

**The effects of education levels  
 on the gender pay gap   
across different industry sectors**

Marc N. Peter

Registration Number: 764996

26 June 2025

Darmstadt

Table of Contents

[1. Introduction 4](#_Toc200816381)

[1.1. Motiviation 4](#_Toc200816382)

[1.2. 4](#_Toc200816383)

[2. Literature Review 5](#_Toc200816384)

[3. SOEP Dataset 6](#_Toc200816385)

[4. Methodoligy 7](#_Toc200816386)

[4.1. The Significance of Skewness in Salary Distribution 7](#_Toc200816387)

[4.2. Econometric Model: Quantile Regression 8](#_Toc200816388)

[5. Results 10](#_Toc200816389)

[5.1. Impact of Gender and Years of Education on Salary 10](#_Toc200816390)

[5.2. Gender Pay Gap Across Multiple Industry Sectors 11](#_Toc200816391)

[6. Conclusion and Discussion 14](#_Toc200816392)

[7. Limitations and Outlook 15](#_Toc200816393)

Figures

[Figure 1: Histogram of yearly salary in € of the working population (SOEP dataset) 7](#_Toc200800669)

Tables

[Table 1: Summary of SOEP Dataset 6](#_Toc200739567)

1. Introduction
   1. Motiviation
2. Literature Review
3. SOEP Dataset

Table 1: Summary of SOEP dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **n** | **%** | **Mean** | **Min/Max** |
| *Female* | 12760 | 54.20 | 0.542 | - |
| *Male* | 10762 | 45.80 | 0.458 | - |
| *Age* | 23522 | - | 48.282 y | 17 y/102 y |
| *People in Household* | 23445 | - | 2.907 | 1/13 |
| *Household with Children* | 9337 | 39.70 | 0.397 | - |
| *Single Parent Household* | 817 | 3.50 | 0.035 | - |
| *Years of Education* | 21911 | - | 12.401 | 7/18 |
| *Gross Salary Year* | 23522 | - | 16775.768 € | 0 €/269425 € |
| *Gross Salay Month* | 23522 | - | 1771.885 € | 0 €/35261 € |
| *Employed Full-Time* | 8700 | 37.00 | 0.370 | - |
| *Employed Part-Time* | 3481 | 14.80 | 0.148 | - |
| *Training Apprenticeship* | 695 | 3.00 | 0.030 | - |
| *Irregular employment or in marginal* | 1446 | 6.10 | 0.061 | - |
| *Unemployed* | 9169 | 39.00 | 0.390 | - |
| *Working in Agriculture* | 176 | 0.70 | 0.007 | - |
| *Working in Manufacturing* | 2599 | 11.00 | 0.110 | - |
| *Working in Energy* | 214 | 0.90 | 0.009 | - |
| *Working in Constructions* | 674 | 2.90 | 0.029 | - |
| *Working in Trade* | 1584 | 6.70 | 0.067 | - |
| *Working in Traffic* | 578 | 2.50 | 0.025 | - |
| *Working in Gastronomy & Hotel* | 457 | 1.90 | 0.019 | - |
| *Working in Telecommunication* | 493 | 2.10 | 0.021 | - |
| *Working in Services* | 2174 | 9.20 | 0.092 | - |
| *Working in Public Administration, Education and Health* | 4637 | 19.70 | 0.197 | - |

(Busch & Holst, 2009) (Jochmann-Döll & Scheele, 2020) (Lang & Groß, 2020)

1. Methodoligy

This chapter outlines the methodological approach employed to investigate the gender pay gap across various industry sectors and to analyze the presence of a "glass ceiling" effect within the provided dataset. The analysis leverages a combination of data preprocessing, exploratory data analysis, and advanced econometric modeling, specifically focusing on Quantile Regression.

* 1. The Significance of Skewness in Salary Distribution

The distribution of salary data is almost universally characterized by positive or right-skewness. This means that a large proportion of individuals earn lower to moderate incomes, while a smaller number of individuals earn significantly higher incomes, creating a long tail to the right of the distribution. This pattern is widely observed in economic studies, including those on regional income and wages in Germany (Immel, 2021).Figure 1 shows the distribution of the yearly salary of the working population based on the SOEP dataset. Based on the distribution, it is clear that the yearly salary is positive skewed. Additionally, a calculated skewness factor (Joanes & Gill, 1998) of **1.6** indicates a strong right-skewness and therefore a small number of inidviduals earn siginifcantly higher incomes than a large proportion of the working population.

Figure 1: Histogram of yearly salary in € of the working population (SOEP dataset)

A graph of a salary

AI-generated content may be incorrect.

Standard regression methods like Ordinary Least Squares (OLS), which model the conditional mean, can provide misleading insights when applied to highly skewed data like salaries.

* 1. Econometric Model: Quantile Regression

Given the inherently skewed nature of salary distributions and the research questions concerning the gender pay gap, particularly the "glass ceiling" effect which by definition occurs in the upper echelons of the income distribution, a modeling approach that can robustly account for these distributional characteristics is essential. This leads directly to the application of Quantile Regression (QR).

As stated in "Quantile regression: A short story on how and why" (Waldmann, 2018), QR "quantifies the association of explanatory variables with a conditional quantile of a dependent variable without assuming any specific conditional distribution. It hence models the quantiles, instead of the mean as done in standard regression.". This underscores that the quantile function QY(τ∣x) = inf y ∈ R∣F(y∣x) ≥ τ (Čížek, 2000) completely describes the conditional distribution of the response variable Y given covariates (potential confounding variables) X. This choice is particularly justified given the characteristics of salary data and the research questions:

**Addressing Skewness and Heterogeneity:** As highlighted before, salary distributions are typically right-skewed. Traditional Ordinary Least Squares (OLS) regression models focus on the conditional mean, which is susceptible to the influence of outliers (high earners) and may not accurately represent the experience of individuals across the entire income spectrum. QR, in contrast, estimates the effects of predictors at various conditional quantiles (percentiles) of the dependent variable's distribution. This allows for the investigation of heterogeneous effects, meaning that the impact of factors like gender, age, or years of education on salary may differ significantly at the low, middle, and high ends of the pay scale.

**Robustness to Outliers:** QR minimizes the sum of asymmetrically weighted absolute residuals, making it inherently more robust to outliers in the salary data compared to OLS (which minimizes squared residuals). Outliers only have an influence on the quantile curves close to them, threrefore extreme values do not disproportionately pull the more central quantile estimates (Waldmann, 2018). This is a critical advantage given the presence of high earners who can exert undue influence on mean-based models.

**No Strict Distributional Assumptions:** QR does not impose strong parametric assumptions about the distribution of the error term (e.g. normality), making it a more flexible and robust tool for real-world datasets where such assumptions often do not hold (Waldmann, 2018).

The specific characteristics of the SOEP dataset and the research questions align perfectly with the strengths of Quantile Regression (Friedrich & Hirtz, 2021).

To analyse the impact of different factors on the gender pay gap and the “glass ceiling” effect,

1. Results and Discussion
   1. Impact of Gender and Years of Education on Salary

A graph on a black background

AI-generated content may be incorrect.

* 1. Gender Pay Gap Across Multiple Industry Sectors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | Manufacturing | Trade | Services | Administration,  Education & Health |
| Persons [n] | 2537 | 1505 | 2074 | 4489 |
| Females [%] | 30.23 | 63.19 | 58.05 | 70.28 |
| Median education [years] | 11.5 | 11.5 | 12 | 13 |
| Minimum education [years] | 7 | 7 | 7 | 7 |
| Maximum education [years] | 18 | 18 | 18 | 18 |
| Median male education [years] | 11.5 | 11.5 | 13.5 | 14.5 |
| Median female education [years] | 11.5 | 11.5 | 12 | 12 |
| Males employed full-time [%] | 92.99 | 84.66 | 77,01 | 83,58 |
| Females employed full-time [%] | 58.02 | 33.02 | 38.12 | 42 |
| Males employed part-time [%] | 1.98 | 4.69 | 12.87 | 10,42 |
| Females employed part-time [%] | 31.03 | 43.32 | 39.37 | 46,28 |
| Males employed irregular/marginal [%] | 0.51 | 6.68 | 6.78 | 3,67 |
| Females employed irregular/marginal [%] | 7.3 | 18.4 | 18.44 | 7.92 |

A graph of different colored lines

AI-generated content may be incorrect.

A graph of different colored lines

AI-generated content may be incorrect.

1. Conclusion
   1. Limitations
   2. Outlook

(Winters, 2018)

(Arulampalam, et al., 2007) (Arulampalam, et al., 2007)

Bibliography

Winters, J., 2018. *IAB-Forum.* [Online]   
Available at: https://iab-forum.de/en/women-still-hit-the-glass-ceiling-despite-operational-promotion-of-equal-opportunities/  
[Accessed 10 06 2025].

Arulampalam, W., Booth, A. L. & Bryan, M. L., 2007. Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *ILR Review,* 60(2), pp. 163-186.

Collischon, M., 2017. *Is there a Glass Ceiling over Germany?,* Erlangen: Friedrich-Alexander-University Erlangen-Nuremberg.

Joanes, D. N. & Gill, C. A., 1998. Comparing Measures of Sample Skewness and Kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician),* 47(1), pp. 183-189.

Waldmann, E., 2018. Quantile regression: A short story on how and why. *Statistical Modelling,* 18(3-4), pp. 203-218.

Čížek, P., 2000. Quantile Regression. In: *XploRe® — Application Guide.* Heidelberg: Springer, pp. 19-48.

Friedrich, A. & Hirtz, S., 2021. Same educational level, same wage returns? An analysis over time and across occupations in Germany. *Journal of Education and Work,* 34(5-6), pp. 632-648.

Immel, L., 2021. *The Regional Distribution of Income and Wages: Causes and Consequences,* Munich: ifo Institute.

Busch, A. & Holst, E., 2009. *Glass Ceiling Effect and Earnings – The Gender Pay Gap in Managerial Positions in Germany,* Berlin: DIW Berlin.

Jochmann-Döll, A. & Scheele, A., 2020. *The Road to Equal Pay for Women and Men,* Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend.

Appendix

|  |
| --- |
|  |