey year (*age/yr*) throughout the rest of this analysis, before any role can be attributed to field of study itself.

5.2. Three Key Patterns

This section highlights the three patterns most central to this thesis: the overall size of the gender wage gap, the extent to which men and women are unevenly represented across fields of study, and the extent to which fields of study differ in their associated wage levels.

5.2.1. The raw Gender Wage Gap

This section presents the raw gender wage gap observed in the sample, both overall and across survey years and degree levels, before turning to its relationship with field of study in the next section. Consistent with the unadjusted gender wage gap defined in Section 2.1.1., the relative gap is computed as the difference between men's and women's mean monthly wage, expressed as a percentage of men's mean wage, without controlling for any observable characteristics.

Table 3 shows an overall gender wage gap that is much larger than the 9.1% unadjusted gap reported in official Belgian statistics (Section 2.1.3.). Since official estimates are based on hourly wages, the working-time-equivalent wage used in this thesis provides the closest comparison. Applying this adjustment reduces the gap from 25.7% to 22.3%, but it remains more than twice the official figure. This difference may reflect the higher gap among tertiary-educated workers (Section 2.1.3.) and the fact that the *Vacature* survey is voluntary and unweighted. However, once age, working time, field of study, sector, and occupation are controlled for, the gap falls to 8.3% (Figure 1), close to the official adjusted estimate of below 8% (Section 2.1.3.). Table 3 also shows a slightly wider gender wage gap among Master's graduates than Bachelor's graduates (25.7% vs 23.2%), a difference of 2.5 percentage points. Rather than narrowing the gap, higher educational attainment is therefore associated with a larger wage difference between men and women. This pattern is consistent with the glass ceiling effect discussed in Section 2.1.3. and with Kampelmann et al.'s (2018) finding that highly educated Belgian women may receive lower returns to their characteristics than their productivity would predict (Section 2.1.4.). A visual representation of the distribution of wages by gender and degree level is presented in Figure A1 (Appendix 5).

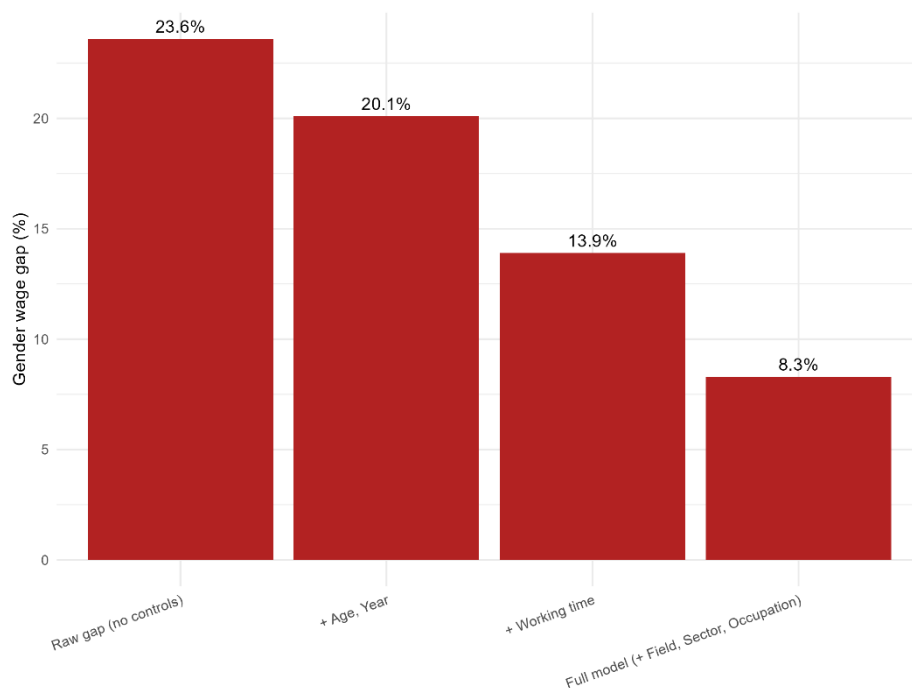
Table 3: The gender wage gap, overall, by year, and by degree

Category	Group	Mean wage - Men (€)	Mean wage - Women (€)	Absolute gap (€)	Relative gap (%)	Relative gap, working-time adjusted (%)
Overall	Overall	3,786	2,812	975	25.7	22.3
By year	2008	3,703	2,642	1,061	28.7	25.5
	2010	3,807	2,824	984	25.8	22.0
	2012	3,830	2,949	881	23.0	19.7
By degree	Bachelor	3,331	2,558	773	23.2	—
	Master	4,136	3,075	1,062	25.7	—

These findings are consistent with the experience gap documented in Section 5.1., which is also large among Master’s graduates. However, it remains unclear whether these differences are linked to the degree level itself or to the fact that men and women tend to study different fields, which offer different labour-market rewards.

Figure 1 below summarises the evolution of the raw gender wage gap as different sets of controls are added, using the final estimation sample ($N = 48,785$) used throughout the rest of the thesis. The raw gap shown here (23.6%) therefore differs slightly from the overall gap in Table 3 (25.7%), calculated on the larger descriptive sample ($N = 105,194$; Section 5.1.). The gap falls to 20.1% after controlling for age and survey year, and then to 13.9% when working time is added. This large reduction is consistent with the large difference in part-time work documented in Section 5.1. Adding field of study, sector, and occupation further reduces the gap to 8.3%. Overall, no single factor explains most of the gap: age, working time, and sorting across fields of study, sectors, and occupations all contribute, together accounting for around two-thirds of the raw gap.

Figure 1: Gender wage gap, from raw to residual



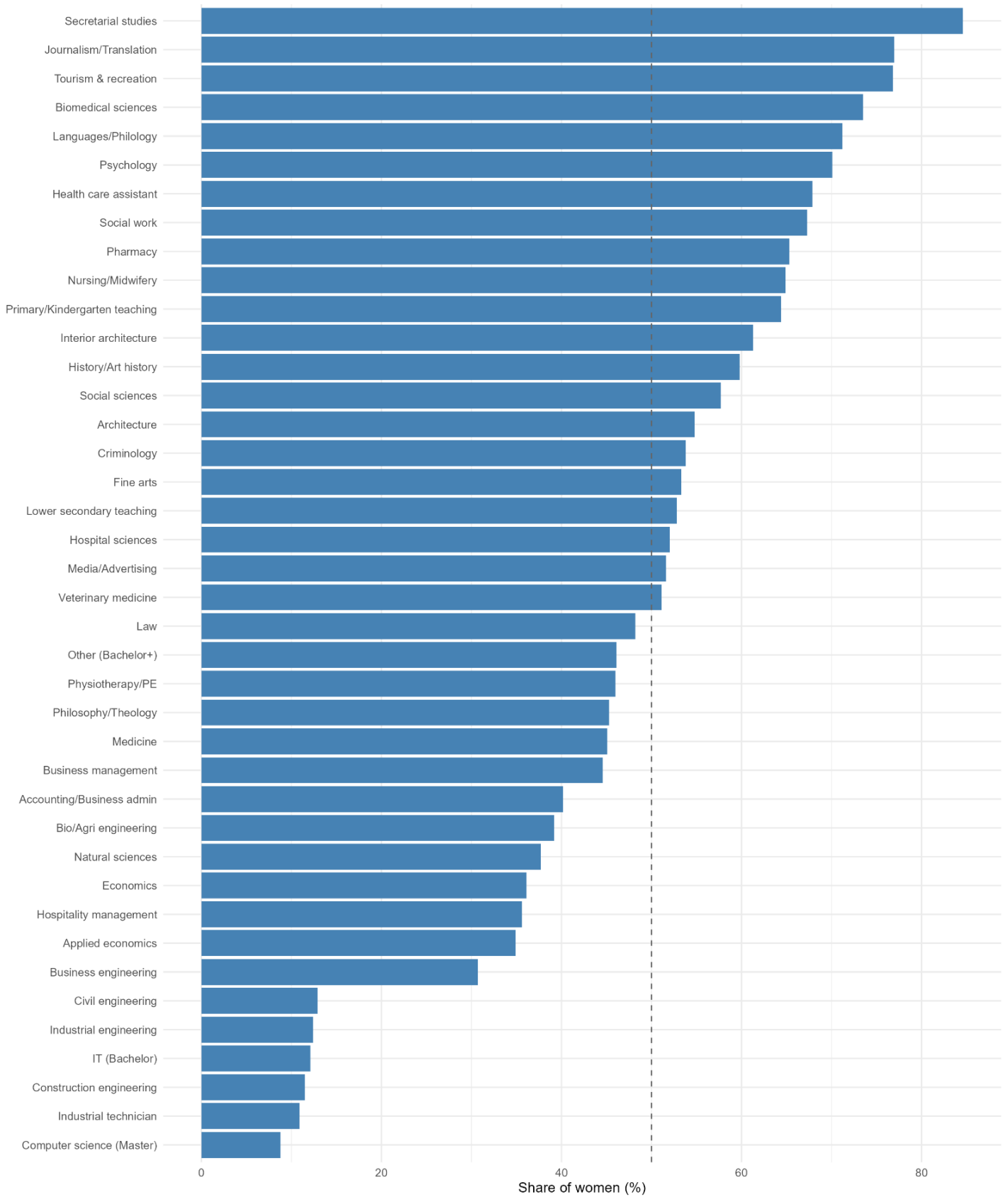
It should be noted that, unlike the Gelbach decomposition, this sequential approach adds controls in a fixed order, meaning that the reduction attributed to each block may partly depend on that order. Nevertheless, it provides a simple first indication of how much of the gender wage gap can be explained by observable characteristics.

5.2.2. Representativeness (ρ): Gender Segregation across Fields of Study

This section turns to field of study itself, the central variable of interest in this thesis, and documents how unevenly men and women are distributed across these fields (sorting effect).

Figure 2 below shows the share of women across all fields of study documented in the dataset, revealing substantial variations: some fields are almost entirely female (Secretarial studies, above 80%), while others are almost entirely male (Computer science and several engineering fields, below 15% of women). This important segregation therefore justifies a more formal analysis using the Gelbach decomposition to assess how much of this uneven distribution contributes to the gender wage gap.

Figure 2: Share of women by field of study



A detailed breakdown of the ten most feminized and masculinized fields, including their wages and within-field gender wage gap, is presented in Table A6 (Appendix 6). It shows that feminized fields generally pay less than masculinized ones, except for Pharmacy, which combines a high share of women with one of the highest wages. Comparing fields by degree level (Table A7, Appendix 7) confirms this pattern on average, but also shows that it is not mechanical: sector and qualification may matter as much as a field's gender composition.

The next section turns to how wages themselves vary across fields of study, independent of their gender composition.

5.2.3. Wage Premium (Γ) across Fields of Study

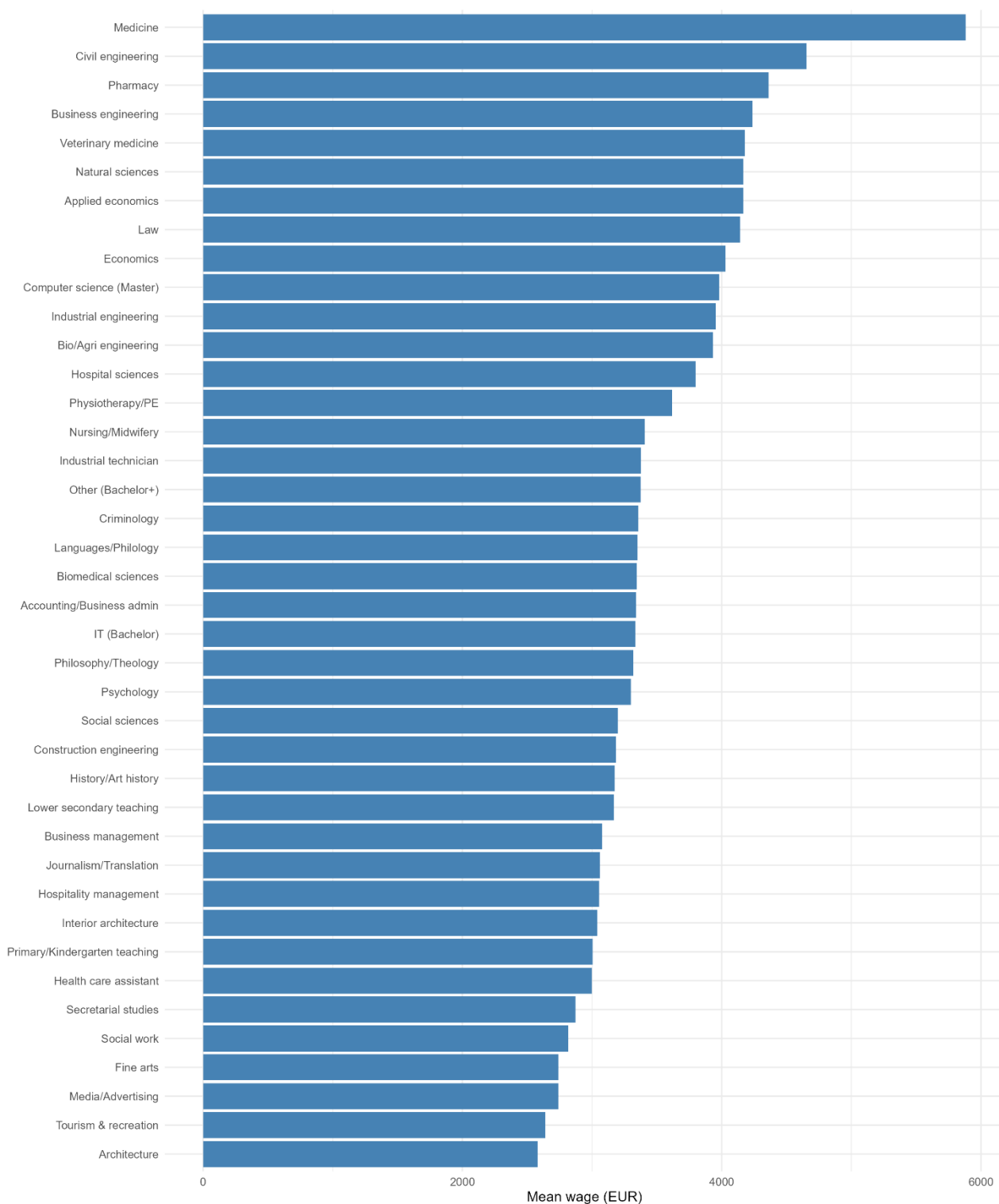
Figure 2 documented how unevenly men and women are represented across fields of study. This section turns to the other side of the relationship: how much fields of study differ in wages, independently of who studies them. Figure 3 below shows the mean gross monthly wage (FTE-equivalent) for each of the 40 fields in the sample.

This pattern is consistent with the returns to specialised, technical and health-related training found in the literature (Sloane et al., 2021; Section 2.2.1.). Fields such as Medicine, Civil Engineering, Pharmacy, and Business Engineering have the highest average wages (above 4,000€), while more creative or service-oriented fields – Architecture, Tourism & Recreation, Media/Advertising, and Fine Arts – rank lowest (around 2,800€/3,000€).

This figure shows unadjusted mean wages, not the wage premium γ_k defined in Section 3.3.1., which accounts for differences in occupation, sector, and working time. It should therefore be interpreted as descriptive evidence of overall wage differences across fields, rather than as estimates of the return to field of study after accounting for other factors; the latter is estimated through the Gelbach decomposition in Section 6.

Taken together, Figures 2 and 3 show the two conditions – unequal representation and unequal pay – that determine whether educational segregation contributes to the gender wage gap (Section 3.3.1.). The next section examines how these two dimensions are related to each other and to the gender wage gap before turning to the presentation of the results obtained through the Gelbach decomposition that formally combines them.

Figure 3: Mean FTE-equivalent monthly wage by field of study



5.3. Field of Study and the Residual Gap

The patterns just described – feminized fields generally paying less, with notable exceptions such as Pharmacy – are confirmed statistically in Table 4 below, computed across the 40 fields of study of the dataset, together covering 97,611 workers.

Table 4 reports two correlations, each computed across the 40 fields of study: the correlation between a field’s share of women and its mean wage, and the correlation between a field’s share of women and its within-field gender wage gap.

Table 4: Correlations between field feminization, mean wage, and within-field gap

Relationship	r (gwage, unweighted)	p-value (unweighted)	r (gwagew, FTE-adjusted)	p-value (FTE-adjusted)
Share of women vs. mean wage of the field	-0.457	0.0030	-0.399	0.0107
Share of women vs. within-field gender wage gap	0.207	0.1992	0.079	0.6284

The first correlation statistically confirms this pattern: fields with a higher share of women tend to have lower average wages, a relationship that remains statistically significant even once working time is accounted for ($r = -0.399$, $p = 0.0107$). The second correlation, by contrast, is close to zero and not statistically significant ($r = 0.079$, $p = 0.6284$): how feminized a field is says nothing about the size of the wage gap between men and women within that same field i.e. that a field with a high share of women can exhibit either a small or a large gender wage gap. Further explaining this residual within-field variation is beyond the scope of this thesis (Section 3.3.1.), though it is briefly discussed as a direction for future research (Section 9).

Taken together, these results point to two distinct mechanisms behind the gender wage gap. The first is **sorting**: men and women are not evenly distributed across fields, and feminized fields tend to lead to lower-paying outcomes overall – consistent with matched employer–employee evidence for Belgium showing that the gender wage gap mainly reflects the unequal distribution of men and women across firms and sectors, rather than wage discrimination within firms (Martins & Vandenberghe, 2010; Vandenberghe, 2011). The second is a within-field gap that appears unrelated to how feminized the field is, and therefore likely stems from other factors – working time, occupation, or discrimination – that this descriptive analysis cannot disentangle on its own. This distinction directly motivates the Gelbach decomposition, which is designed to separate these two channels. Results are presented in Section 6.

The segregation documented across fields of study also appears to extend to occupations: the dissimilarity index, which measures the share of workers who would need to change category for men and women to be equally distributed (as introduced in the literature review, Section 2.1.6.), is nearly identical for field of study and for occupation (37.4%), and considerably lower for sector of activity (21.9%; Table A8, Appendix 8). This near-identical level of segregation across field of study and occupation suggests that occupational sorting is largely driven by educational sorting: men and women do not appear to be substantially re-sorted into different occupations beyond what their initial field of study already predicts. This is consistent with the discussion in the literature review, which highlighted occupational sorting as a key mechanism underlying the gender wage gap in Belgium. It reinforces the idea that these disparities emerge well before labour-market entry, through the unequal distribution of men and women across fields of study.

As explained above, the remaining 8.3% (Figure 1) is the unexplained part of the wage gap after accounting for demographic characteristics, working time, field of study, sector, and occupation. What this residual reflects – whether unobserved productivity differences, negotiation behaviour, or discrimination – is beyond the scope of this thesis and is left to the broader literature (Section 2.1.5.). Instead, this thesis focuses on how much of the reduction from the raw gap to this residual can be attributed to each factor, and particularly to field of study. The Gelbach decomposition results presented in the following section formally separate these contributions.

6. Main Results

This section presents the results of the Gelbach decomposition introduced in Section 3. It first presents the two nested wage models and the resulting raw and residual gender wage gaps (Section 6.1., Table 5). It then examines how the gap is explained by working time, field of study, occupation, and sector, first in aggregate (Section 6.2., Tables 6a and 6b) and then separately for each dimension (Sections 6.3.–6.5.), starting with field of study, the main focus of this thesis. Section 6.6. brings these dimensions together to highlight the common pattern, while Section 6.7. tests the main findings through two robustness checks. Statistical significance is assessed throughout using bootstrap resampling (Section 3.5.), with results reported for the aggregate decomposition (Table 7) and the field-by-field breakdown (Table 8, Section 6.3.).

6.1. Model and Sample

As introduced in Section 3.3.1., the empirical strategy relies on two nested wage regressions. The baseline model regresses log wages on the gender indicator, age and survey year:

$$\ln(wage_i) = \alpha^b + \beta^b \cdot fem_i + AGE_i' \theta + YR_i' \varphi + \varepsilon_i^b \quad [10]$$

The full model further adds working time, occupation, sector, and field of study:

$$\ln(wage_i) = \alpha^f + \beta^f \cdot fem_i + AGE_i' \theta + YR_i' \varphi + wkpc_i + occh_i + secth_i + dipl_i + \varepsilon_i^f \quad [11]$$

where β^b is the raw gender wage gap and β^f is the gap that remains once all controls are included. Age and survey year are entered as full sets of dummy variables, meaning that each age and each survey year can have its own effect on wages. This allows the model to capture non-linear age-wage patterns and differences between survey years without imposing a specific relationship between age and wages. Degree level is not included separately, as it is already captured by field of study, whose categories distinguish between Bachelor's ("b-") and Master's ("m-") fields. Age, rather than potential experience, is used to control for career stage: the two are highly collinear in this sample (age explains about 90% of the variation in experience), while adding experience leaves β^f nearly unchanged and reduces the sample by about a quarter (Table A9, Appendix 9).

The estimation sample (*work_decomp*) further restricts the 105,164 respondents introduced in Section 5.1. (with non-missing wage, working time, and age) to those with non-missing occupation, sector, and field of study as well, leaving 48,785 observations. Most of this additional reduction stems from missing occupation data (49.1%, against 8.4% for field of study and under 2% for wage and working time; Table A10, Appendix 10), concentrated in the earlier survey years and slightly more common among men (51.9% vs. 45.4%; Section 4.3.). Because occupation drives this attrition, any result depending on it – including occupation's own contribution and the part of field of study's contribution running through it – is re-estimated on the 2012 subsample as a robustness check and will be developed further in Section 6.7.

Two further checks support the validity of the categorical variables. First, field of study records only each respondent's first diploma. The sample shows that only 3.7% report a second one, with no meaningful gender differences (3.61% vs. 3.82%; Table A11, Appendix 11). Second, as shown theoretically in Section 3.3.1., the Gelbach decomposition's aggregate contribution is invariant to the choice of reference category. This is confirmed empirically for sector: although the default reference category (Agriculture & Food Industry) is small, replacing it with the

largest category leaves both β^f and the overall sector contribution unchanged (Tables A12a & A12b, Appendix 12).

Table 5 below reports the resulting raw and residual gender wage gap, estimated on the same estimation sample (*work_decomp*, $N = 48,785$) used throughout Figure 1. The baseline model – including only gender, age, and survey year – gives a wage gap of 20.1%, matching the corresponding step in the sequential decomposition shown in Figure 1 (Section 5.2.1.). After jointly adding working time, occupation, sector, and field of study, the gap falls to 8.3%, a reduction of 11.8 percentage points, again matching the residual reported in Figure 1.

Table 5: Raw and residual gender wage gap, baseline vs. full model¹⁶

Model	Coefficient on fem (log points)	Implied wage gap (%)
Baseline (fem + age + yr)	$-0.2239 = \beta^b$	20.1
Full model (+ wkpc + occh + secth + dipl)	$-0.0865 = \beta^f$	8.3
Reduction	$-0.1373 = \beta^b - \beta^f$	11.8

The next section decomposes this reduction from 20.1% to 8.3% (a change of 0.1373 log points) into the aggregate contribution of each dimension, namely working time, field of study, occupation, and sector, and further decomposes the field of study contribution into the contribution of each individual field.

6.2. Overview: Aggregate Contributions

Following the logic set out in Sections 3.3. and 3.4., this section applies equations [5]-[7], to attribute the 11.8-percentage-point reduction documented in Table 5 above – the 0.1373 log-point gap between β^b and β^f – to each of the four dimensions of the wage equation.

Table 6a reports each dimension’s raw contribution δ_z , computed as $\rho_k * \gamma_k$ summed across categories (cf. equation [7]) or, for working time, directly as $\rho_{wkpc} * \Gamma_{wkpc}$ (cf. equation [6]), before reallocating any indirect effect running through field of study. Working time dominates (48.4%), followed by field of study (25.2%) and occupation (17.4%), with sector contributing the least (9.0%).

¹⁶ Implied wage gap (%) = $(1 - e^\beta) * 100$, where β is the coefficient on fem. This transformation is non-linear, so the “Reduction” row is the difference between the two implied percentages, not a separate transformation of $\beta^b - \beta^f$. The same logic applies to the following tables.

Table 6a: Raw, dimension-by-dimension contribution to the gender wage gap

Component	Delta (log points)	Share of total explained (%)
Working time (wkpc)	-0.0665	48.4
Field of study (dipl)	-0.0346	25.2
Occupation (occh)	-0.0238	17.4
Sector (secth)	-0.0124	9.0
Total (z)	-0.1373	100.0

However, as shown conceptually in Section 3.4. (equations [8]-[9]), part of what Table 6a attributes to working time, occupation, and sector is in fact driven by field of study itself – through gender sorting into working-time arrangements, occupations, and sectors alike. Following equation [8], ρ_k is re-estimated net of field of study for each category, using the real variables introduced in Section 6.1. The resulting indirect and direct contributions follow equation [9] exactly. Table 6b reports the final grouped decomposition, with field of study's contribution defined as the sum of its direct effect and three indirect components, while working time, occupation, and sector are each reported net of field of study.

Table 6b: Final grouped contribution to the gender wage gap, net of field of study

Dimension	Delta (log points)	Share of total explained (%)
Working time	-0.0576	41.9
Field of study (direct + indirect)	-0.0550	40.0
Occupation	-0.0191	13.9
Sector	-0.0056	4.1
Total	-0.1373	100.0

The percentages reported in Tables 6a and 6b show each dimension's share of the 0.1373 log-point reduction ($\beta^b - \beta^f$, equation [6]). They should not be interpreted directly as shares of the percentage-point difference between the raw and residual gaps (20.1% and 8.3%), because the relationship between log wage coefficients and percentage wage gaps is non-linear: converting a log-point reduction into percentage-point one depends on the level from which the reduction

starts, so shares computed in log points do not translate into identical shares in percentage points.

This shift is a key methodological result of the decomposition. Once field of study's indirect effect on the other dimensions is properly attributed to it, working time (41.9%) and field of study (40.0%) become nearly tied at the top, jointly accounting for 82% of the gap attributable to observable characteristics, while occupation and sector together explain only the remaining 18%. Field of study emerges as a driver of the gender wage gap nearly as important as working time – the factor most consistently identified as dominant in the broader literature (Section 2.1.4.). This supports the central argument of this thesis: educational segregation is not simply a secondary source of the gender wage gap; once its contribution is properly isolated, it appears to be one of the two main drivers of the gap among tertiary-educated workers in Belgium.

This near-equality deserves close examination: two estimates can appear similar in a single sample simply by chance. Table 7 below tests whether this is the case using the bootstrap procedure introduced in Section 3.5. The estimation sample is resampled with replacement 300 times, and the full decomposition is re-estimated for each resample. This produces a standard error and 95% confidence interval for each of the eleven quantities. All confidence intervals exclude zero, confirming that each contribution is statistically different from zero and is unlikely to result from sampling variation alone.

Point estimates in Table 7 differ very slightly from those in Tables 6a and 6b (e.g. -0.0548 versus -0.0550 for field of study's total contribution) because Table 7 reports the average across the 300 bootstrap resamples rather than the single point estimate from the full sample. These differences are negligible and do not affect any conclusion.

More importantly, the confidence intervals for working time (net of field of study) and field of study's total contribution do not overlap ($[-0.0604, -0.0549]$ vs. $[-0.0582, -0.0512]$). Even when comparing the highest possible value for working time with the lowest possible value for field of study, the intervals remain separate. This confirms that the near-equal size of the two contributions is a genuine feature of the data rather than a result of sampling variation.

Table 7: Bootstrap estimates, standard errors, and 95% confidence intervals

Label	Estimate	Bootstrap SE	95% CI	Significant at 5%?
Raw gap (beta_b)	-0.2237	0.0029	[-0.2289, -0.2184]	Yes
Residual gap (beta_f)	-0.0865	0.0026	[-0.0919, -0.0816]	Yes
Total explained (delta_z)	-0.1372	0.0025	[-0.1421, -0.1322]	Yes
Working time (delta_wkpc)	-0.0666	0.0015	[-0.0695, -0.0637]	Yes
Occupation (delta_occh)	-0.0237	0.0016	[-0.0269, -0.0208]	Yes
Sector (delta_secth)	-0.0124	0.0009	[-0.0141, -0.0107]	Yes
Field of study, direct (delta_dipl)	-0.0345	0.0019	[-0.0381, -0.0309]	Yes
Working time, net of dipl	-0.0577	0.0014	[-0.0604, -0.0549]	Yes
Occupation, net of dipl	-0.0191	0.0011	[-0.0212, -0.0171]	Yes
Sector, net of dipl	-0.0056	0.0006	[-0.0068, -0.0046]	Yes
Field of study, total (direct + indirect)	-0.0548	0.0019	[-0.0582, -0.0512]	Yes

* All eleven quantities are significant at 5%.

The following sections examine each dimension, starting with field of study.

6.3. Field of Study

Table 8 below reports the direct contribution of all 40 fields of study – that is, each field’s effect on the gap, excluding the portion transmitted indirectly through working time, occupation, or sector (reallocated separately in Table 6b) – ranked by degree level, together with sample size, representation (ρ_k), wage premium (γ_k), and statistical significance based on a nonparametric bootstrap¹⁷ (50 replications, Section 3.5.). Summed across all fields, these direct contributions reproduce the aggregate direct effect of field of study reported in Table 6a (-0.0346, i.e. 25.2%), as confirmed by the total row at the bottom of the table. 32 of the 40 fields are statistically significant at the 5% level. The remaining fields – including the omitted reference category, whose contribution is zero by construction – tend to be smaller fields with less precisely estimated contributions.

¹⁷ A **nonparametric bootstrap** re-estimates the model on repeated resamples of the data (drawn with replacement) to construct standard errors and confidence intervals without assuming a specific distribution.

Table 8: Direct contribution of each field of study to the gender wage gap

Degree	Field of study	N	ρ_k (representation)	γ_k (wage premium)	$\delta_z, \rho_k * \gamma_k$
Bachelor	b-Computing_IT(A1)/Computer science (A1)	2,705	-0.07624	0.0653	-4.981e-03*
	b-Technician (audio, photography, electricity)/Industrial sciences (audiovisual technology, photography, graphic arts, (bio)chemistry, electricity, electromechanics, telecommunications, imaging, textiles)	2,834	-0.07703	0.0308	-2.370e-03*
	b-Construction, wood/Construction engineering, wood technology	290	-0.00896	0.0633	-5.669e-04*
	b-Architecture/Architecture	230	0.00171	-0.0474	-8.091e-05*
	b-Tourism, entertainment/Tourism and recreation	499	0.01442	-0.0053	-7.624e-05
	b-Accommodation, restaurant/Hospitality and hotel management	209	-0.0019	0.0388	-7.365e-05*
	b-Media, publicity/Media, advertising	432	0.00215	-0.0017	-3.703e-06
	b-Secretary/Secretarial studies-Languages	2,934	0.09999	0	0.000e+00
	b-Teaching(lower secondary)/Teacher training for lower secondary education	1,327	0.01161	0.0079	9.139e-05
	b-Teaching(primary, kindergarten)/Kindergarten and primary school teacher	793	0.01572	0.0118	1.847e-04
	b-Business assistant (bookkeeping, taxation, marketing, insurance)/Business management (accountancy, taxation, marketing, legal practice, finance and insurance, communication management, logistics management)	7,582	0.00575	0.0452	2.598e-04*
	b-Medical assistant/Health care (laboratory technician, speech therapy, physiotherapy, occupational therapy, nutrition, dietetics)	870	0.02003	0.0357	7.146e-04*
	b-Social assistant/Social work (social assistant, psychological assistant, special education)	1,684	0.03595	0.0324	1.164e-03*
	b-Nurse, midwife/Nursing, midwifery	1,453	0.02974	0.0651	1.937e-03*
Master	m-Industrial engineers/Industrial engineering, surveying	4,385	-0.11535	0.2058	-2.374e-02*
	m-Civil Engineer/Civil engineering, architectural engineering	1,647	-0.04328	0.34	-1.471e-02*
	m-Computing science/Computer science	557	-0.017	0.2292	-3.896e-03*
	m-Sales engineer/Business engineering, business engineering in information management	1,146	-0.01133	0.3089	-3.500e-03*
	m-Business/Applied economics	2,000	-0.01298	0.2638	-3.425e-03*
	m-Accountant/Sales engineer/Public administration, business engineering, business sciences	2,327	-0.00831	0.1426	-1.185e-03*
	m-Economics/Economics	841	-0.00445	0.2004	-8.910e-04*
	m-Bio Engineer, agricultural engineer/Bio-engineering, agricultural engineering	605	-0.00232	0.2519	-5.854e-04*
	m-Natural sciences (biology, math, chemistry, physics, geology, geography)/Natural sciences (biology, mathematics, chemistry, physics, geology, geography)	858	-0.00251	0.2014	-5.062e-04*
	m-Fine arts (theater, design, music, visual arts)/Arts (drama, music, design, visual arts)	677	0.006	-0.027	-1.618e-04*
	m-Philosophy, Theology/Philosophy, theology	135	-7e-05	0.102	-7.425e-06
	m-Other	422	0.00013	0.1488	1.946e-05

Degree	Field of study	N	ρ_k (representation)	γ_k (wage premium)	$\delta_z, \rho_k * \gamma_k$
	m-Veterinarians/Veterinary medicine	56	0.00068	0.1745	1.189e-04*
	m-Physical education, physiotherapy, speech therapy/Physical education, physiotherapy, speech therapy	289	9e-04	0.1466	1.325e-04
	m-Interior architecture/Interior architecture	310	0.00437	0.0458	2.002e-04*
	m-Medico-social sciences (nurses)/Hospital sciences	202	0.00214	0.2108	4.520e-04*
	m-Criminology/Criminology	363	0.00291	0.1661	4.839e-04*
	m-Biomedical sciences/Biomedical sciences	188	0.00364	0.1995	7.270e-04*
	m-Medical studies (Physicians, Dentists)/Medicine (physicians, dentists)	175	0.00158	0.48	7.580e-04*
	m-History, History of art, Musicology, Archeology/History, art history, musicology, archaeology	530	0.00778	0.1148	8.931e-04*
	m-Pharmacists/Pharmacy	259	0.00487	0.3649	1.778e-03*
	m-Law/Law	1,594	0.00743	0.2714	2.018e-03*
	m-Journalist, translator/Journalism, interpretation, translation	975	0.02475	0.1113	2.755e-03*
	m-Social Science/Social sciences (sociology, political science, communication, public administration)	2,042	0.02227	0.1406	3.131e-03*
	m-Psychology/Psychology, education sciences	1,016	0.0246	0.1692	4.161e-03*
	m-Languages, Philology/Languages, philology	1,344	0.03058	0.1376	4.206e-03*
Total	All 40 fields	48,785	—	—	-3.458e-02

* significant at the 5% level (95% **bootstrap** CI excludes 0), based on 50 replications.

The largest negative contributors are two Master's-level engineering fields. Industrial Engineering (N = 4,385) combines strong female under-representation ($\rho = -0.115$) with a substantial wage premium ($\gamma = 20.6\%$). It therefore makes the largest contribution to the gender wage gap ($\delta = -0.0237$). Civil Engineering (N = 1,647) shows a more moderate representation gap ($\rho = -0.043$), but has one of the highest wage premia in the sample ($\gamma = 34.0\%$). It nevertheless produces the second-largest contribution ($\delta = -0.0147$).

It must be noted that sample size matters as much as representation or wage premium (shadow price) in determining a field's realized impact. A field can have a strong representation gap or a large wage premium, but its contribution will remain small if it includes only a few individuals. Veterinary Medicine is a clear example. It is the smallest field in the sample (N = 56) and has a wage premium similar to that of Languages/Philology ($\gamma = 17.5\%$ vs. 13.8% , respectively). However, its contribution to the aggregate gap is negligible ($\delta \approx 0.0001$) because few individuals are affected. By contrast, Languages and Philology includes more than 20 times as many individuals (N = 1,344). Combined with female over-representation ($\rho = +0.031$) and

positive wage premium, its larger size makes it one of the largest positive contributors, narrowing the gender wage gap ($\delta = +0.0042$).

Notably, both Veterinary Medicine's negligible contribution and Languages and Philology's larger contribution are statistically significant at 5%, showing that statistical significance alone does not imply practical importance. Sample size affects how much a field's representation and wage premium translate into an actual contribution to the gap.

Beyond the specific fields discussed above, the relationship between representation and wage premium can also be assessed across all 40 fields of study. As shown in Table 9, the resulting correlation ($r = -0.223$) is negative but modest, and notably weaker than for sector ($r = -0.310$), though stronger than for occupation ($r = -0.167$, discussed in Section 6.4.). Fields with more women do tend to pay somewhat less, but the relationship is far from mechanical: several of the highest- and lowest-paying fields are close to gender parity. A field-by-field breakdown is provided in Table A13 and Figure A2 (Appendix 13).

Table 9: Correlation between gender representation and wage premium, by dimension

Dimension	N categories	Correlation ρ_k / γ_k
Sector (secth)	22	-0.310
Field of study (dipl)	40	-0.223
Occupation (occh)	60 ¹⁸	-0.167

As shown in Table 6b, field of study's total contribution to the gap ($\delta = -0.0550$) combines a direct effect (- 0.0346, decomposed field by field in Table 8) with an indirect effect of - 0.0204, transmitted through working time, occupation, and sector but measured only at the aggregated level rather than attributed to specific fields. The next two sections examine the reverse direction instead: how much of occupation's and sector's own contributions are inherited from field of study, category by category.

A natural extension would decompose field of study's **indirect** contribution at the level of individual fields – for instance, how much of Industrial Engineering's effect on the gap operates through occupational sorting specifically. This would require a further layer of decomposition beyond the scope of the Gelbach framework applied here and is left for future research (Section 7 and 9).

¹⁸ Note that the column “N categories” excludes categories with fewer than 30 observations (same threshold as Table 8 and Table A14), as their γ_k estimates are too imprecise. This reduces occupation from 77 to 60 categories; field of study (40) and sector (22) are unaffected. Excluded categories remain part of the aggregate contribution reported in Table 6b, so this exclusion has a negligible impact on the results.

6.4. Occupation

As shown in Table 6b, occupation accounts for 13.9% of the gap when measured net of field of study ($\delta = -0.0191$), which is less than field of study and driven almost entirely by two occupations (Table A14, Appendix 14). Company managers ($N = 1,441$) combine moderate female under-representation ($\rho_{CompanyManager} = -0.027$) with the highest wage premium in the sample ($\gamma_{CompanyManager} = 44.5\%$), producing the largest contribution ($\delta_{CompanyManager} = -0.0119$, 50% of occupation's raw total). IT specialists ($N = 4,460$) show the strongest gender imbalance ($\rho_{ITspecialist} = -0.124$) with a moderate premium ($\gamma_{ITspecialist} = 7.5\%$), contributing substantially to the gender wage gap ($\delta_{ITspecialist} = -0.0094$; 39.5%). Together, these two occupations account for 89.5% of occupation's total contribution and, relative to the raw gender wage gap, close to a tenth of it overall (9.5%), while the remaining 75 occupations largely offset one another. This concentration is reflected in the modest correlation between ρ_k and γ_k across occupations ($r = -0.167$, Table 9), which is weaker than for field of study ($r = -0.223$) because it is driven by these two extreme cases rather than a broader pattern (Figure A3, Appendix 14).

Splitting this contribution into direct and indirect components, following the same logic as equation [8] (Section 3.4., applied with real variables in Section 6.2.), reveals two distinct mechanisms. Company managers' imbalance is almost entirely direct, changing little once field of study is controlled for and pointing to an occupation-specific barrier. In contrast, 61% of the IT specialist contribution is indirect, largely inherited from the male-dominated composition of computing and engineering fields of study. Overall, occupation's contribution is highly concentrated rather than diffuse and, aside from these two exceptions, reflects field-of-study segregation more than occupation-specific sorting.

6.5. Sector

Sector accounts for the smallest share of the wage gap when measured net of field of study ($\delta = -0.0056$, 4.1%; Table 6b), but its main interest lies in *why*. Table A15 (Appendix 15) reports all 22 sectors, with none falling below the 30-observations threshold. The four largest contributors — Training & Education, Healthcare & Social Services, the Socio-cultural sector, and (Petro)chemicals/Pharmaceuticals — together account for about 59% of sector's raw contribution, making the pattern more diffuse than occupation's concentration in two categories.

The key finding is the direct/indirect split: 54.8% of sector's contribution (-0.0068 of -0.0124) is indirect and inherited from field-of-study segregation, compared with only 19.9% for occupation. Sector therefore mostly reflects sorting that has already occurred through

educational choices, rather than a separate mechanism operating within the labour market. This is particularly clear in Healthcare & Social Services, where women's over-representation falls by 71% once field of study is controlled for (ρ_k : 0.082 $\rightarrow$ 0.024), and in (Petro)chemicals, where men's over-representation falls to almost zero (ρ_k : -0.016 $\rightarrow$ -0.0001). In both cases, the sector's raw imbalance is largely a downstream consequence of who studies what. This mediation is statistical rather than strictly causal, as it does not rule out an unobserved factor affecting both field and sector choices, which is further discussed in Section 7.

This inheritance also helps explain sector's relatively strong raw ρ_k/γ_k correlation (-0.310, Table 9), stronger than for field of study (-0.223) or occupation (-0.167), despite sector not being the most systematic sorting mechanism. It largely reflects field of study's own pattern being carried through to sectors. As with occupation, the direction of the effect depends on both ρ_k and γ_k : Telecom & ICT, the most male-dominated sector ($\rho_k = -0.085$), narrows rather than widens the gap because its wage premium net of other controls is slightly negative ($\gamma_k = -0.0121$). As with the other dimensions, sector's contribution is robust to the 2012 compositional skew (Section 6.7.).

Sector's contribution therefore lies between the other two dimensions: broader than occupation's concentration in a few cases, but much less independent of field of study. Much of the sectoral imbalance appears to be inherited from educational segregation rather than reflecting a separate sorting process at labour-market entry.

6.6. Cross-Dimension Synthesis

Table 10 draws together all four dimensions of the decomposition on a common basis: each dimension's raw contribution to the gender wage gap (Table 6a), its direct/indirect split with respect to field of study, and its ρ_k/γ_k correlation (Table 9).

Table 10: Direct and indirect contribution of working time, field of study, occupation, and sector to the gender wage gap

Dimension	δ_k (raw)	Direct	% direct	Indirect	% indirect	Correlation ρ_k/γ_k
Working time	-6.65e-02	-5.76e-02	86.6%	-8.91e-03	13.4%	— (continuous variable)
Field of study	-3.46e-02	-3.46e-02	—	-2.04e-02	—	-0.223
Occupation	-2.38e-02	-1.91e-02	80.1%	-4.73e-03	19.9%	-0.167
Sector	-1.24e-02	-5.64e-03	45.5%	-6.77e-03	54.8%	-0.310

One point in this table requires clarification. For field of study, the raw and direct contributions are identical (-0.0346, the same total already reported field by field in Table 8) because, unlike occupation and sector, it cannot receive an indirect contribution from itself. Therefore, no direct/indirect split or percentage breakdown is reported for this row. The indirect figure of -0.0204 should not be interpreted as a division of the -0.0346 into two parts. Instead, it represents the indirect contribution that field of study receives through occupation and sector. Adding this amount to the field of study's own contribution (-0.0346) gives its total contribution of -0.0550 (40%), in line with what is reported in Table 6b.

Working time follows a pattern closer to occupation than to sector: 86.6% of its raw contribution is direct, and only 13.4% (-0.00891) is inherited from field of study – likely reflecting gender norms around part-time work that differ somewhat across fields (Section 2.1.4.), though this channel is not decomposed field by field here (cf. Section 9). Therefore, out of the three dimensions besides field of study itself, working time is the least dependent on educational sorting.

The direct/indirect split shows a clear pattern across the three dimensions. **Occupation** is mostly direct (80.1% of -0.0238), suggesting that much of its contribution comes from sorting mechanisms that operate independently of educational choices, such as the barrier identified for company managers. **Sector** shows the opposite pattern: 54.8% of its contribution (-0.0068 of -0.0124) is indirect, indicating that it largely reflects field-of-study segregation rather than a separate sorting mechanism. **Field of study** is therefore the starting point for these inherited effects. Its own contribution (-0.0346) is entirely direct by construction, while the additional -0.0204 in the final grouped decomposition (Table 6b) corresponds exactly to the indirect contributions transferred from working time, occupation, and sector.

This pattern – working time and occupation largely direct, sector mostly inherited, and field of study as the common source of these inherited effects – is the central takeaway emerging from this section. It also helps explain an apparent anomaly in Table 10: sector has the strongest raw ρ_k/γ_k correlation (-0.310), ahead of field of study (-0.223) and occupation (-0.167), even though its sorting is not the most systematic. This does not indicate a strong sector-specific mechanism. Rather, because more than half of sector's contribution is inherited from field-of-study segregation, its raw correlation partly reflects the same sorting pattern already present in field of study.

Overall, these results support the main argument of this thesis: field of study (pre-labour-market component) plays a central role in shaping occupational and, even more strongly, sectoral segregation. Educational choices made before entering the labour market influence where men

and women eventually work and hence, their potential earnings. The next section tests whether these findings remain robust to alternative specifications and samples before discussing their broader implications.

6.7. Robustness Check

One concern with the results above is that occupation data are missing much more often in 2008 and 2010 than in 2012 (49.1% overall; Section 6.1.), meaning that the pooled estimation sample is disproportionately made up of 2012-observations. If this imbalance were driving the results, the contributions of field of study, occupation, and sector – all relying on occupation data, directly or indirectly – could reflect differences in survey coverage across waves rather than genuine labour-market patterns. Table 11 below tests this by re-estimating these three dimensions using only the 2012 subsample (N = 25,337), where occupation data are most complete, and comparing the results with the pooled sample.

Table 11: Robustness to the 2012 compositional skew

Dimension	Pooled (N = 48,785)	2012 only (N = 25,337)	Change
Field of study (total, direct + indirect)	-0.0550	-0.0599	+8.9%
Occupation (raw)	-0.0238	-0.0256	+7.6%
Sector (raw)	-0.0124	-0.0148	+19.4%

Restricting the sample to 2012 does not weaken any of the three results. If anything, each contribution increases slightly in absolute terms, while the larger relative change for sector reflects its smaller initial contribution rather than a large absolute difference. This confirms that the pooled estimates are not driven by the disproportionate share of 2012-observations in the sample.

A second check addresses the public-sector share of the sample discussed in Section 5.1. As already mentioned, public-sector wages are mainly determined by seniority and diploma level rather than field of study, which may reduce wage differences across fields. Therefore, including the 8.7% of public-sector respondents should, if anything, lead to a lower estimate of field of study's contribution. Table 12 below tests this by re-estimating the decomposition after excluding these observations. Both the direct and total contributions of field of study increase in absolute value when the public sector is excluded – from -0.0346 to -0.0357 for the direct contribution, a 3.2% increase in magnitude. This modest but consistent change is in line with the interpretation discussed in Section 5.1., namely that the inclusion of public-sector respondents would, if anything, lead this thesis to underestimate rather than overestimate the

true contribution of field of study, rather than suggesting that the pooled estimates overstate its importance.

Table 12: Robustness to the exclusion of the public sector

Quantity	Full sample (N = 48,785)	Private sector only (N = 44,540)	Change
Field of study, direct	-0.0346	-0.0357	+3.2%
Field of study, total (direct + indirect)	-0.0550	-0.0570	+3.6%

Together with the bootstrap results in Section 6.2., these two checks confirm that field of study is a robust and quantitatively important contributor to the gender wage gap among tertiary-educated workers in Belgium.

7. Limitations

Several limitations should be kept in mind when interpreting the results of this thesis. Some concern data quality issues already examined through the robustness checks presented above; others are more fundamental and cannot be addressed within the scope of this thesis.

7.1. Data quality limitations already addressed

Several data-quality concerns were identified earlier in this thesis and tested directly; none appear to affect the main conclusions. First, occupation data are missing for 49.1% of the sample, mainly in 2008 and 2010 (Section 4.3.), but restricting the analysis to 2012, when occupation data are most complete, does not weaken any of the three contributions, with each coefficient instead showing a slight increase (Table 11, Section 6.7.). Second, field of study is measured using only each respondent's first diploma and is slightly more often missing among women than men (9.48% vs. 7.52%; $\chi^2 = 130.04$, $p < 0.001$, Section 4.3.). Although this difference is statistically significant, it is small, and only 3.7% of respondents report a second diploma, with no meaningful gender difference (Section 4.2.). Third, some fields of study may have too few observations for their estimated contributions to be reliable. This does not appear to be a concern here: the smallest field, Veterinary Medicine, has 56 observations, and none of the 40 fields falls below the 30-observations threshold used elsewhere in this thesis. This contrasts with occupation, where 17 of 77 categories were excluded for having fewer than 30 observations (Section 6.4.). The bootstrap-based significance test added to Table 8 provides further support: out of 40 fields, only 8 have statistically insignificant contributions, and these

are generally smaller fields or fields with modest point estimates. None affects the main conclusions.

Overall, neither issue is likely to introduce systematic gender bias into the measurement of field of study, although some selection bias and imperfect match between first diploma and later specialisation for this small subgroup cannot be fully ruled out.

7.2. Fundamental Limitations

Several further limitations are more fundamental and cannot be addressed within the scope of this thesis. They fall into four categories.

❖ Data limitations

The data come from the *Vacature SalarisEnquête*, a large-scale voluntary online survey completed directly by employees, and covering the 2008-2012 period. As with any voluntary survey, participation is self-selected and may not be random with respect to gender, wage level, or characteristics studied here, such as field of study, occupation, and sector. If participation differs systematically by gender, this could affect both the estimated wage gap and its decomposition. This concern is partly reduced by the focus on tertiary-educated workers, who are closer to the segment of the labour market most likely to be reached by this type of online survey. In Belgium, tertiary-educated workers have stronger labour-market attachment and are less likely to work part-time than the general workforce (Eurostat, 2023), making them more comparable to the employed and digitally reachable respondents typically attracted by the online survey.

Also, the 2008-2012 period limits the extent to which the magnitude of the estimates can be generalized to the current labour market. This concern is partly mitigated by the survey's long-standing use in Belgian academic research and salary reporting (Delmotte et al., 2002; Delmotte et al., 2010; VersoDirect, 2012). Moreover, the patterns identified in this thesis reflect structural features of the Belgian education and labour market that are unlikely to change substantially over a short period of time.

Finally, the dataset does not include sampling weights (Section 4.1.). All descriptive statistics and regression results presented in this thesis are therefore unweighted and reflect the composition of survey respondents rather than a weighted estimate of the Belgian tertiary-educated workforce. If respondent characteristics are related to the likelihood of participating in the survey, unweighted estimates may not fully represent the broader tertiary-educated population.

❖ Identification limitations

The Gelbach decomposition is an accounting exercise: it identifies how much of the gender wage gap is statistically associated with each dimension, given the current wage structure. However, it does not establish a causal effect of educational choices, occupation, or sector on wages, i.e. on the gender wage gap. Relatedly, the decomposition is a partial equilibrium exercise: it assumes that the current wage structure remains unchanged. Therefore, the reported contributions – for instance, those of company managers and IT specialists (Section 6.4.) – show how the observed wage gap is currently distributed across occupations, rather than what would happen if the share of men and women in these occupations changed. Such changes could affect wages and sorting patterns elsewhere, which this framework does not capture.

Relatedly, this thesis focuses on the sorting mechanism rather than the devaluation mechanism (Section 3.1.): it does not test whether a field's wage premium falls as its share of women increases over time. Testing this would require tracking changes in both wages and gender composition across a longer period, which cannot be addressed with the four-years period covered by the dataset (2008-2012). Moreover, the model assumes that the return to age, working time, and each categorical control is the same for men and women. As discussed in Section 3.3.1., this assumption is not directly tested for every control. For field of study, however, the descriptive evidence shows no significant relationship between the within-field gender wage gap and a field's degree of feminization ($r = 0.079$, $p = 0.6284$; Table 4). Even if this is suggestive rather than a formal test, it indicates that the assumption is unlikely to be strongly violated for field of study. Whether the same applies specifically to occupation and sector is not examined here and remains an area for further research (Section 9). If these returns do differ by gender within occupation or sector categories, such differences would not be captured by the decomposition.

Finally, as with any decomposition based on observable characteristics, part of the residual 8.3% gap may reflect unobserved factors, such as negotiation behaviour, risk aversion or employer discrimination, rather than genuine differences in productivity-related characteristics that the dataset gives access to. This limitation instead defines how the decomposition should be interpreted: as evidence of robust statistical associations rather than proven causal channels.

❖ Measurement limitations

Field of study is measured at only one point in time, using each respondent's first diploma. As noted earlier, some workers may later change direction through further studies or career changes, which this thesis cannot capture. In addition, each dimension is measured at a relatively broad level – 40 fields of study, 77 occupations, and 22 sectors – meaning that differences within these categories are not observed. For instance, the category “Natural Sciences” pools together sub-fields such as biology, physics, and geology, which may differ substantially in their gender composition (ρ) and wage premiums (Γ), but this internal variation cannot be distinguished within the data used here. This limitation may make some gender differences harder to observe, as workers with potentially different characteristics are grouped into the same categories. They are therefore unlikely to explain the strong patterns found in the results, although some differences within categories may remain hidden.

Similarly, this thesis estimates field of study's indirect contribution only at the aggregate level, not field by field. The extent to which, say, Industrial Engineering's effect operates through occupation, sector, or working time specifically cannot be determined here (Section 9).

❖ Scope limitations

The sample is restricted to tertiary-educated workers in Belgium who are already employed. Therefore, it is hard to generalize the results to workers with lower education levels, to other countries, or to gender differences in labour market participation. This restriction nevertheless allows the analysis to focus more precisely on educational segregation within a relatively homogeneous population and institutional context, without mixing it with the different dynamics of lower-educated workers or cross-country institutional differences. The scope of the thesis should therefore be seen as a deliberate choice: it provides a more precise analysis of a clearly defined population rather than a broader but less focused comparison across very different groups.

Overall, all the cited limitations above mainly affect how the results should be interpreted and generalised rather than their internal consistency. The decomposition identifies statistical associations rather than causal effects, while the use of broad categories, first diploma, a voluntary survey, and a sample restricted to employed tertiary-educated workers limits the generalisability of findings. However, none of these limitations undermine the central conclusion that field of study is an important contributor to the gender wage gap among tertiary-educated workers in Belgium.

8. Discussion

This thesis investigates how much of the gender wage gap among tertiary-educated workers in Belgium between 2008 and 2012 can be explained by educational segregation, both directly and through its influence on working time, occupation, and sector. More specifically, it examines whether field of study is simply one channel among several or a key driver of the gap.

The results presented in Section 6 provide a clear answer. Once its indirect effects operating through occupation and sector are correctly reallocated to it, field of study (*dipl*) accounts for 40% of the gap explained by observable characteristics – nearly matching working time (41.9%) and considerably exceeding occupation (13.9%) and sector (4.1%). This finding challenges the emphasis of much of the empirical literature on the gender wage gap, which has traditionally identified working time and occupational segregation (labour-market determinants) as major observable sources of the gap (Bardasi & Gornick, 2008; Block, 2023; Matteazzi et al., 2018), while treating field of study (pre-labour-market determinant) more often as a background characteristic than an independent channel.

This emphasis in the literature may partly reflect how the different sources of the gap are typically measured: field of study's substantial indirect effect through occupation and sector (Section 6.2.) means that, if left unattributed, its true contribution is underestimated.

The decomposition of occupation's and sector's own contributions further nuances this picture and reveals that gender segregation does not operate through a single, uniform mechanism. Occupational segregation (Section 6.4.) is narrow and concentrated in just two occupations in this dataset, with opposite origins: one reflects a real occupation-specific barrier, largely unrelated to field of study, while the other is largely inherited from male-dominated fields of study. Sector segregation (Section 6.5.), by contrast, is more diffuse and largely inherited from field of study: no single sector in this dataset dominates the way the two main occupations do, and more than half of its contribution reflects segregation already present in field of study rather than sorting occurring when workers enter a sector.

Taken together, these findings suggest that the main contribution of this thesis is to show empirically that field of study is not simply a background control variable, but a quantitatively important dimension of the gender wage gap in its own right – one whose full contribution becomes visible only once the indirect effects it transmits through occupation and sector are distinguished from its direct effect. By separating what occupation and sector contribute independently from what they inherit from field of study, the analysis shows that a substantial

share of the segregation observed in the labour market is already associated with educational sorting that occurs before labour-market entry, rather than mainly with a small number of barriers or sorting mechanisms within particular jobs or sectors, i.e. the kind of labour-market barriers that typically receive the most policy attention in the literature on gender segregation and inequality (Bettio et al., 2009; Hegewisch, 2014). This shifts the focus of the gender wage gap towards the earlier educational choices that influence later labour-market outcomes.

These findings should nevertheless be interpreted with the caution already discussed in Section 7.: as an accounting exercise rather than a causal estimate. This does not, however, prevent a more exploratory question from being asked. The strength and consistency of the association documented here, combined with a growing experimental literature showing that field-of-study choices can be causally influenced (Li, 2018; Porter & Serra, 2020), makes it worth asking the following question that arises from these findings:

“If part of the gender wage gap originates in educational choices made around the age of 18, can it also be reduced through interventions at that stage?”

This is precisely the question this section explores, in light of the findings mentioned above. Unlike working time, which is shaped by a complex mix of household bargaining, childcare availability, and labour-market norms that are difficult to influence directly, field of study is chosen once, relatively early, and before labour-market entry. This makes it, in principle, more responsive to policy interventions. Field-of-study choices are particularly interesting from a policy perspective because they occur before many of the later sorting mechanisms identified in this thesis. Intervening at this stage could therefore potentially reduce the likelihood of subsequent segregation (occupational and sectoral), rather than attempting to correct it only after it has emerged, i.e. on the labour market.

This possibility is supported by a body of behavioural economics evidence on how such choices can be shaped, through what Thaler and Sunstein (2008) describe as “**nudges**”, i.e. interventions that change behaviour in a predictable way without forbidding any options or significantly changing economic incentives, and that are easy and cheap to avoid.

Several field experiments suggest that such interventions can meaningfully affect gender-imbalanced major choices. Li (2018) finds that a low-cost, targeted nudge – which was a message encouraging female students who performed above the class median to consider majoring in economics – was associated with a higher probability of choosing that major, suggesting that nudges of this kind can be a cost-effective tool for narrowing gender imbalances

in major choice. Porter and Serra (2020) find a considerably larger effect from exposing female students to successful, related role models already working in a male-dominated field. Among women who had initially planned to choose lower-earning fields, this exposure increased the probability of majoring in economics by 8 percentage points, suggesting that such interventions can influence both educational choices and hence future earnings.

However, not all interventions are equally effective. Owen (2023) shows that giving students information about their academic performance relative to their peers can reduce gender differences in *perceptions of ability* but has little impact on their actual choice of major. This suggests that information alone may not be sufficient to shift behaviour and that the design of an intervention matters as much as its content.

Overall, this literature combined with the results presented above suggests that the substantial contribution of field of study to the gender wage gap identified in this thesis is not necessarily an unavoidable feature of the labour market. Instead, part of it may be addressed through targeted, low-cost interventions (e.g. nudges) earlier in the education process (pre-labour-market), as long as they are well designed. Such interventions could complement labour-market policies that act mostly on the labour-market determinants, once educational and occupational segregation documented in this thesis has already emerged.

9. Next Steps

While this thesis provides a Belgian estimate of the direct and indirect contribution of field of study to the gender wage gap, several questions remain beyond its scope and could be explored in future research.

❖ Testing the effect of wage transparency reforms

A first area for future research concerns the EU Pay Transparency Directive (2023/970), which requires employers to disclose salary ranges and, for larger firms, gender pay gaps. By providing jobseekers with information about pay at little cost, the directive could potentially reduce the gender wage gap by helping women sort into higher-paying occupations and sectors and potentially reach a more gender-balanced labour market. However, evidence from Austria suggests that information alone may not be enough: Bamieh and Ziegler (2025) found no meaningful change in gender sorting into higher-paying jobs following mandatory wage disclosure. This highlights the importance of policy design, as discussed in the previous section in relation to Owen (2023). The effect of the directive cannot yet be studied in Belgium because it has not been fully transposed into Belgian law and no post-reform data are available. Moreover, once such data become available, identifying its causal effect would require a research design different from the decomposition approach used in this thesis.

❖ Extending external validity beyond tertiary-educated Belgian workforce

This thesis focuses on tertiary-educated workers in Belgium to keep educational attainment broadly comparable and isolate the role of field of study. Future research could extend the analysis to workers with lower levels of education or to other countries to determine whether the patterns identified here – particularly the similar contribution of field of study and working time to the gender wage gap – also hold in other contexts. Such extensions would require comparable data across countries or education levels, which is beyond the scope of this single-country thesis.

❖ Testing nudge-based interventions in the Belgian context

The discussion of Section 8 presents experimental evidence showing that low-cost interventions, such as exposure to role models and encouraging messages, can influence gender-imbalanced field-of-study choices (Li, 2018; Porter & Serra, 2020). However, these findings, coming from the US, may not apply in the same way to the Belgian education system, given differences in curricula and in when students choose their field of study. Testing this in Belgium would require a field experiment that follows students long enough to observe their study

choices and, ideally, their later labour-market outcomes. Such data collection is beyond the scope of this thesis, which relies on existing survey data.

❖ Testing the common wage-return assumption across dimensions

This thesis makes two related simplifying assumptions about wage returns (Section 3.1., Section 3.3.1.): first, that the wage return to a given field, occupation, or sector is the same for men and women; second, that this return does not change over time as a category's gender composition shifts (the devaluation mechanism). Neither assumption is fully tested here. For the first, only field of study was checked directly – the within-field gender wage gap shows no significant relationship with a field's degree of feminization (Table 4) – but occupation and sector were not examined specifically. Future research could apply the same test to these two dimensions. For the second, testing whether wages in a category fall as its share of women increases would require tracking both wages and gender composition over a longer period than the four years covered by this thesis's data (2008-2012), using a different approach from the cross-sectional decomposition applied here.

Relatedly, the absence of a significant relationship (Table 4) does not explain what drives the remaining differences across fields. Future research could investigate whether they reflect occupational sorting within fields, differences in career interruptions, or unobserved productivity differences.

❖ Decomposing field of study's indirect contribution field by field, including through working time

Finally, the indirect contribution of field of study is estimated only at the aggregate level (Section 6.6.), for occupation, sector and working time, whose contribution is split into a direct and an indirect component only in aggregate. Future research could extend this analysis field by field, examining, for each field of study, how much of its indirect effect operates through each of these three channels specifically. For instance, this could show whether the effect of Industrial Engineering operates mainly through occupational sorting, sector sorting, or differences in working time. Such an extension would require an additional layer of decomposition beyond the approach used in this thesis.

10. Conclusion

This thesis asked how much of the gender wage gap among tertiary-educated workers in Belgium can be linked to the field of study men and women choose before entering the labour market. Using the *Vacature SalarisEnquête* (2008–2012) and Gelbach's (2016) decomposition, the results show that field of study's total contribution accounts for 40% of the explained gender wage gap. This is almost as much as working time (41.9%) and considerably more than occupation (13.9%) or sector (4.1%).

The main contribution of this thesis is to show that educational choices are closely connected to where women and men end up working, i.e. their occupation and sector. Once the indirect effects of field of study through occupation, sector and even working time are taken into account, some interesting results emerge. More than half of the sector contribution and a substantial share of the occupation contribution can be traced back to educational sorting. Part of what appears to be labour-market segregation may therefore already be shaped before individuals enter the labour market.

These results should nevertheless be interpreted with care. The decomposition identifies statistical associations rather than causal effects, and the analysis is limited to employed tertiary-educated workers in Belgium between 2008 and 2012. However, the consistency of the findings across different specifications, samples, and fields of study suggests that the role of educational segregation is not simply a statistical coincidence.

These findings also raise an important question for future research: if part of the gender wage gap is shaped by educational choices, can it be reduced before individuals enter this market? Future research could examine whether interventions such as role models, targeted encouragement, or other forms of nudges can influence students' choices in Belgium, and study how the wage returns to fields of study change as their gender composition evolves over time (devaluation effect).

Overall, this thesis shows that understanding the gender wage gap requires looking beyond the workplace itself. Where men and women work matters, but so does where they started: the field of study they chose years before entering the labour market.

11. Bibliography

1. Alonso-Villar, O., & Del Río, C. (2023). Disentangling Occupational Sorting from Within-Occupation Disparities: Earnings Differences Among 12 Gender–Race/Ethnicity Groups in the U.S.. *Population Research and Policy Review*, 42, 1-28. <https://doi.org/10.1007/s11113-023-09791-1>
2. Arceo-Gómez, E. O., Campos-Vázquez, R. M., Badillo, R. Y., & Lopez-Araiza, S. (2022). Gender stereotypes in job advertisements: What do they imply for the gender salary gap?. *Journal of Labor Research*, 43, 65 - 102. <https://doi.org/10.1007/s12122-022-09331-4>
3. Bamieh, O., & Ziegler, L. (2025). Can wage transparency alleviate gender sorting in the labour market? *Economic Policy*, 40(122), 401–426. <https://doi.org/10.1093/epolic/eiae025>
4. Baraku, I., & Busetta, G. (2024). Statistical and taste-based discrimination in the labor market: an analysis of European Countries to identify optimal policy interventions. *Quality & Quantity*, 59, 621 - 638. <https://doi.org/10.1007/s11135-024-01944-4>
5. Bardasi, E., & Gornick, J. (2008). Working for less? Women's part-time wage penalties across countries. *Feminist Economics*, 14, 37 - 72. <https://doi.org/10.1080/13545700701716649>
6. Bar-Haim, E., Chauvel, L., Gornick, J., & Hartung, A. (2018). The Persistence of the Gender Earnings Gap: Cohort Trends and the Role of Education in Twelve Countries. *Social Indicators Research*, 165, 821-841. <https://doi.org/10.1007/s11205-022-03029-x>
7. Barron, K., Dittmann, R. K., Gehrig, S., & Schweighofer-Kodritsch, S. (2024). Explicit and Implicit Belief-Based Gender Discrimination: A Hiring Experiment. *Manag. Sci.*, 71, 1600-1622. <https://doi.org/10.2139/ssrn.4097858>
8. Becker, G. S. (1971). The Economics of Discrimination. <https://doi.org/10.7208/chicago/9780226041049.001.0001>
9. Bettio, F., Verashchagina, A., Mairhuber, I., Meulders, D., Beleva, I., Panayiotou, A., Křížková, A., & Emerek, R. (2009). Gender segregation in the labour market : root causes, implications and policy responses in the EU. <https://doi.org/10.2767/1063>

10. Blau, F. D., & Kahn, L. M. (2003). Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21(1), 106–144.
<https://doi.org/10.1086/344125>
11. Blau, F. D., & Kahn, L. M. (2006). The U.S. Gender Pay Gap in the 1990S: Slowing Convergence. *Industrial & Labor Relations Review*, 60, 45 - 66.
<https://doi.org/10.1177/001979390606000103>
12. Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *SSRN Electronic Journal*. <https://doi.org/10.1257/jel.20160995>
13. Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8, 436-455. <https://doi.org/10.2307/144855>
14. Block, P. (2023). Understanding the self-organization of occupational sex segregation with mobility networks. *Soc. Networks*, 73, 42-50.
<https://doi.org/10.1016/j.socnet.2022.12.004>
15. Boll, C., & Lagemann, A. (2019). The Gender Pay Gap in EU Countries — New Evidence Based on EU-SES 2014 Data. *Intereconomics*, 54, 101-105.
<https://doi.org/10.1007/s10272-019-0802-7>
16. Bonaccolto-Töpfer, M., Castagnetti, C., & Rosti, L. (2023). Changes in the gender pay gap over time: the case of West Germany. *Journal for Labour Market Research*, 57, 1-17. <https://doi.org/10.1186/s12651-023-00338-7>
17. Borrowman, M., & Klasen, S. (2020). Drivers of Gendered Sectoral and Occupational Segregation in Developing Countries. *Feminist Economics*, 26, 62 - 94.
<https://doi.org/10.1080/13545701.2019.1649708>
18. Brynin, M., & Perales, F. (2016). Gender Wage Inequality: The De-gendering of the Occupational Structure. *European Sociological Review*, 32, 162-174.
<https://doi.org/10.1093/esr/jcv092>
19. Buligescu, B., Borghans, L., & Fouarge, D. (2020). THE IMPACT OF OCCUPATIONAL SEGREGATION ON OCCUPATIONAL GENDER PAY GAP IN THE EUROPEAN UNION. *Journal of Community Positive Practices*.
<https://doi.org/10.35782/jcpp.2020.4.05>
20. Busch, F. (2018). Occupational Devaluation Due to Feminization?: Causal Mechanics, Effect Heterogeneity, and Evidence from the United States, 1960 to 2010. *Social Forces*, 96, 1351 - 1376. <https://doi.org/10.1093/sf/sox077>
21. Cech, E. (2016). Mechanism or Myth?. *Gender & Society*, 30, 265 - 288.
<https://doi.org/10.1177/0891243215608798>

22. Cha, Y., & Weeden, K. A. (2014). Overwork and the slow convergence in the gender gap in wages. In American Sociological Association, *American Sociological Review*.
<https://doi.org/10.1177/0003122414528936>
23. Cho, I. (2011). Do gender differences in risk preferences explain gender differences in labor supply, earnings or occupational choice?. *Research Papers in Economics*.
<https://ideas.repec.org/p/isu/genstf/201112020800001099.html>
24. Choi, K., Renna, F., & Choe, C. (2023). The impact of field of study on the gender wage gap: evidence from the first job offer out of college. *Applied Economics*, 56, 6713 - 6729. <https://doi.org/10.1080/00036846.2023.2276078>
25. Christofides, L., Polycarpou, A., & Vrachimis, K. (2013). Gender wage gaps, ‘sticky floors’ and ‘glass ceilings’ in Europe ☆. *Labour Economics*, 21, 86-102.
<https://doi.org/10.1016/j.labeco.2013.01.003>
26. Ciminelli, G., Schwellnus, C., & Stadler, B. (2021). Sticky floors or glass ceilings? The role of human capital, working time flexibility and discrimination in the gender wage gap. <https://doi.org/10.1787/02ef3235-en>
27. De Poli, S., & Maier, S. (2024). Enforcing ‘Equal Pay for Equal Work’ in the EU: what would it take?. *The Journal of Economic Inequality*, 23, 1219 - 1254.
<https://doi.org/10.1007/s10888-024-09654-5>
28. Delmotte, J., Sels, L., Faculteit Economie en Bedrijfswetenschappen, K.U.Leuven, Vandekerckhove, S., Vandenbrande, T., & HIVA - Onderzoeksinstituut voor Arbeid en Samenleving. (2010). De samenstelling van de loonkloof in België [WSE Report]. Steunpunt Werk en Sociale Economie.
https://www.steunpuntwerk.be/files/publications/WSE/WSE-Reports/wse-report_2010_07.pdf
29. Delmotte, J., Van Gyes, G., De Troyer, M., Hoger instituut voor de arbeid KATHOLIEKE UNIVERSITEIT LEUVEN, HIVA-K.U.Leuven, TEF-ULB, & Federale Diensten voor Wetenschappelijke, Technische en Culturele Aangelegenheden (DWTC). (2002). *Voorbereidende studie voor het oprichten van een observatorium voor het kaderpersoneel*.
https://www.belspo.be/belspo/organisation/publ/pub_ostc/agora/ragobs_nl.pdf
30. Deschacht, N. (2017). Part-Time Work and Women’s Careers: a Decomposition of the Gender Promotion Gap. *Journal of Labor Research*, 38, 169-186.
<https://doi.org/10.1007/s12122-017-9242-y>

31. Edelsztein, V., & Waisbrot, S. A. (2023). Breaking down the Gender Pay Gap through a machine learning model. *Convergencia Revista de Ciencias Sociales*.
<https://doi.org/10.29101/crcs.v30i0.20656>
32. Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC.
33. European Commission (2024). *The gender pay gap situation in the EU*. (n.d.).
https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en
34. Eurostat. (2023). *Part-time and full-time employment – statistics* [Statistics Explained]. European Commission. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Part-time_and_full-time_employment_-_statistics
35. Fan, X., & Sturman, M. (2019). Has Higher Education Solved the Problem? Examining the Gender Wage Gap of Recent College Graduates Entering the Workplace. *Compensation & Benefits Review*, 51, 12 - 5.
<https://doi.org/10.1177/0886368719856268>
36. Finger, C., Solga, H., Ehlert, M., & Rusconi, A. (2020). Gender differences in the choice of field of study and the relevance of income information. Insights from a field experiment. *Research in Social Stratification and Mobility*, 65, 100457.
<https://doi.org/10.1016/j.rssm.2019.100457>
37. Fontenay, S., & Tojerow, I. (2024). Sickness absences among young mothers and the child penalty in employment. *Review of Economics of the Household*, 23, 625 - 660.
<https://doi.org/10.1007/s11150-024-09720-x>
38. Fontenay, S., Murphy, T., & Tojerow, I. (2021). Child penalties across industries: why job characteristics matter. *Applied Economics Letters*, 30, 488 - 495.
<https://doi.org/10.1080/13504851.2021.1994518>
39. Francesconi, M., & Parey, M. (2018). Early gender gaps among university graduates. *European Economic Review*.
<https://doi.org/10.1016/j.euroecorev.2018.02.004>
40. Galos, D., & Coppock, A. (2023). Gender composition predicts gender bias: A meta-reanalysis of hiring discrimination audit experiments. *Science Advances*, 9.
<https://doi.org/10.1126/sciadv.ade7979>
41. Galos, D., & Kulic, N. (2022). Variations of Gender Gaps in the Labour Market Outcomes of Graduates across Fields of Study: A (Combined) Test of Two Theories. *Sociology*, 57, 882 - 903. <https://doi.org/10.1177/00380385221122400>

42. Gangl, M., & Ziefle, A. (2009). Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States. *Demography*, 46, 341-369.
<https://doi.org/10.1353/dem.0.0056>
43. Garner, A., Kampelmann, S., & Rycx, F. (2014). Part-Time Work, Wages, and Productivity. *Industrial & Labor Relations Review*, 67, 926 - 954.
<https://doi.org/10.1177/0019793914537456>
44. Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much?. *Journal of Labor Economics*. <https://doi.org/10.1086/683170>
45. Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *The American Economic Review*, 104, 1091-1119. <https://doi.org/10.1257/aer.104.4.1091>
46. Hägglund, A. (2024). Same degrees, different outcomes? Fields of study choices and gender wage inequality in Finland and Germany.. *Social science research*, 122, 103029 . <https://doi.org/10.1016/j.ssresearch.2024.103029>
47. Hardies, K., Lennox, C., & Li, B. (2020). Gender Discrimination? Evidence from the Belgian Public Accounting Profession*. *Contemporary Accounting Research*.
<https://doi.org/10.1111/1911-3846.12667>
48. Harman, J. (2023). Gender Pay Gap in the Slovak Labour Market. *Ekonomické rozhľady – Economic Review*. <https://doi.org/10.53465/er.2644-7185.2023.1.29-47>
49. Hegewisch, A., & Hartmann, H. (2014). Occupational Segregation and the Gender Wage Gap: A Job Half Done.
<https://ecommons.cornell.edu/entities/publication/ad9dba41-41a5-4864-9134-d975d8ef3e5a>
50. Jagannathan, R., & Camasso, M. (2025). Gendered Wage Penalties for the Overeducated: The Experiences of Young Men and Women in Ten European Countries. *Comparative Economic Research. Central and Eastern Europe*.
<https://doi.org/10.18778/1508-2008.28.30>
51. Jepsen, M., O'dorchai, S., Plasman, R., & Rycx, F. (2005). The wage penalty induced by part-time work: the case of Belgium. *Brussels economic review*, 48, 73-94.
<https://ideas.repec.org/a/bxr/bxrceb/y2005v48i1-2p73-94.html>
52. Kampelmann, S., Rycx, F., Saks, Y., & Tojerow, I. (2018). Does education raise productivity and wages equally? The moderating role of age and gender. *IZA Journal of Labor Economics*, 7, 1-37. <https://doi.org/10.1186/s40172-017-0061-4>

53. Kleven, H., Landais, C., & Leite-Mariante, G. (2023). The Child Penalty Atlas. *SSRN Electronic Journal*. <https://doi.org/10.3386/w31649>
54. Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). Child Penalties Across Countries: Evidence and Explanations. *Labor: Demographics & Economics of the Family eJournal*. <https://doi.org/10.1257/pandp.20191078>
55. Kolesnikova, N. A., & Liu, Y. (2011). The gender wage gap. *Economic Synopses, 2011*. <https://doi.org/10.20955/es.2011.25>
56. Leuze, K., & Strauss, S. (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 & Society, 30*, 802 - 820. <https://doi.org/10.1177/0950017015624402>
57. Li, H.-H. (2018). Do mentoring, information, and nudge reduce the gender gap in economics majors?. *Economics of Education Review*. <https://doi.org/10.1016/j.econedurev.2018.04.004>
58. Li, Y. (2025). Who Chooses? Exploring Family, Gender, and Aspirations in Chinese High School Subject Selection. *Lecture Notes in Education Psychology and Public Media*. <https://doi.org/10.54254/2753-7048/2025.nd26189>
59. Martins, D.B., Vandenberghe, V. (2010). Using Firm-Level data to assess gender wage discrimination in the Belgian labour market. *ideas.repec.org*. <https://ideas.repec.org/p/ctl/louvir/2010007.html>
60. Matteazzi, E., Pailhé, A., & Solaz, A. (2018). Part-time employment, the gender wage gap and the role of wage-setting institutions: Evidence from 11 European countries. *European Journal of Industrial Relations, 24*, 221 - 241. <https://doi.org/10.1177/0959680117738857>
61. Mavrikiou, P. M., & Angelovska, J. (2020). The Impact Of Sex Segregation By Economic Activity On The Gender Pay Gap Across Europe. *UTMS Journal of Economics, 11*, 1-12. <https://www.econstor.eu/bitstream/10419/281867/1/1735913650.pdf>
62. Mincer, J. & Ofek, H. (1980). Interrupted Work Careers. *National Bureau of Economic Research*. <https://doi.org/10.3386/w0479>
63. Morgan, S., Gelbgiser, D., & Weeden, K. A. (2013). Feeding the pipeline: Gender, occupational plans, and college major selection.. *Social science research, 42* 4, 989-1005 . <https://doi.org/10.1016/j.ssresearch.2013.03.008>

64. Murphy, E., & Oesch, D. (2016). The Feminization of Occupations and Change in Wages: A Panel Analysis of Britain, Germany, and Switzerland. *Social Forces*, 94, 1221 - 1255. <https://doi.org/10.1093/sf/sov099>
65. Naszodi, A. (2021). *Decomposition scheme matters more than you may think* (arXiv:2104.09141). arXiv. <https://arxiv.org/abs/2104.09141>
66. Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14, 693-709. <https://doi.org/10.2307/2525981>
67. OECD (2021). Why do more young women than men go on to tertiary education?. *Education Indicators in Focus*. <https://doi.org/10.1787/6f7209d1-en>
68. Olivetti, C., & Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *SSRN Electronic Journal*. <https://doi.org/10.1146/annurev-economics-080614-115329>
69. Owen, S. (2023). College major choice and beliefs about relative performance: An experimental intervention to understand gender gaps in STEM. *Economics of Education Review*. <https://doi.org/10.1016/j.econedurev.2023.102479>
70. Parada-Contzen, M., & Jara, F. (2025). Gender wage gap among the educated: evidence from fields of study in Chile. *Humanities and Social Sciences Communications*, 12. <https://doi.org/10.1057/s41599-025-05312-x>
71. Perales, F. (2013). Occupational sex-segregation, specialized human capital and wages: evidence from Britain. *Work, Employment & Society*, 27, 600 - 620. <https://doi.org/10.1177/0950017012460305>
72. Petrenko, O., Błażejowski, M., Kwiatkowski, J., & Čadil, J. (2026). Do more women majoring in STEM subjects bridge the gender pay gap in the EU? Implications for building a competitive, socially sustainable economy. *Equilibrium. Quarterly Journal of Economics and Economic Policy*. <https://doi.org/10.24136/eq.3539>
73. Petrongolo, B., & Ronchi, M. (2020). Gender gaps and the structure of local labor markets. *Labour Economics*. <https://doi.org/10.1016/j.labeco.2020.101819>
74. Porter, C., & Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*. <https://doi.org/10.1257/app.20180426>
75. Rozhkova, K., Roshchin, S., & Yemelina, N. (2024). The dynamics of the early career gender wage gap among university graduates: The case of Russia. *European Journal of Education*. <https://doi.org/10.1111/ejed.12730>

76. Ruijter, J., Doorne-Huiskes, A., & Schippers, J. (2003). Size and Causes of the Occupational Gender Wage-gap in the Netherlands. *European Sociological Review*, 19, 345-360. <https://doi.org/10.1093/esr/19.4.345>
77. Sánchez-Mangas, R., & Sánchez-Marcos, V. (2021). Wage Growth across Fields of Study among Young College Graduates: Is There a Gender Gap?. *CESifo Economic Studies*. <https://doi.org/10.1093/cesifo/ifaa021>
78. Scarborough, W. J., & Moeder, J. (2022). Culture's Gendered Consequences: The Relationship Between Local Cultural Conditions and the Gender Wage Gap. *Social Currents*, 9, 526 - 557. <https://doi.org/10.1177/23294965211045088>
79. Seehuus, S., & Strømme, T. (2024). Gendered Returns to Education: The Association between Educational Attainment, Gender Composition in Field of Study and Income. *Sociology*, 59, 503 - 523. <https://doi.org/10.1177/00380385241303448>
80. Sloane, C., Hurst, E. G., & Black, D. (2021). College Majors, Occupations, and the Gender Wage Gap. *Journal of Economic Perspectives*. <https://doi.org/10.1257/jep.35.4.223>
81. Smyk, M., Tyrowicz, J., & Van Der Velde, L. A. (2018). A Cautionary Note on the Reliability of the Online Survey Data: The Case of Wage Indicator. *Sociological Methods & Research*, 50, 429 - 464. <https://doi.org/10.1177/0049124118782538>
82. Statbel. (2026, 26 mars). *Légère progression des jeunes titulaires d'un diplôme de l'enseignement supérieur*. <https://statbel.fgov.be/fr/themes/emploi-formation/formation-et-enseignement/niveau-dinstruction>
83. Strawiński, P., Majchrowska, A., & Broniatowska, P. (2018). Occupational segregation and wage differences: the case of Poland. *International Journal of Manpower*. <https://doi.org/10.1108/ijm-07-2016-0141>
84. Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press (Book).
85. Tverdostup, M., & Paas, T. (2022). Gender disparities in wage returns to human capital components: how different are European labour markets?. *Baltic Journal of Economics*, 22, 28 - 48. <https://doi.org/10.1080/1406099x.2022.2033418>
86. Vandenberghe, V. (2011). Firm-level Evidence on Gender Wage Discrimination in the Belgian Private Economy. *Labour*, 25, 330-349. <https://doi.org/10.1111/j.1467-9914.2011.00524.x>

87. VersoDirect. (2012). VersoDirect. In *VersoDirect*.
https://docs.vlaamsparlement.be/docs/biblio/opendigibib/tijdschrift/versodirect/versodirect2012_4.pdf
88. Warner, M., & Clair, M. (2026). Gendered expectations of future work-family conflict in the legal profession. *American Journal of Cultural Sociology*.
<https://doi.org/10.1057/s41290-026-00287-x>
89. Weber, T. (2015). *Belgium: Reporting the remuneration of national civil servants for Art65 purposes*. Eurostat.
https://ec.europa.eu/eurostat/documents/6939681/9010625/Belgium+-+Civil+Service+Remuneration_1.pdf/7b13c8e7-9d57-45f4-aaa2-0a8c5d26a96b
90. Zając, T., Magda, I., Bożykowski, M., Chłoń-Domińczak, A., & Jasiński, M. (2024). Gender pay gaps across STEM fields of study. *Studies in Higher Education*, 50, 126 - 139. <https://doi.org/10.1080/03075079.2024.2330667>
91. Zheng, H., & Weeden, K. A. (2023). How Gender Segregation in Higher Education Contributes to Gender Segregation in the U.S. Labor Market.. *Demography*.
<https://doi.org/10.1215/00703370-10653728>

12. Appendices

❖ Appendix 1

Table A1: Top 10 female- and male-dominated occupations

Female-dominated occupations			Male-dominated occupations		
Occupation	% women	% of fem. LF	Occupation	% men	% of male LF
Nursery school teacher	100	0.43	Const. carpenter	100	1.52
Medical receptionist	99	1.19	Road worker	100	0.56
Homecare worker	98	0.90	Bricklayer	99	0.91
Secretary	98	3.58	Sanitary installer	99	0.98
Homecare assistant	98	0.87	Shutter carpenter	99	1.06
Pharmaceutical assistant	97	0.65	Const. works foreman	99	0.43
Medical secretary	97	0.86	Road engineer	99	0.43
Geriatric helper	96	1.19	Welder	99	0.38
Orderly	95	1.92	Mech. eng. foreman	99	0.52
Beautician	94	0.81	Electrician	99	1.02
Total		12.40	Total		7.82

Source: Authors' calculations based on the Structure of Earnings Survey 1997.

Source: Ruijter, J., Doorne-Huiskes, A., & Schippers, J. (2003). *Size and Causes of the Occupational Gender Wage-gap in the Netherlands*. *European Sociological Review*, 19, 345-360. <https://doi.org/10.1093/esr/19.4.345>

❖ Appendix 2

Table A2: Sample size by field of study, broken down by Bachelor and Master degrees

Degree	Field of study	N
Bachelor	b-Business assistant (bookkeeping, taxation, marketing, insurance)/Business management (accountancy, taxation, marketing, legal practice, finance and insurance, communication management, logistics management)	14,409
	b-Technician (audio, photography, electricity)/Industrial sciences (audiovisual technology, photography, graphic arts, (bio)chemistry, electricity, electromechanics, telecommunications, imaging, textiles)	5,786
	b-Secretary/Secretarial studies-Languages	5,414
	b-Computing_IT(A1)/Computer science (A1)	4,530
	b-Social assistant/Social work (social assistant, psychological assistant, special education)	3,122
	b-Teaching(lower secondary)/Teacher training for lower secondary education	2,243
	b-Nurse, midwife/Nursing, midwifery	1,880
	b-Medical assistant/Health care (laboratory technician, speech therapy, physiotherapy, occupational therapy, nutrition, dietetics)	1,348
	b-Teaching(primary, kindergarten)/Kindergarten and primary school teacher	1,200
	b-Tourism, entertainment/Tourism and recreation	1,009
	b-Media, publicity/Media, advertising	777
	b-Construction, wood/Construction engineering, wood technology	488
	b-Accommodation, restaurant/Hospitality and hotel management	451

Degree	Field of study	N
Master	b-Architecture/Architecture	414
	m-Industrial engineers/Industrial engineering, surveying	9,719
	m-Accountant/Sales engineer/Public administration, business engineering, business sciences	5,076
	m-Business/Applied economics	4,902
	m-Social Science/Social sciences (sociology, political science, communication, public administration)	4,581
	m-Civil Engineer/Civil engineering, architectural engineering	3,979
	m-Sales engineer/Business engineering, business engineering in information management	2,971
	m-Law/Law	2,684
	m-Languages, Philology/Languages, philology	2,672
	m-Natural sciences (biology, math, chemistry, physics, geology, geography)/Natural sciences (biology, mathematics, chemistry, physics, geology, geography)	2,311
	m-Economics/Economics	2,002
	m-Psychology/Psychology, education sciences	1,995
	m-Journalist, translator/Journalism, interpretation, translation	1,921
	m-Bio Engineer, agricultural engineer/Bio-engineering, agricultural engineering	1,701
	m-Fine arts (theater, design, music, visual arts)/Arts (drama, music, design, visual arts)	1,145
	m-History, History of art, Musicology, Archeology/History, art history, musicology, archaeology	1,127
	m-Computing science/Computer science	1,057
	m-Other	792
	m-Criminology/Criminology	750
	m-Physical education, physiotherapy, speech therapy/Physical education, physiotherapy, speech therapy	592
	m-Interior architecture/Interior architecture	522
	m-Pharmacists/Pharmacy	503
	m-Biomedical sciences/Biomedical sciences	465
	m-Medico-social sciences (nurses)/Hospital sciences	385
	m-Medical studies (Physicians, Dentists)/Medicine (physicians, dentists)	299
	m-Philosophy, Theology/Philosophy, theology	252
	m-Veterinarians/Veterinary medicine	137
Bachelor	Bachelor - Total	43,071
Master	Master - Total	54,540
All	Total	97,611

❖ Appendix 3

Table A3: Sample construction and missing data

Step	N	Dropped	Dropped (%)
<i>Raw dataset (all education levels, gender non-missing)</i>	158,577	—	—
<i>Restricted to tertiary educated (Bachelor/Master) = 'work'</i>	106,520	52,057	32.8
+ <i>non-missing gwage</i>	105,194	1,326	1.24
+ <i>non-missing wkpc</i>	105,194	0	0.00
+ <i>non-missing dipl (field of study)</i>	96,440	8,754	8.32
+ <i>non-missing occh (occupation)</i>	48,799	47,641	49.4
+ <i>non-missing secth (sector)</i>	48,799	0	0.00
+ non-missing age	48,785	14	0.03

Notes: Each row applies an additional restriction, on top of all restrictions listed above it, to the tertiary-educated sample (work). The final sample (N = 48,785) corresponds to the estimation sample used in the Gelbach decomposition (work_decomp).

❖ Appendix 4

Table A4: Missing values on field of study (dipl), by gender

Gender	N_{total}	$N_{missingDipl}$	% _{missing}
Men	60,816	4,576	7.52
Women	45,704	4,333	9.48

Notes: Sample is the tertiary-educated working sample (work, N = 106,520). A Pearson's chi-squared test with Yates' continuity correction indicates that missingness on field of study is significantly associated with gender ($\chi^2 = 130.04$, $df = 1$, $p < 0.001$).

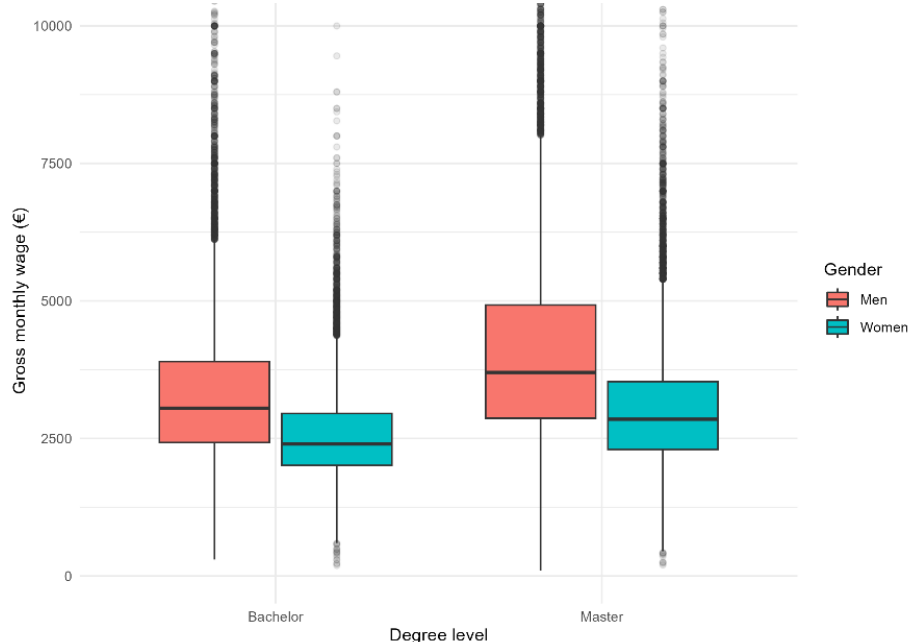
Table A5: Missing values on occupation (occh), by gender

Gender	N_{total}	$N_{missingOcch}$	Pct _{missing}
Men	60,816	31,565	51.90
Women	45,704	20,757	45.42

Notes: Sample is the tertiary-educated working sample (work, N = 106,520). A Pearson's chi-squared test with Yates' continuity correction indicates that missingness on occupation is significantly associated with gender ($\chi^2 = 439.01$, $df = 1$, $p < 0.001$).

❖ Appendix 5

Figure A1: Distribution of gross monthly wages by gender and degree level



❖ Appendix 6

Table A6: Top 10 most feminized and most masculinized fields

Category	Field of study	N	Share of women (%)	Mean wage, FTE-equivalent (€)	Relative gap (%)
Most feminized	b-Secretary/Secretarial studies-Languages	5,339	84.6	2,873	17.5
	m-Journalist, translator/Journalism, interpretation, translation	1,896	77.0	3,060	18.2
	b-Tourism, entertainment/Tourism and recreation	1,000	76.8	2,639	14.4
	m-Biomedical sciences/Biomedical sciences	457	73.5	3,347	16.5
	m-Languages, Philology/Languages, philology	2,637	71.2	3,352	23.3
	m-Psychology/Psychology, education sciences	1,969	70.1	3,301	19.2
	b-Medical assistant/Health care (laboratory technician, speech therapy, physiotherapy, occupational therapy, nutrition, dietetics)	1,324	67.9	2,998	17.9
	b-Social assistant/Social work (social assistant, psychological assistant, special education)	3,060	67.3	2,817	17.9
	m-Pharmacists/Pharmacy	496	65.3	4,364	20.5
	b-Nurse, midwife/Nursing, midwifery	1,798	64.9	3,407	15.7
Most masculinized	m-Computing science/Computer science	1,046	8.8	3,982	9.1
	b-Technician (audio, photography, electricity)/Industrial sciences (audiovisual technology, photography, graphic arts, (bio)chemistry, electricity, electromechanics, telecommunications, imaging, textiles)	5,720	10.9	3,379	18.2

Category	Field of study	N	Share of women (%)	Mean wage, FTE-equivalent (€)	Relative gap (%)
	b-Construction, wood/Construction engineering, wood technology	479	11.5	3,184	14.5
	b-Computing_IT(A1)/Computer science (A1)	4,498	12.1	3,334	11.2
	m-Industrial engineers/Industrial engineering, surveying	9,670	12.4	3,956	17.4
	m-Civil Engineer/Civil engineering, architectural engineering	3,950	12.9	4,655	21.0
	m-Sales engineer/Business engineering, business engineering in information management	2,957	30.7	4,240	21.1
	m-Business/Applied economics	4,852	34.9	4,166	22.6
	b-Accommodation, restaurant/Hospitality and hotel management	444	35.6	3,055	23.1
	m-Economics/Economics	1,981	36.1	4,031	23.1

❖ Appendix 7

Table A7: Wage comparison across feminized and masculinized fields, by matched degree level

Degree	Composition	Field of study	N	Share of women (%)	Mean wage - Men, FTE-equivalent (€)	Mean wage - Women, FTE-equivalent (€)	Relative gap (%)
Bachelor	Feminized	Nurse, midwife/Nursing, midwifery	1,798	64.9	3,794	3,199	15.7
	Masculinized	Computing_IT(A1)/Computer science (A1)	4,498	12.1	3,380	3,003	11.2
Master	Feminized	Psychology/Psychology, education sciences	1,969	70.1	3,814	3,082	19.2
	Feminized	Pharmacists/Pharmacy	496	65.3	5,037	4,007	20.5
	Masculinized	Civil Engineer/Civil engineering, architectural engineering	3,950	12.9	4,785	3,780	21.0
	Masculinized	Computing science/Computer science	1,046	8.8	4,014	3,650	9.1

❖ Appendix 8

Table A8: Dissimilarity index by field of study, occupation, and sector

Dimension	Dissimilarity index
Field of study	37.41
Occupation	37.36
Sector	21.88

❖ Appendix 9

Table A9: Robustness check — age vs. experience as control for career stage

Check	Value
Sample size (main spec, age only)	48,785
Sample size (with exp added)	37,602
R ² of exp explained by age alone	0.9
Coef. on fem — without exp	-0.0828
Coef. on fem — with exp	-0.0839
Coef. on exp	0.0199
Share missing exp — men (%)	26.7
Share missing exp — women (%)	18.1

❖ Appendix 10

Table A10: Missing values by variable, tertiary sample (2008–2012)

Variable	Missing (n)	Missing (% of tertiary sample)
Wage (gwage/lngwage)	1,326	1.2
Working time (wkpc)	10	0.0
Field of study (dipl)	8,909	8.4
Occupation (occh)	52,322	49.1
Sector (secth)	0	0.0

❖ Appendix 11

Table A11: Prevalence of a second recorded diploma and consistency with the field-of-study variable

Check	Value
Share with 2nd diploma recorded — overall (%)	3.70
Share with 2nd diploma recorded — men (%)	3.61
Share with 2nd diploma recorded — women (%)	3.82
N with 2nd diploma, restricted to tertiary sample (educat 5/6, yr 2008-2012)	5,864
Match rate: dipl = dipl_1 (%)	100.0
Match rate: dipl = dipl_2, among those with a 2nd diploma (%)	0.0

❖ Appendix 12

Table A12a: Omitted (reference) category for each categorical control

Variable	Reference category (code)	Label	N in reference category	Rank by size (out of total categories)
Field of study (dipl)	101	Secretary/Secretarial studies-Languages	2,934	38th / 40
Occupation (occh)	1	Administrative staff	4,967	Largest / 91
Sector (secth)	30	Agriculture & Food Industry	1,374	7th smallest / 22

Table A12b: Invariance check — releveing the sector reference category

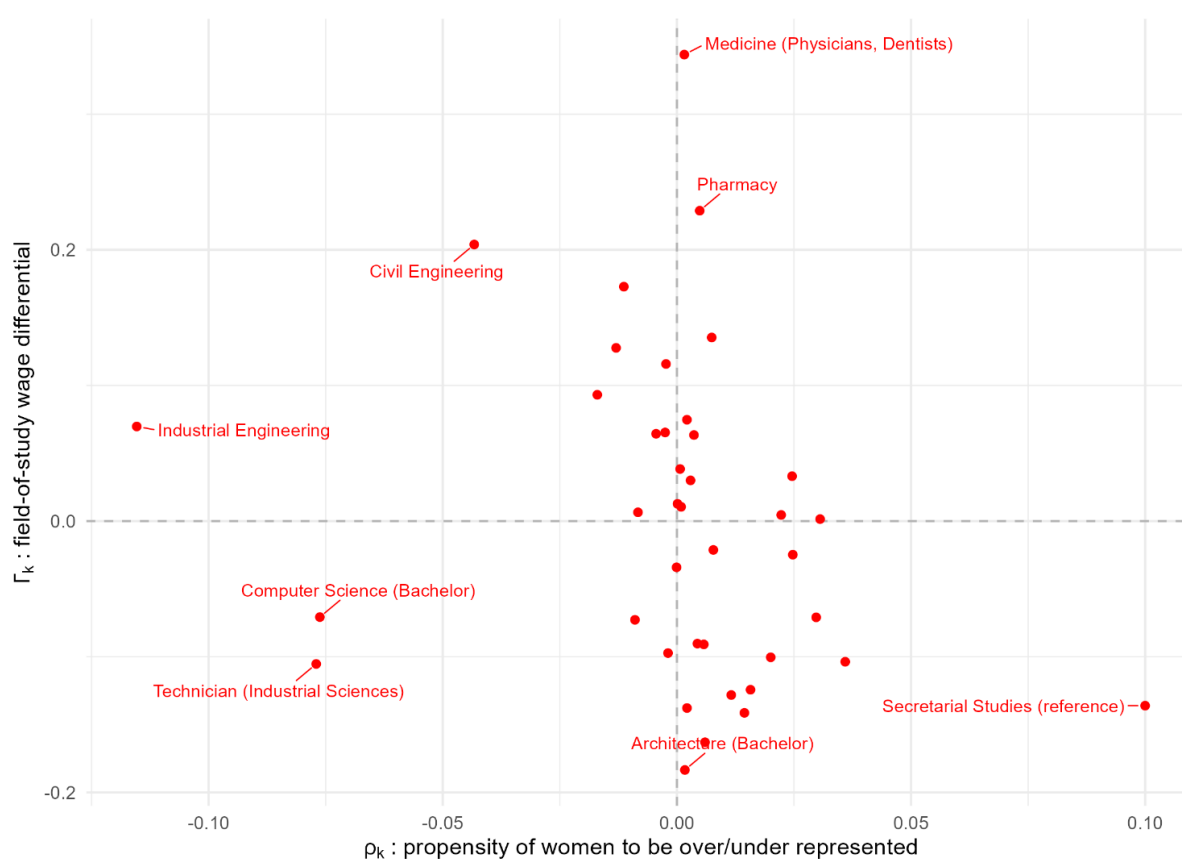
Reference	Coef. on fem, full model	δ_{secth} (Gelbach contribution)
Agriculture & Food Industry (original)	-0.0865	-0.01241
Healthcare & Social Services (largest category, releve)	-0.0865	-0.01241

❖ Appendix 13

Table A13: Fields most extreme in gender representation or wage premium (values centered on the sample average)

Field of study	N	ρ_{kc}	γ_{kc}
m-Industrial engineers/Industrial engineering, surveying	4,385	-0.11535	0.06977
b-Secretary/Secretarial studies-Languages	2,934	0.09999	-0.13607
b-Technician (audio, photography, electricity)/Industrial sciences (audiovisual technology, photography, graphic arts, (bio)chemistry, electricity, electromechanics, telecommunications, imaging, textiles)	2,834	-0.07703	-0.10530
b-Computing_IT(A1)/Computer science (A1)	2,705	-0.07624	-0.07074
m-Civil Engineer/Civil engineering, architectural engineering	1,647	-0.04328	0.20392
m-Pharmacists/Pharmacy	259	0.00487	0.22883
b-Architecture/Architecture	230	0.00171	-0.18345
m-Medical studies (Physicians, Dentists)/Medicine (physicians, dentists)	175	0.00158	0.34390

Figure A2: Gender representation (ρ_k) versus wage premium (γ_k), by field of study



❖ Appendix 14

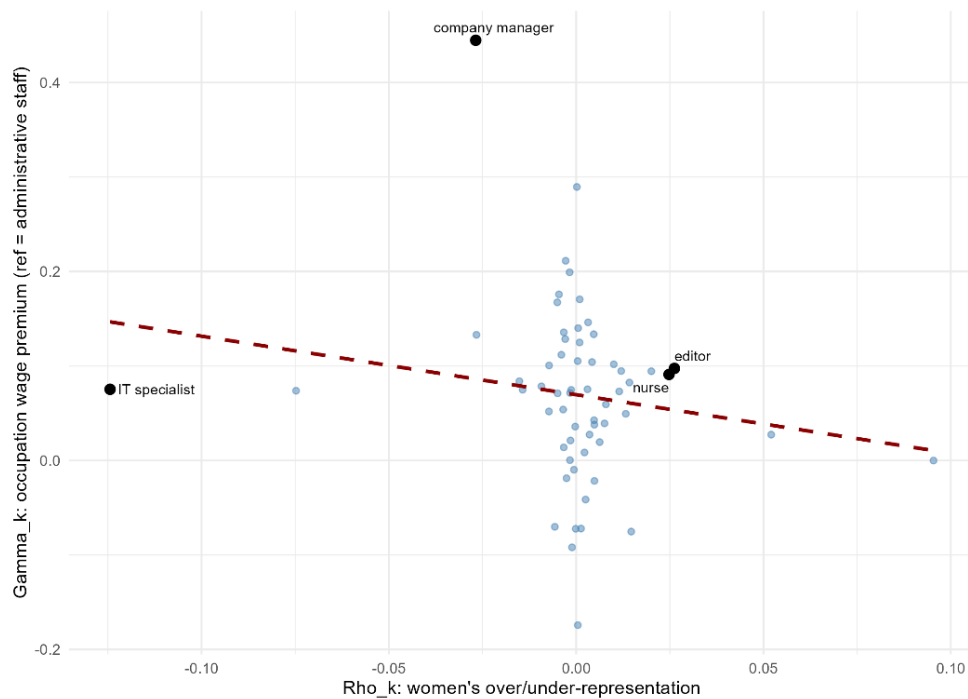
Table A14: Direct contribution of each occupation to the gender wage gap

Occupation	N	ρ_k (raw)	γ_k	δ_k (total)	ρ_k (net of dipl)	γ_k (direct)	γ_k (indirect via dipl)
company manager	1,441	-0.02681	0.4448	-1.19e-02	-0.02501	-1.11e-02	-8.01e-04
IT specialist	4,460	-0.12441	0.0754	-9.38e-03	-0.04891	-3.69e-03	-5.69e-03
technician	2,658	-0.07477	0.0739	-5.52e-03	-0.01899	-1.40e-03	-4.12e-03
project manager	1,203	-0.02663	0.1331	-3.54e-03	-0.0087	-1.16e-03	-2.39e-03
maintenance staff	430	-0.01516	0.084	-1.27e-03	-0.00539	-4.53e-04	-8.21e-04
receptionist	545	0.01469	-0.0752	-1.10e-03	0.00942	-7.08e-04	-3.97e-04
sales representative	2,674	-0.01429	0.0749	-1.07e-03	-0.02712	-2.03e-03	9.61e-04
office manager	258	-0.00505	0.1674	-8.45e-04	-0.00412	-6.89e-04	-1.56e-04
architect	384	-0.00459	0.1758	-8.07e-04	-0.00272	-4.79e-04	-3.29e-04
accountant	4,870	-0.00927	0.0785	-7.28e-04	-0.00274	-2.16e-04	-5.13e-04
craftsman/tradesperson	340	-0.00723	0.1007	-7.28e-04	-0.00327	-3.29e-04	-3.99e-04
military personnel	84	-0.0028	0.2114	-5.92e-04	-0.00358	-7.56e-04	1.64e-04
police officer	185	-0.00331	0.1357	-4.50e-04	-0.0057	-7.73e-04	3.23e-04
buyer	901	-0.00395	0.112	-4.43e-04	-0.00092	-1.03e-04	-3.39e-04
site manager	221	-0.00727	0.0519	-3.77e-04	-0.00239	-1.24e-04	-2.53e-04
security guard	106	-0.00292	0.1286	-3.76e-04	-0.00195	-2.51e-04	-1.25e-04
prevention advisor	312	-0.00496	0.0711	-3.53e-04	-0.00113	-8.05e-05	-2.72e-04
manager	4,123	-0.00176	0.1992	-3.50e-04	-0.00594	-1.18e-03	8.33e-04
dispatcher	770	-0.00348	0.054	-1.88e-04	-0.00301	-1.62e-04	-2.54e-05
electrician	42	-0.00163	0.0716	-1.17e-04	-0.00075	-5.34e-05	-6.34e-05
traffic controller *	3	-0.00013	0.9027	-1.14e-04	—	—	—
travel agent	157	0.00487	-0.0215	-1.05e-04	0.00146	-3.13e-05	-7.35e-05
temporary employment consultant	72	0.00252	-0.0413	-1.04e-04	0.00206	-8.52e-05	-1.86e-05
production worker	80	-0.00133	0.0748	-9.97e-05	-0.00017	-1.28e-05	-8.69e-05
hospitality staff	154	0.00129	-0.072	-9.30e-05	0.00028	-2.03e-05	-7.27e-05
cashier *	24	0.00037	-0.2315	-8.64e-05	—	—	—
animator	41	0.00044	-0.1742	-7.59e-05	1e-05	-1.99e-06	-7.39e-05
firefighter *	7	-0.00027	0.1831	-4.91e-05	—	—	—
childcare worker *	17	0.00056	-0.0847	-4.77e-05	—	—	—

Occupation	N	ρ_k (raw)	γ_k	δ_k (total)	ρ_k (net of dipl)	γ_k (direct)	γ_k (indirect via dipl)
cost estimator	184	-0.00332	0.0139	-4.62e-05	-4e-04	-5.51e-06	-4.07e-05
quality controller	372	-0.00153	0.0212	-3.24e-05	0.00375	7.96e-05	-1.12e-04
cleaner *	15	0.00031	-0.0931	-2.88e-05	—	—	—
forklift driver *	1	5e-05	-0.3764	-1.81e-05	—	—	—
concierge *	2	-4e-05	0.3019	-1.16e-05	—	—	—
insurance agent	305	-0.00026	0.0359	-9.24e-06	-0.00116	-4.18e-05	3.26e-05
domestic helper *	7	6e-05	-0.0596	-3.77e-06	—	—	—
assembler/mechanic *	14	-5e-04	0.0048	-2.40e-06	—	—	—
baker *	5	0.00013	-0.0104	-1.38e-06	—	—	—
conductor	161	-0.00166	4e-04	-6.71e-07	-0.0028	-1.13e-06	4.59e-07
butcher *	2	-8e-05	0.0024	-1.91e-07	—	—	—
administrative staff	4,967	0.09538	0	0.00e+00	0.06224	0.00e+00	0.00e+00
prison guard *	9	-2e-05	-0.0401	7.57e-07	—	—	—
packer *	18	-1e-05	-0.085	7.60e-07	—	—	—
mail carrier	54	-6e-04	-0.0097	5.86e-06	-0.00025	2.42e-06	3.44e-06
printer *	7	-0.00021	-0.045	9.29e-06	—	—	—
salesperson	515	-0.00013	-0.0722	9.30e-06	-0.0023	1.66e-04	-1.57e-04
librarian	89	0.00222	0.0085	1.90e-05	0.00155	1.33e-05	5.71e-06
bricklayer *	3	-0.00013	-0.184	2.42e-05	—	—	—
surveyor *	19	-0.00055	-0.0579	3.18e-05	—	—	—
merchandise	67	4e-04	0.1053	4.20e-05	0.00029	3.06e-05	1.13e-05
warehouse worker	433	-0.00258	-0.0187	4.82e-05	-0.00451	8.44e-05	-3.62e-05
magistrate	67	0.00018	0.2896	5.19e-05	-0.00022	-6.43e-05	1.16e-04
lawyer *	29	0.00039	0.1787	7.01e-05	—	—	—
journalist	141	0.00053	0.1402	7.41e-05	-0.00155	-2.17e-04	2.91e-04
laboratory technician	667	0.00357	0.0275	9.79e-05	0.00617	1.70e-04	-7.16e-05
driver	48	-0.00111	-0.0919	1.02e-04	-0.00103	9.48e-05	7.21e-06
physician	175	0.00091	0.125	1.14e-04	-0.00059	-7.35e-05	1.88e-04
caregiver	240	0.00627	0.0195	1.22e-04	0.00266	5.18e-05	7.03e-05
notary	40	0.00094	0.1706	1.60e-04	0.00093	1.58e-04	1.97e-06
graphic designer	581	0.00485	0.0379	1.84e-04	0.0051	1.93e-04	-9.34e-06
pharmacist	203	0.00477	0.0426	2.04e-04	0.00134	5.73e-05	1.46e-04

Occupation	N	ρ_k (raw)	γ_k	δ_k (total)	ρ_k (net of dipl)	γ_k (direct)	γ_k (indirect via dipl)
translator	186	0.00301	0.0754	2.27e-04	-0.00149	-1.13e-04	3.40e-04
social worker	297	0.00756	0.0393	2.97e-04	0.00269	1.06e-04	1.91e-04
draftsperson	283	-0.0057	-0.0701	4.00e-04	-0.00229	1.61e-04	2.39e-04
researcher	1,168	0.00423	0.1043	4.41e-04	0.00444	4.63e-04	-2.18e-05
teacher tertiary	319	0.00318	0.1462	4.65e-04	0.00182	2.66e-04	1.99e-04
customer service employee	987	0.00794	0.0594	4.72e-04	0.00488	2.90e-04	1.82e-04
teacher lower secondary	754	0.00469	0.1337	6.27e-04	-0.00044	-5.86e-05	6.85e-04
educator	636	0.01325	0.0494	6.55e-04	0.00276	1.37e-04	5.18e-04
recruiter	425	0.0115	0.0731	8.41e-04	0.00712	5.21e-04	3.20e-04
legal expert	1,213	0.01003	0.1019	1.02e-03	0.00564	5.75e-04	4.47e-04
teacher primary	482	0.01202	0.0947	1.14e-03	0.00368	3.48e-04	7.90e-04
teacher	1,206	0.01421	0.0825	1.17e-03	0.00557	4.60e-04	7.13e-04
secretarial staff	1,506	0.05206	0.0274	1.43e-03	0.03066	8.41e-04	5.87e-04
paramedic	713	0.02006	0.0945	1.90e-03	0.00858	8.11e-04	1.09e-03
nurse	1,309	0.02478	0.0911	2.26e-03	-0.00052	-4.74e-05	2.30e-03
editor	1,299	0.02622	0.0975	2.56e-03	0.01621	1.58e-03	9.76e-04
Total	48,785	—	—	-2.38e-02	—	-1.89e-02	-4.73e-03

Figure A3: Gender representation (ρ_k) versus wage premium (γ_k), by occupation



❖ Appendix 15

Table A15: Direct and Indirect contribution of each sector to the gender wage gap

Sector	N	ρ_k (raw)	γ_k	δ_k (total)	ρ_k (net of dipl)	γ_k (direct)	γ_k (indirect via dipl)
Training, Education & Scientific Research	3,756	0.04546	-0.0546	-2.48e-03	0.02095	-1.14e-03	-1.34e-03
Healthcare & Social Services	4,390	0.08181	-0.0225	-1.84e-03	0.02397	-5.40e-04	-1.30e-03
Socio-cultural sector	674	0.01249	-0.1261	-1.57e-03	0.00633	-7.98e-04	-7.76e-04
(Petro)chemicals, pharmaceuticals & science	3,837	-0.01579	0.0915	-1.45e-03	-9e-05	-8.06e-06	-1.44e-03
Other	2,611	0.0249	-0.0527	-1.31e-03	0.02141	-1.13e-03	-1.84e-04
Human Resources	992	0.02001	-0.0471	-9.43e-04	0.01375	-6.48e-04	-2.95e-04
Tourism, hospitality and leisure	783	0.00837	-0.1044	-8.74e-04	0.00061	-6.35e-05	-8.10e-04
Notarial services & Law	229	0.00745	-0.1161	-8.65e-04	0.00574	-6.66e-04	-1.99e-04
Energy & Environment	1,068	-0.01272	0.0671	-8.54e-04	-0.00465	-3.12e-04	-5.42e-04
Media, Entertainment, Communication & Marketing	1,915	0.00993	-0.0678	-6.73e-04	0.00089	-6.05e-05	-6.13e-04
Electronics & Electrical Engineering	2,469	-0.04144	0.0139	-5.74e-04	-0.01925	-2.67e-04	-3.08e-04
Other services to businesses and individuals	1,420	0.00806	-0.0687	-5.53e-04	0.0076	-5.22e-04	-3.13e-05
Banking, Finance & Insurance	3,342	-0.00979	0.0331	-3.24e-04	-0.01278	-4.23e-04	9.89e-05
Government/Public sector	4,245	0.01714	-0.0175	-3.00e-04	-0.00259	4.53e-05	-3.45e-04
Other industrial sectors	2,684	-0.02288	0.0044	-1.02e-04	-0.00382	-1.70e-05	-8.48e-05
Retail & Wholesale Trade	2,002	0.00024	-0.0933	-2.27e-05	-0.00589	5.49e-04	-5.72e-04
Fast Moving Consumer Goods	534	-6e-05	0.0672	-4.16e-06	-0.00098	-6.59e-05	6.17e-05
Agriculture & Food Industry	1,374	-0.00449	0	-0.00e+00	0.00085	0.00e+00	-0.00e+00
Logistics, Transport & Distribution	2,212	-0.00988	-0.0047	4.63e-05	-0.0147	6.89e-05	-2.26e-05
Construction	2,246	-0.02522	-0.0234	5.90e-04	-0.00287	6.72e-05	5.23e-04
Engineering consultancy/Engineering	1,759	-0.00897	-0.0757	6.79e-04	0.00196	-1.49e-04	8.28e-04
Telecom, ICT & Internet	4,243	-0.08461	-0.0121	1.02e-03	-0.03645	4.40e-04	5.81e-04
Total	48,785	—	—	-1.24e-02	—	-5.64e-03	-6.77e-03