Decomposing the wage impact of education-occupation mismatches: the contributions of worker, firm, and job title heterogeneity[†]

Tiago Silva[*]

fep.rc, Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal

Anabela Carneiro

fep.rc, Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal

Paulo Guimarães

fep.rc, Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal

[*] Email: up201804678@up.pt (corresponding author).

1. Introduction

In this study, we quantify the contributions of worker, firm, and job title permanent observed and unobserved heterogeneity to the total real hourly wage impact of vertical and horizontal education-occupation mismatches. We perform this quantification for two wage effects commonly estimated in the literature: the wage impact of vertical and horizontal mismatches when comparing mismatched workers to well-matched ones with the same education, and the wage impact when comparing mismatched workers to well-matched ones in the same occupation. To do so, we first estimate two benchmark models that are inspired by the two seminal models in this literature: the Duncan and Hoffman (1981) model and the Verdugo and Verdugo (1989) model. Then, we introduce worker, firm, and job title fixed effects to them. Additionally, we also estimate a model that further accounts for worker-firm match quality heterogeneity. Finally, we apply the decomposition proposed by Gelbach (2016) (henceforth, Gelbach decomposition), which provides a theoretically sound way to measure the relative contribution of each of the three sources of heterogeneity to the wage effects estimated by previous literature, because it yields unambiguous results where the order of inclusion of covariates does not matter, unlike the sometimes used approach of introducing one fixed effect at a time.

2. Data

In this work, we use data from *Quadros de Pessoal* (QP) for the period between 2010 and 2023. QP is a Portuguese administrative linked employer-employee dataset of high quality and representativeness. We focus on individuals aged 16-66 with post-secondary non-tertiary or higher education (ISCED 4 or higher), since those are the levels of education that tend to offer more field-specific skills (which are essential to study horizontal mismatch). After the data treatment, the dataset ended up containing 4,456,742 worker-year observations, which are composed of 731,059 individual workers, 77,005 individual firms, and 29,918 unique job titles.[1]

We classify a worker as undereducated(overeducated) if the number of years required for their current attained level of education is below(above) the mode number of years of education of newly hired workers (i.e., workers with less than one year of tenure) in their occupation in that year, and as well-matched otherwise. Similarly, a worker is classified as horizontally mismatched if their field of education differs from the mode of new hires in their occupation.

We construct a measure of hourly earnings by dividing regular earnings (i.e., base wages plus other regular wage components)[2] by the regular hours worked. We deflate wages by the total Consumer Price Index (base 2012).

---

[1]However, some workers, firms, and job titles appear only once in our 14 years of data, and, thus, they will not be used for the regressions, since singletons do not add any information to a model with fixed effects.

[2]As various other authors that used QP data before, like Card et al. (2016) and Portugal et al. (2024), we leave out non-regular components of pay, as well as overtime pay, because they may bias the estimates, since QP data only reflects the reference month of October, when it is collected.

3. Methods

Our benchmark model is the following wage equation:

$$w_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + \mathbf{F}'_{jt}\boldsymbol{\gamma} + \mathbf{Z}'_{kt}\boldsymbol{\delta} + \boldsymbol{\theta}_{VM_{it} \times HM_{it}} + \lambda_t + \varepsilon_{it}, \tag{1}$$

where w_{it} is the natural logarithm of the real hourly wage of worker i, in period t; $\mathbf{X}'_{it}$ is a vector of time-varying observable worker characteristics; $\mathbf{F}'_{jt}$ is a vector of time-varying observable firm characteristics; $\mathbf{Z}'_{kt}$ is a vector of time-varying observable characteristics of workers' jobs; VM_{it} is a vector of dummies that indicate whether a worker is vertically mismatched; HM_{it} is a dummy that indicates whether a worker is horizontally mismatched; λ_t are year fixed effects; and ε_{it} is an error term.

The vector of time-varying worker characteristics includes quadratic and cubic terms on age.[3] The vector of time-varying firm characteristics includes linear and quadratic terms on the log of the number of employees, dummies for the NUTS II region of the firm, and dummies for the NACE Rev. 2 four-digit sector of economic activity of the firm. The vector of time-varying job characteristics includes a dummy variable that takes the value 1 for part-timers, linear and quadratic terms on firm tenure, and a dummy variable that takes the value 1 for temporary contracts.

Additionally, depending on whether we aim to compare workers in the same occupation (henceforth, within-occupation analysis) or workers with the same education (henceforth, within-education analysis), we also include, in equation (1), four-digit ISCO-08 occupation fixed effects or education fixed effects (i.e., fixed effects for each pair of field and level of education), respectively.

Our full specification consists of equation (1) enhanced with worker, firm, and job title fixed effects. In the spirit of Carneiro et al. (2012), we define job titles as the combination of a worker's professional category and the applicable collective bargaining agreement.

After having estimated the benchmark regression and the full specification, we perform the decomposition proposed in Gelbach (2016).[4]

4. Results

Table 1 presents the results of our baseline models. Panel A concerns the within-occupation analysis, while Panel B covers the within-education analysis. Columns (1) and (2) present, respectively, the benchmark and full model coefficients, while columns (3), (4), and (5) display, respectively, the contributions of the worker, firm, and job title heterogeneity obtained through the Gelbach decomposition.

The estimates of the wage impact of education-occupation mismatches presented in Table 1 columns (1) and (2) are in line with international evidence.

---

[3]The inclusion of a linear term on age is impossible in the model with worker fixed effects due to perfect multicollinearity between age and year.

[4]For a detailed explanation of how this decomposition can be applied to high-dimensional fixed effects, please see Portugal et al. (2024).

Regarding the results of the Gelbach decomposition, overall, the main conclusion seems to be that individual heterogeneity is the most relevant contributor to the total wage impact of education-occupation mismatches estimated through the benchmark models. On average, around seventy percent of the total wage impact is accounted for by worker heterogeneity. This shows that workers do not randomly sort into education-occupation match status, and that such non-random assignment has important wage consequences. This is in line with the evidence found by Card et al. (2016) for the gender pay gap, and by Portugal et al. (2024) for the returns to schooling.

Nevertheless, job title sorting is also found to play a non-negligible role in this form of wage heterogeneity. If individuals were randomly assigned to job titles, everything else constant, the wage penalty of undereducated but horizontally well-matched workers, compared to well-matched ones in the same occupation, would reduce by about 1.23 log points, and the wage penalty of undereducated(overeducated) but horizontally well-matched workers, compared to well-matched ones with the same education, would be attenuated by around 2.63(2.33) log points. While previous studies like Büchel and Mertens (2004), Roller et al. (2019), Sicherman (1991), and Wen and Maani (2019) had already shown that overeducation affects the probability of upward job mobility (when measured as wage growth), to the best of our knowledge, we are the first to quantify the portion of the total estimated wage effects of education-occupation mismatches that is attributable to job title heterogeneity.

Additionally, we show that, once worker and job title heterogeneity are controlled for, firm sorting plays a much smaller role in this form of wage heterogeneity than what has been found for others, such as the returns to schooling (Portugal et al., 2024) and the gender pay gap (Card et al., 2016). These results also suggest that the higher importance of firm sorting for the wage impact of vertical education-occupation mismatches found in Araújo and Carneiro (2023) may have been driven by the omission of controls for job title heterogeneity (since firm heterogeneity and job title heterogeneity can be correlated).

Finally, we assessed whether the quality of the match between workers and firms plays a significant role in the wage impact of education-occupation mismatches. To do so, we estimated full models with worker-firm and job title fixed effects. Such models nest our baseline specification, since they control for worker, firm, and job title permanent heterogeneity, but also control for an additional source of heterogeneity: the quality of the match between workers and firms. This exercise showed that the quality of the matching between workers and firms plays no significant role in this form of wage heterogeneity, which is in line with the evidence presented in Portugal et al. (2024) for the returns to schooling.[5]

---

[5]The results are available upon request.

TABLE 1. Gelbach decompositions for baseline models

Panel A - Gelbach Decomposition of within-occupation coefficients

	Benchmark Model	Full Model	Contribution of:		
			Worker FE	Firm FE	Job Title FE
	(1)	(2)	(3)	(4)	(5)
H. Mismatched	-0.0001	0.0003	-0.0014	0.0012	-0.0002
	(0.0014)	(0.0007)	(0.0012)	(0.0027)	(0.0007)
Undereducated	-0.0805***	-0.0437***	-0.0175***	-0.0069	-0.0125***
	(0.0022)	(0.0013)	(0.0019)	(0.0051)	(0.0035)
Overeducated	0.0568***	0.0099***	0.0206***	0.0248***	0.0015
	(0.0014)	(0.0007)	(0.0012)	(0.0029)	(0.0017)
H Mis.#Under.	-0.0277***	0.0204***	-0.0527***	-0.0069	0.0115***
	(0.0049)	(0.0022)	(0.0044)	(0.0088)	(0.0028)
H Mis.#Over.	-0.0341***	-0.0018**	-0.0265***	-0.0059**	0.0002
	(0.0019)	(0.0009)	(0.0017)	(0.0028)	(0.0012)

Panel B - Gelbach Decomposition of within-education coefficients

	Benchmark Model	Full Model	Contribution of:		
			Worker FE	Firm FE	Job Title FE
	(1)	(2)	(3)	(4)	(5)
H. Mismatched	0.0321***	0.0026***	0.0253***	0.0002	0.0040*
	(0.0016)	(0.0008)	(0.0014)	(0.0031)	(0.0022)
Undereducated	0.1606***	-0.0419***	0.2333***	-0.0040	-0.0269***
	(0.0030)	(0.0013)	(0.0029)	(0.0063)	(0.0077)
Overeducated	-0.1415***	-0.0109***	-0.1085***	0.0008	-0.0230***
	(0.0015)	(0.0007)	(0.0014)	(0.0032)	(0.0043)
H Mis.#Under.	-0.2154***	0.0218***	-0.2507***	0.0077	0.0057
	(0.0056)	(0.0022)	(0.0051)	(0.0093)	(0.0066)
H Mis.#Over.	-0.0431***	-0.0015*	-0.0277***	-0.0051*	-0.0088***
	(0.0021)	(0.0009)	(0.0019)	(0.0029)	(0.0033)

Notes: The table presents the results of the Gelbach decomposition for our baseline models. Columns (1) and (2) display the estimated coefficients from the benchmark model and the full model, respectively, while columns (3), (4), and (5) display the contributions of worker, firm, and job title heterogeneity, respectively. Panel A concerns the within-occupation regressions, while Panel B regards the within-education regressions.

Standard errors (clustered at the respective level- worker, firm, or job title) in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: Author's calculations based on *Quadros de Pessoal* data.

5. References

Araújo, I., & Carneiro, A. (2023). Educational mismatches of newly hired workers: Short- and medium-term effects on wages. *International Labour Review, 162*(3), 355–383. https://doi.org/10.1111/ilr.12374.

Büchel, F., & Mertens, A. (2004). Overeducation, undereducation, and the theory of career mobility. *Applied Economics, 36*(8), 803–816. https://doi.org/10.1080/0003684042000229532.

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics, 131*(2), 633–686. https://doi.org/10.1093/qje/qjv038.

Carneiro, A., Guimarães, P., & Portugal, P. (2012). Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity. *American Economic Journal: Macroeconomics, 4*(2), 133–152. https://doi.org/10.1257/mac.4.2.133.

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review, 1*(1), 75–86. https://doi.org/10.1016/0272-7757(81)90028-5.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics, 34*(2), 509–543. https://www.jstor.org/stable/26553211.

Portugal, P., Reis, H., Guimarães, P., & Cardoso, A. R. (2024). What Lies behind the Returns to Schooling: The Role of Labor Market Sorting and Worker Heterogeneity. *The Review of Economics and Statistics.* https://doi.org/10.1162/rest_a_01482.

Roller, C., Rulff, C., & Tamminga, M. M. (2019). *Educational mismatch and mobility* (Ruhr Economic Papers, No. 798). RWI – Leibniz-Institut für Wirtschaftsforschung. https://doi.org/10.4419/86788926.

Sicherman, N. (1991). "Overeducation" in the Labor Market. *Journal of Labor Economics, 9*(2), 101–122. https://www.jstor.org/stable/2535236.

Verdugo, R. R., & Verdugo, N. T. (1989). The Impact of Surplus Schooling on Earnings: Some Additional Findings. *The Journal of Human Resources, 24*(4), 629–643. https://doi.org/10.2307/145998.

Wen, L., & Maani, S. A. (2019). Job mismatches and career mobility. *Applied Economics, 51*(10), 1010–1024. https://doi.org/10.1080/00036846.2018.1524569.