What Lies Behind the Returns to Schooling: The Role of Worker Sorting and Peer Quality

Ana Rute Cardoso (ICS - Instituto de Ciências Sociais)
Paulo Guimarães (Banco de Portugal and Universidade do Porto)
Pedro Portugal (Banco de Portugal and Nova SBE)
Hugo Reis (Banco de Portugal and Universidade Católica)

Joint Bank of Portugal and Bank of Spain Workshop on Labor Economics
Tavira
June 2022

Motivation 1: Measuring the wage returns to education

- Fits into a long tradition of explaining wage differentials by looking at worker-side attributes
  - mainly competitive background (Becker, 1962): wage differentials reflect differences in workers' productivity
  - which depends on workers' skills: education, training, seniority, experience, motivation, ability, ...
  - key role of education
- Empirically, major progress over past decades
- In large part, dealing with the endogeneity of educational decisions (IVs, RD, SDDC)
- Line of literature fully detached from that on the firm shaping wages
  - possibly constrained, for a long time, by the availability of data from household surveys only
  - time to redress this neglect?

Motivation 2: "The return of the firm to labor economics"

- Firm-side explanations for wage differentials: a profusion of theories
  - why firms find it profitable to pay non-competitive wages
  - they design incentive schemes to attract better workers, retain them, and enhance their productivity
  - efficiency wages; implicit contracts; rent-sharing; insider-outsider; labor market frictions & job search and matching;
- Empirically
  - early case studies in 1950s: firms' pay standard diverges within narrowly defined regions and industries (Lester, 1952) (Reynolds, 1951)
  - in 1990s, firm back on the spotlight: intra-industry wage differentials or gender pay gap (Groshen, 1991, 1991a)
  - the boom that followed Abowd, Kramarz and Margolis (1999) estimation of firm and worker fixed effects in wage regressions
- Interest in the role of the firm has been growing
  - increasingly richer linked employer-employee longitudinal datasets
  - trends in wage dispersion; matching of workers and firms (Torres et al., 2018); gender pay gap (Cardoso et al., 2016; Card et al., 2015));

Motivation 3: The role of peer and spillover effects

- A new branch of literature based on peer and spillover effects
  - Theoretical models allowing for the role of learning networks
  - Having in mind assortative matching (Gary Becker, again)
  - Key role of education, again (Acemoglu and Angrist, 2000)
- Empirically, some progress recently advanced
- Attempts made to measure peer and spillover effects
  - Focused on the spillovers from education
  - Along the geographical dimension (Moretti, 2004)
  - Despite the perils of peer effect (Manski - many papers - Angrist, 2014) and the curse of reflection effects
  - An elegant way to overcome reflection effects (Arcidiacono et al., 2012)
  - An extension to the sources of wage variation (Cornelissen et al., 2017; Batisti, 2013)

Aims

- How much of the economy-wide return to education operates through the allocation of workers with different schooling levels to firms with different pay standards?
  - if firms are heterogeneous and workers are not randomly allocated,
  - part of the overall return to education could indeed be a return to working for a good firm
- Assess the returns to education taking into account:
  - **who** the worker is —observed and unobserved ability
  - **what** he does —detailed task (job)
  - **for whom** —the employer
  - **with whom** —peers (in the same establishment, job-title and year)
    - Measure peer effects; Measure the role of peers education
    - How the return on peers education relates with other peer effects
- A decomposition exercise on the sources of the returns to education
  - Gelbach (2016) unambiguous conditional decomposition of the impact of omitted covariates on the coefficient of interest

Institutional wage setting

Wage bargaining system in Portugal prevailing over the sample period:

- Mandatory national minimum wage
- Collective bargaining takes place mainly at the industry level, and voluntary or mandatory extensions are commonplace
- Around 30,000 job-title wage floors bind wages in a given year
- Despite very low unionization rates (around 10 percent)
- Firms often pay wages above bargained wages (wage cushion)

Data

- Longitudinal linked employer-employee data: *Quadros de Pessoal*
- Years: 1995 to 2013 (2001 not available)
- Variables:
  - worker's gender, birth date, schooling, occupation, date of hire, earnings (several components), hours of work (normal and overtime), collective bargaining agreement, worker category in the agreement
  - firm's industry, location, etc.
  - unique identifiers for workers, establishments, and job titles
- Final dataset (after constraints on two peers, full-timers, base wage at least 80% min wage, non-missing relevant data, etc; largest connected set):
  - 19 million observations
    3.7 million workers, 282 thousand firms, and 48 thousand job titles
    3.9 million peer groups
- Hourly wage = (overall monthly earnings, incl, overtime)/(sum of normal and overtime hours)

A HD wage regression equation with worker and establishment/job title/year fixed effects

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \theta_{\mathbf{F}\times\mathbf{J}(i,t)} + \varepsilon_{it} \quad , \tag{1}$$

- where y_{it} is the logarithm of the hourly wage for each worker i $(i = 1, ..., N)$ at year t $(t = 1, ..., T)$;
- $\mathbf{x}_{it}$ is a vector of observed time-varying characteristics of workers and firms;
- α_i is a time-invariant worker fixed effect;
- $\theta_{\mathbf{E}\times\mathbf{J}(i,t)}$ is a unique establishment/job-title specific time-varying fixed effect;
- and ε_{it} is the disturbance term of the regression.

We assume strict exogeneity, $E(\varepsilon_{it}|\mathbf{x}_{it}, \alpha_i, \theta_{\mathbf{E}\times\mathbf{J}(i,t)}, \mu_t) = 0$, to ensure unbiasedness of the regression coefficient estimates.

The vector of explanatory variables, $\mathbf{x}_{it}$, comprises a quadratic on age of the worker, a quadratic on tenure, as well as a measure of firm size.

Estimation algorithm, Guimarães and Portugal (2010)

Controlling for worker, establishment/job title/year -specific effects requires the introduction of two high-dimensional fixed effects in the linear regression model. In matrix form:

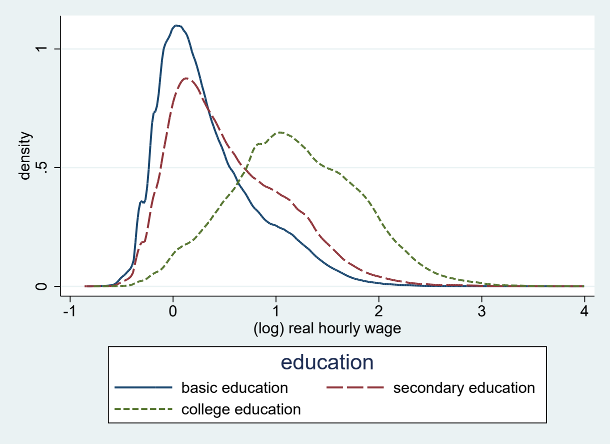
$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{D_1}\alpha + \mathbf{D_2}\theta + \mu \tag{2}$$

where X is a matrix of time-varying explanatory variables and D_1 and D_2 are high-dimensional matrices for the worker, establishment/job title/year effects. The normal equations may be rewritten as

$$\begin{bmatrix} \beta \\ \alpha \\ \theta \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'(Y - D_1\alpha - D_2\theta) \\ (D_1'D_1)^{-1}D_1'(Y - X\beta - D_2\theta) \\ (D_2'D_2)^{-1}D_2'(Y - X\beta - D_1\alpha) \end{bmatrix}$$

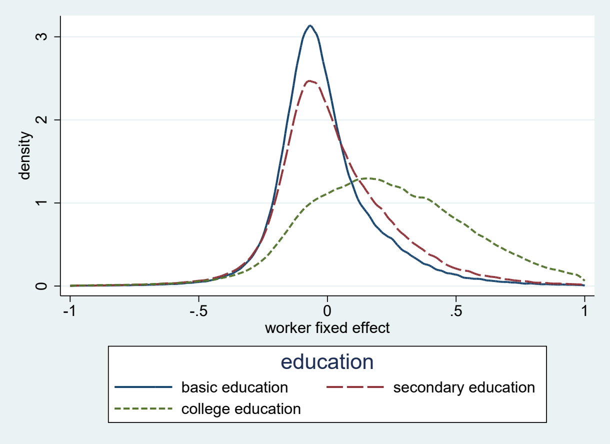
suggesting an iterative solution that alternates between estimation of β, α, and θ.

Graphical overview of the impact of education on wages



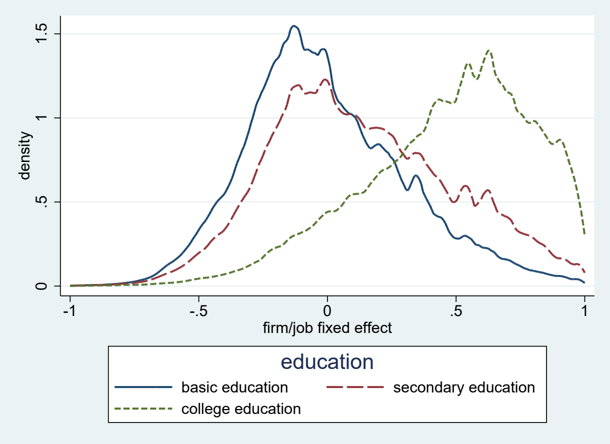
More educated workers collect higher wages

Graphical overview of the worker fixed effect



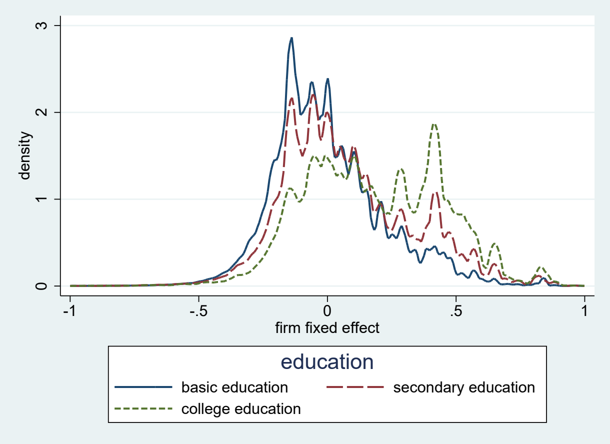
Better educated workers exhibit higher worker fixed effects

Graphical overview of the establishment/job fixed effects



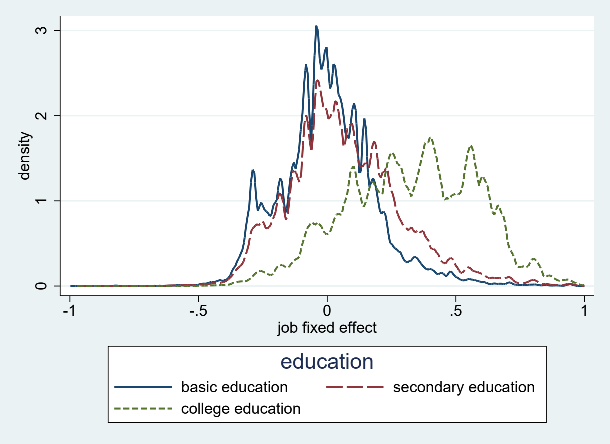
Better educated workers populate high paying firms (establishments) and high paying job titles

Graphical overview of the establishment fixed effects

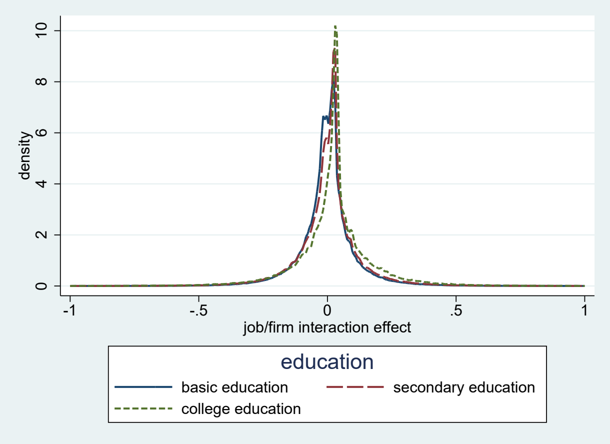


Better educated workers populate high paying establishments.

Graphical overview of the job title fixed effects



Graphical overview of the interaction between establishments and job title fixed effects



What role do establishments and jobs play shaping the returns to education? Gelbach (2016) decomposition

The idea is quite simple. It starts out with the well-known omitted variable bias formula:

- Assume there are two sets of right-hand side variables in the full model, X_1 and X_2
- X_1 contains the regressor of interest, worker's level of education (plus basic controls)
- X_2 contains covariates traditionally omitted in wage regressions, the establishment/job title/year and the worker-fixed effects
- Consider the base regression of (log) wages on X_1 only:

$$\mathbf{Y} = \mathbf{X_1} \boldsymbol{b}_1^{base} + \varepsilon \tag{3}$$

$$\begin{aligned} \widehat{\boldsymbol{b}}_1^{base} &= \widehat{\beta}_1^{full} + (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{X}_2 \widehat{\beta}_2 & (4) \\ &= \widehat{\beta}_1^{full} + \boldsymbol{\delta} & (5) \end{aligned}$$

Following Gelbach (2016) decomposition of the difference in the estimated coefficients of X_1 between the base and the full models, we know that

$$\widehat{\boldsymbol{b}}^{base} - \widehat{\boldsymbol{\beta}}^{full} = \mathbf{P}_{\mathbf{X_1}} \mathbf{D}_{fj} \widehat{\boldsymbol{\theta}} + \mathbf{P}_{\mathbf{X_1}} \mathbf{D}_{i} \widehat{\boldsymbol{\alpha}} \quad , \tag{6}$$

where $\mathbf{P}_{\mathbf{X_1}} = (\mathbf{X_1}'\mathbf{X_1})^{-1}\mathbf{X_1}'$ and (taking the firm/job effects/year as illustration):

- $\mathbf{D}_{fj}$ is a design matrix for the establishment/job title fixed effects
- $\widehat{\boldsymbol{\theta}}$ is a vector of the estimates of the establishment/job title fixed-effects in the full model
- therefore, $\mathbf{D}_{fj}\widehat{\boldsymbol{\theta}}$ is a vector with the estimates of the establishment/job title fixed effects
- and $\mathbf{P}_{\mathbf{X_1}}\mathbf{D}_{f}\widehat{\boldsymbol{\theta}}$ simply the coefficients of the regression of the establishment fixed effects on the variables $\mathbf{X}_1$ in the base model

Thus, we can rewrite the above equation more succinctly as

$$\widehat{\boldsymbol{b}}_{base} - \widehat{\boldsymbol{\beta}}_{full} = \widehat{\boldsymbol{\delta}}_{\boldsymbol{\alpha}} + \widehat{\boldsymbol{\delta}}_{\boldsymbol{\theta}} \tag{7}$$

- The change in the coefficient of interest is partitioned into the role of the different additional covariates

Our first application of Gelbach decomposition

Base regressor of interest in X_1:

- education (plus basic controls: age, tenure, firm size, time)

Additional covariates in X_2 (omitted)

- establishment/job title/year fixed effects
- worker fixed effects

Purpose – identify the conditional contribution of the worker, firm and the job title to the returns to schooling

Mandatory Schooling

Table: List of key legal documents on compulsory schooling

Year	Document	Years of compulsory schooling	First cohort affected	School entry age
1911	DL from March 29	3 years	Boys and girls born after 1904 (to be enrolled in the 1st grade in 1911)	7
1956	DL 40964	4 years	Boys born in or after 1949 (to be enrolled in the 1st grade in 1956)	7
1960	DL 42994	4 years	Girls born in or after 1953 (to be enrolled in the 1st grade in 1960)	7
1964	DL 45810	6 years	Boys and girls born in or after 1957 (to be enrolled in the 1st grade in 1964)	7
1986	Law 46/86	9 years	Boys and girls born in or after 1980 (to be enrolled in the 1st grade in 1986)	6

Source: *Diário da República*.

Note: Law 5/73 from 1973 increased mandatory schooling from 6 to 8 years, but it was not put into practice because subsequent legislation required for its impementation ended up not being enacted.

Wage Equation

	OLS	IV	Full
Age	0.0403	0.0403	
	(0.0002)	(0.0003)	
Age squared	-0.0003	-0.0003	-0.0002
	(0.0000)	(0.000)	(0.0000)
Tenure	0.0181	0.0180	0.0058
	(0.0002)	(0.0003)	(0.0000)
Tenure squared	-0.0002	-0.0002	-0.0001
	(0.0000)	(0.000)	(0.0000)
Firm size (log)	0.0604	0.0611	0.0263
	(0.0014)	(0.0026)	(0.0000)
Gender (Female=1)	-0.2721	-0.2716	
	(0.0013)	(0.0022)	
Schooling	**0.0791**	**0.0774**	
	(0.0003	(0.0032)	
Time effects	✓	✓	
Worker effects			✓
Establishment/Job-title/Year effects			✓

Decomposing the returns to education, a first pass

Table: Conditional Decomposition of the Return to Education - OLS

Benchmark regression	Individual	Establishment/Job
0.0791		
	0.0303	0.0488

Note: Decompositions based on Gelbach (2016).

Decomposing the Establishment/Job-title contribution

Table: Decomposition of the firm/job - OLS

Total contribution	Establishment	Job title	Interaction
0.0488			
	0.0201	0.0267	0.0020

Decomposing the returns to education, a first pass

Table: Conditional Decomposition of the Return to Education - OLS

Benchmark regression	Individual	Establishment/Job
0.0791		
	0.0303	0.0488

Note: Decompositions based on Gelbach (2016).

Table: Conditional Decomposition of the Return to Education - IV

Benchmark regression	Individual	Establishment/Job
0.0774		
	0.0232	0.0542

Note: Decompositions based on Gelbach (2016).

A wage regression equation accounting for peer effects

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \eta\overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad , \tag{8}$$

where

- $\overline{\alpha}_{-it}$ is the average of worker fixed effects over of worker i at time t
- δ is the associated coefficient.
- $\theta_{\mathbf{P}(i,t)}$ is the Establishment/Job/Year effect

- Estimation: Arcidiacono et al. (2012) empirical procedure to account for peers' unobservable attributes

The notion of peer

- Work in the same establishment
- Covered by the same collective agreement
- Work in the same job-title (categoria profossional)
- In the same year (in October)
-
- only Hugo is my peer
-
- 3.9 million peer groups
- 4.9 workers per peer group
- 14 peer groups by firm
- 47.8 peer groups by job title

A wage regression equation accounting for peer effects - Identification

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \eta\overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad , \tag{9}$$

- identification of η would come strictly from changes on the size of the peer groups (N_p), eliminating any endogenous contamination from sorting into establishment and job-title

Table: Identification - Illustrative example

	Worker	α_i					N_p	$\overline{\alpha}_{it}$	$\overline{\alpha}_{-it}$	change in peer quality
		Ana	*Rute*	*Pedro*	*Paulo*	*Hugo*				
2010	*Ana*	2	0				1	1	0	
2011	*Ana*	2	0	1			2	1	1/2	0,5
2012	*Ana*	2	0	1	1		3	1	2/3	0,17
2013	*Ana*	2	0	1	1	1	4	1	3/4	0,08

Estimating a wage regression equation accounting for peer effects

The estimating equation can be written:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \eta_0 \mathbf{W}\mathbf{D}\boldsymbol{\alpha} + \epsilon = \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} + \eta_0 \mathbf{W}]\mathbf{D}\boldsymbol{\alpha} + \epsilon \qquad (10)$$

and from the first order conditions for minimization of *SSR* we get:

$$\begin{aligned}
\frac{\partial S(.)}{\partial \widehat{\boldsymbol{\beta}}} &= \mathbf{X}'\mathbf{e} = \mathbf{0} \\
\frac{\partial S(.)}{\partial \widehat{\eta}_0} &= \widehat{\boldsymbol{\alpha}}'\mathbf{D}'\mathbf{W}\mathbf{e} = \mathbf{0} \\
\frac{\partial S(.)}{\partial \widehat{\boldsymbol{\alpha}}} &= \left[\mathbf{D}' + \widehat{\eta}_0 \mathbf{D}'\mathbf{W}\right]\mathbf{e} = \mathbf{0}
\end{aligned}$$

Estimating a wage regression equation accounting for peer effects

- *Step 1* Given $\widehat{\alpha}$ run an OLS regression on $\mathbf{X}$, $\mathbf{D}\widehat{\alpha}$, and $\mathbf{WD}\widehat{\alpha}$. The coefficients on $\mathbf{X}$ will provide an estimate of β, while the coefficient on $\mathbf{WD}\widehat{\alpha}$ is an estimate of η_0. $\mathbf{D}\widehat{\alpha}$ should have a coefficient of 1.
- *Step 2* Given $\widehat{\beta}$ and $\widehat{\eta}_0$ estimate α using the updating equation

$$\widehat{\alpha}_{[h]} = \mathbf{M_D}\left[\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\left[\mathbf{Y} - \mathbf{X}\widehat{\beta}\right] - \widehat{\eta}_0\mathbf{M_D}\left[2\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{WD}\widehat{\alpha}_{[h-1]} \quad (11)$$

There is, however, a faster approach to solve the f.o.c. $\left[\mathbf{D}' + \widehat{\eta}_0\mathbf{D}'\mathbf{W}\right]\mathbf{e} = \mathbf{0}$. We can then rewrite the equation as

$$\mathbf{D}'\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}\mathbf{D}\widehat{\alpha} = \mathbf{D}'\widetilde{\mathbf{W}}\left[\mathbf{Y} - \mathbf{X}\widehat{\beta}\right]$$

and since this is now written as a system of linear equations we apply the conjugate gradient method to obtain a solution for $\widehat{\alpha}$.

Estimating the standard errors from a wage regression equation accounting for peer effects

As shown in Davidson and Mackinnon (2004), once we obtain the NLS estimates for the parameters $(\boldsymbol{\beta}^o, \eta_0^o, \boldsymbol{\alpha}^o)$, we can easily estimate the corresponding variance-covariance matrix.
The idea consists of using the associated Gauss-Newton regression (GNR).
After some simplication, the GNR becomes,

$$\mathbf{y} + \eta_0^o \mathbf{W}\mathbf{D}\boldsymbol{\alpha}^o = \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} + \eta_0^o \mathbf{W}]\mathbf{D}\boldsymbol{\alpha} + \eta_0 \mathbf{W}\mathbf{D}\boldsymbol{\alpha}^o + \varepsilon \tag{12}$$

The estimation of this linear regression is complicated by the inclusion of the regressors $[\mathbf{I} + \eta_0^o \mathbf{W}]\mathbf{D}$ as well as other high-dimensional fixed effects.
Since this is a linear regression we can take advantage of the Frisch-Waugh-Lovell theorem and partial out the effects of all high dimensional variables including the set of variables $[\mathbf{I} + \eta_0^o \mathbf{W}]\mathbf{D}$.

The role of peers

Table: Wage Equation Accounting for Human Capital Spillovers

HC spillovers ($\overline{\alpha}_{-it}$)	0.2050 (0.0006)
Worker effects (α_i)	✓
Establishment/Job-title/Year effects	✓

Note: Regression includes a quadratic term on age and tenure, gender, and firm size as covariates.

Some moments of the empirical joint distribution of fixed effects

Table: Statistical Moments from Wage Distribution - firm/job-title/year specification

Panel A - Variance Decomposition	
α_i - worker	0.2780
$\eta_0 \overline{\alpha}_{-it}$ - coworker	0.1087
$\theta_{\mathbf{F} \times \mathbf{J}(i,t)}$ - firm/job-title	0.4546
$Z_{it}\widehat{\gamma}$	0.1171
Panel B - Correlations	
$\rho(\alpha_i, \overline{\alpha}_{-it})$	0.7060
$\rho(\alpha_i, \theta_{\mathbf{F} \times \mathbf{J}(i,t)})$	0.1366
$\rho(\overline{\alpha}_{-it}, \theta_{\mathbf{F} \times \mathbf{J}(i,t)})$	-0.0133
Panel C - Fixed Effect Heterogeneity	
σ_{α_i}	0.2455
$\sigma_{\overline{\alpha}_{-i}}$	0.2193
$\overline{\sigma}_{\alpha_{i\mathbf{F} \times \mathbf{J}(i,t)}}$	0.1139
$\sigma_{\theta_{\mathbf{F} \times \mathbf{J}(i,t)}}$	0.3530

Decomposing the returns to education, a second pass

Table: Conditional Decomposition of the Return to Education - OLS

Base	Individual	Est./Job/Year	HC Spillovers
0.0791	0.0313	0.0422	0.0056

Note: Decompositions based on Gelbach (2016).

Table: Conditional Decomposition of the Return to Education - IV

Base	Individual	Est./Job/Year	HC Spillovers
0.0774	0.0236	0.0481	0.0057

Note: Decompositions based on Gelbach (2016).

Using Occupation instead of Job-Title

Table: Wage Equation with Human Capital Spillovers - Occupations

	IV	
	Base	Full
Schooling	0.0688	0.0228
	(0.0001)	(0.0001)
HC spillovers ($\overline{\alpha}_{-it}$)	-	0.3085
	-	(0.0012)
Time effects	✓	
Worker effects		✓
Establishment/Occupation/Year effects		✓

Using a College Dummy instead of Years of Schooling

Table: Wage Equation with Human Capital Spillovers - College

	IV	
	Base	Full
College	1.1689	0.0940
	(0.1214)	(0.0395)
HC spillovers ($\overline{\alpha}_{-it}$)	-	0.2035
	-	(0.0012)
Time effects	✓	
Worker effects		✓
Establishment/Job-title/Year effects		✓

Robustness: Covariates, peer definition, and fixed effects

Table: Sensitivity of the Human Capital Spillovers

	HC spillovers ($\overline{\alpha}_{-it}$) (1)	(2)	(3)	(4)	(3')
Full Specification	0.4864 (0.0006)	0.3556 (0.0005)	0.2035 (0.0012)	0.1484 (0.0008)	0.3085 (0.0012)
Excluding Covariates	0.4394 (0.0006)	0.3417 (0.0005)	0.1984 (0.0012)	0.1457 (0.0008)	0.3165 (0.0011)
Time effects (μ_t)	✓		✓		
Worker effects (α_i)	✓	✓	✓	✓	✓
Establishment/Job-title effects ($\theta_{\mathbf{F}\times\mathbf{J}(i,t)}$)	✓			✓	
Establishment/year effects		✓		✓	
Job-title/year effects		✓		✓	
Establishment/job-title/year effects			✓		
Establishment/occupation/year effects					✓

Accounting for co-worker education

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta s_i + a_i + \eta_1 \overline{s}_{-it} + \eta_0 \overline{a}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad , \tag{13}$$

where we are separating schooling (s_i) from the other covariates ($\mathbf{z}_{it}$). Here $\overline{s}_{-it}$ is the average education of the coworkers of worker i at time t, and $\overline{a}_{-it}$ is the equivalent measure for ability. The η parameters are the associated coefficients.

The above equation can be written equivalently as,

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + s_i(\delta - \omega) + (\eta_1 - \eta_0\omega)\overline{s}_{-it} + \alpha_i + \eta_0 \overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad . \tag{14}$$

where ω can be any real value and the worker fixed effect, α_i, is obtained as $\alpha_i = s_i\omega + a_i$. In this setting, $\overline{\alpha}_{-it}$ can be interpreted as a measure of coworker quality.

Accounting for co-worker education

The above equation remains overparameterized and some restrictions are needed to make it identifiable.

Arcidiacono et al. (2012) impose the restriction that both coefficients on s_i and $\overline{s}_{-it}$ are zero ($\delta = \omega$ and $\eta_1 = \eta_0 \omega$), which amounts to assuming that the importance of own characteristics is proportional to that of coworker characteristics ($\eta_1 = \eta_0 \delta$).

Although convenient, the imposition of these two conditions is unnecessarily restrictive.

In the analysis that follows we report results for the following specification:

$$y_{it} = \mathbf{z}_{it}\gamma + \delta^* s_i + \alpha_i + \eta_0 \overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad . \tag{15}$$

New Decomposition Accounting for Peer's Education

	Base	Individual	Est./Job/Year	HC Spillovers
Schooling	0.0411	0.0181	0.0203	0.0026
Peer Schooling	0.0574	0.0198	0.0330	0.0045

- 0.0198 - reflection effect (homophily). More educated co-workers mirror the individual effect (illusion).

New Decomposition Accounting for Peer's Education

	Base	Individual	Est./Job/Year	HC Spillovers
Schooling	0.0411	0.0181	0.0203	0.0026
Peer Schooling	0.0376	-	0.0330	0.0045

- 0.0181 - individual effect (including abiliy) - unrelated with sorting in jobs and establishments, and unrelated with peers.
- 0.0203 - more educated workers sort into firms with more generous wage policies and better paid job titles
- 0.0026 - more educated workers benefit from matching with higher skilled peers

New Decomposition Accounting for Peer's Education

	Base	Individual	Est./Job/Year	HC Spillovers
Schooling	0.0411	0.0181	0.0203	0.0026
Peer Schooling	0.0376	-	0.0330	0.0045

- 0.0045 - human capital spillover effect. One additional year of the mean education of peers increases individual wages by 0.5 percent - unrelated with sorting in jobs and establishments.
- 0.0330 - more educated peers match with firms with more generous wage policies and better paid job titles.

Alternative interpretation based on equation

$$y_{it} = \mathbf{z}_{it}\gamma + s_i(\delta - \omega) + (\eta_1 - \eta_0\omega)\overline{s}_{-it} + \alpha_i + \eta_0\overline{\alpha}_{-it} + \theta_{\mathbf{P}(i,t)} + \varepsilon_{it} \quad . \quad (16)$$

- Under the assumption of no external returns, the implied own return to education would be a meager 0.3 log point.
- Assuming a 6 log points return to education, the model would imply a 1.2 log points return to coworker education.
- For a 0.5 log point external return, the own return to education would be 2.7 log points.
- etc.

Take away (I)

- We decompose the (Mincerian-type) 7.9 percent return per year of schooling.
- Acccounting for the endogeneity of education (via IV) decreases the return to education by just 0.2 percentage points.
- Labour Market Sorting is very important
    - Allocation of better educated workers into better paying firms and job titles accounts for three fifths of the overall conditional return on education. (2.0 percent to firm allocation and 2.7 to job title allocation).

Take away (II)

- **1.8 percent** of the return is associated exclusively with the **individual**
  - unrelated with sorting in jobs and establishments, and unrelatd with peers

- Peers education directly influences wages: **Human capital spillovers**
  - one additional year of peer education increases individual wages by **0.5 percent**
  - unrelated with sorting in jobs and establishments

- **Peer effects have a strong impact on wages**:
  - a 10 percent increase in the measure of peer quality leads to a wage increase of around 2 percent.

New Decomposition Accounting for Peer's Education

	Base	Individual	Est./Job/Year	HC Spillovers
Schooling	0.0411	0.0181	0.0203	0.0026
Peer Schooling	0.0376	-	0.0330	0.0045

- by memory: standard equation with peers decomposition

	Base	Individual	Est./Job/Year	HC Spillovers
Schooling	0.0791	0.0283	0.0422	0.0057

Note: Decompositions based on Gelbach (2016).

IV - First Stage

Table: First Stage - Schooling

	(1)
3 years Mandatory Schooling Cohort (born year 1948 to 1950) (Male)	2.035
	(0.0976)
3 years Mandatory Schooling Cohort (born year 1948 to 1950) (Female)	0.9717
	(0.0689)
4 years Mandatory Schooling Cohort (Female)	0.0260
	(0.0445)
4 years Mandatory Schooling Cohort (Male)	0.7675
	(0.0521)
6 years Mandatory Schooling Cohort (All)	0.2886
	(0.0334)
Time effects (μ_t)	✓
F-Test	146.18
	(0.0000)
N	19,051,268
R Squared	0.2173

BPlim

- The Bank of Portugal is launching a new research unit to promote the analysis of micro data.
  - The researchers of the unit are specialized in the production of micro econometric estimators and techniques.
  - The datasets used and produced at the Bank of Portugal are going to be available to the international academic community.
  - Through a network of powerful virtual computers (sand boxes) that can be accessed anywhere in the world.
  - Provided that a research project was submitted and approved by the head of the research unit.
  - Issues of statistical secret are dealt on a case by case basis.
  - Portuguese micro data is very, very rich.

Estimation and Identification

- Estimation:
- Zigzag algorithm - Guimarães and Portugal (2010)

- Identification:
- Driven by the entry and exit of workers into particular job-titles within firm - changes in the composition overt time
  - isolating the endogeneous sorting of workers into firm/job-title cells