The returns to schooling unveiled

Ana Rute Cardoso
IAE (CSIC)
Barcelona GSE

Paulo Guimarães
Banco de Portugal
Universidade do Porto

Pedro Portugal
Banco de Portugal
NOVA SBE

Hugo Reis
Banco de Portugal
Católica Lisbon School of Business & Economics

March 2018

Acknowledgements: We thank the Portuguese Ministry of Employment, Statistics Department, for access to the data, Lucena Vieira for very competent research assistance, and the comments by seminar participants at the University of Oslo and the Banco de Portugal, as well as conference participants at SOLE-EALE World Conference in Montreal. Cardoso acknowledges the financial support from the AEI-Spain and FEDER-UE (grant ECO2016-76734-P), the Spanish Ministry of the Economy and Competitiveness (Severo Ochoa Programme for Centres of Excellence in R&D, SEV20150563) and the Government of Catalonia (grant 2017SGR1359).

E-mail: ; pfguimaraes@bportugal.pt; pportugal@bportugal.pt; hfreis@bportugal.pt

Abstract
We bring together the strands of literature on the returns to education, its spillovers, and the role of the employer shaping the wage distribution. The aim is to analyze the labor market returns to education taking into account *who* the worker is (worker unobserved ability), *what* he does (the job title), *with whom* (the coworkers) and, also crucially, *for whom* (the employer). We combine data of remarkable quality — exhaustive longitudinal linked employer-employee data on Portugal — with innovative empirical methods, to address the homophily or reflection problem, selection issues, and common measurement errors and confounding factors. Our methodology combines the estimation of wage regressions in the spirit of Abowd, Kramarz, and Margolis (1999), Gelbach's (2016) unambiguous conditional decomposition of the impact of various omitted covariates on an estimated coefficient, and Arcidiacono et al.'s (2012) procedure to identify the impact of peer quality. We first uncover that peer effects are quite sizeable. A one standard deviation increase in the measure of peer quality leads to a wage increase of 2.1 log points. Next, we show that education grants access to better-paying firms and job titles: one fourth of the overall return to education operates through the firm channel and a third operates through the job-title channel, while the remainder is associated exclusively with the individual worker. Finally, we unveil that an additional year of average education of coworkers yields a 0.5 log points increase in a worker's wage, after we net out a 2.0 log points return due to homophily (similarity of own and peers' characteristics), and 3.3 log points associated with worker sorting across firms and job titles.

JEL: J31; J24; I26

Keywords: wage distribution; human capital spillovers; returns to education; peer effects; linked employer-employee data; high-dimensional fixed effects; firm; job title.

1. Introduction

The literature on the returns to schooling has made remarkable progress over the last 50 years; nevertheless, the role of worker sorting across employers or jobs has been neglected. It is time to redress this omission, given that both theoretical models and empirical analyses have highlighted the importance of the demand side of the market shaping the wage distribution. Indeed, we are now equipped with several explanations about why firms may find it profitable to deviate from a market-wide wage standard: efficiency wages, implicit contracts, rent-sharing, principal-agent models, and the frictions contemplated in search and matching models. To the extent that workers with different levels of education are not randomly allocated to firms and firms' pay standards are heterogeneous, they could be a key factor channeling the returns to education. A similar argument could be built over job titles. There are remarkable wage contrasts across narrow occupations, possibly driven by differences in their degree of riskiness, the amount of specific training required, or the technology used. Provided that education can grant a "passport" to better paying job titles, part of the overall return on education would operate through a job title channel. The first aim of the current analysis is therefore to pinpoint the role of firm- and job-level heterogeneity shaping the returns to schooling. We will quantify the impact of sorting of workers across firms and job titles on the returns to education.

Analysis of the role of firm- and job-level heterogeneity structuring the returns to education begs another question: What if peers matter? Fundamentally, the quality of a firm will depend on the quality of its human resources. Are there spillovers of education operating within the firm and, possibly, the job title? This fits into the controversial line of investigation on the social returns to education, which has seen little evidence reported at the firm level (Battu, Belfield, and Sloane, 2003; Wirz, 2008; Martins and Jin, 2010; Nix, 2016). Taking the city or region as the unit in which education spillovers could operate, the available results diverge widely. Whereas Acemoglu and Angrist (2000) failed to find evidence on external returns to education, Ciccone and Peri (2006) found negative spillovers and Moretti (2004b, 2004c), instead, reports on significant positive impacts of graduates on the wages of workers in the same city. In the meantime, Manski (1993) and Angrist (2014) have called attention to problems that can potentially plague the analysis of the impact of peers, namely: the "reflection problem", sorting of workers across firms and detailed occupations, and other confounding factors and measurement errors.

The lines of research on wage heterogeneity across employers and the returns to education and its spillovers have remained unarticulated. We perform the first analysis of the returns to education and its spillovers at the firm level accounting for the unobservable quality of the worker, his peers, and worker assortative matching to firms and job titles. We thus adopt a multifaceted approach to explore the labor market returns to education, taking into account *who* the worker is (worker unobserved ability), *what* he does (the job title), *with whom* (the coworkers) and, also crucially, *for whom* (the employer). To do so, we rely on innovative empirical methods, propelled by remarkable data quality.

We start out with the estimation of wage regressions augmented to include sets of high-dimensional fixed effects, in the spirit of the seminal work of Abowd, Kramarz, and Margolis (1999) (hereinafter AKM). We account for firm/job-title and worker unobserved time-invariant heterogeneity. We then adapt Gelbach's (2016) unambiguous conditional decomposition of the impact of various omitted covariates on an estimated coefficient, to quantify how much of the return to education operates through a firm and a job title channel, as opposed to a worker individual channel. The exercise undertaken can be interpreted very intuitively taking the example of the firm channel —it brings to light differences in firm wage effects across schooling levels. Therefore, it quantifies the relevance of worker sorting across firms in shaping the returns to education.

We then progress to explicitly acknowledge that work within a firm or a job title is not undertaken in isolation, but with coworkers. Peers may matter. Hence, we enrich the analysis along two dimensions. We account for the level of education of coworkers performing the same task. The underlying idea is that there may be a greater scope for spillovers within the same job title than across all job levels, in the spirit of the economics of education literature, which reports stronger peer effects at the classroom level than at the school level. In the terminology of Acemoglu (2014) and Moretti (2004b), we are concerned with technological productivity spillovers of education operating within the firm. At the same time, we account for peers' unobserved quality. The state-of-the-art empirical procedure to account for peers' unobservable attributes is that of Arcidiacono et al. (2012). While relying on their procedure to estimate the impact of peers' unobserved quality on a worker's wage, we relax their assumption of constant proportionality between the returns on own and on peers' attributes, whether observed or unobservable. We free the relationship between the returns on own education and on peers' education, and estimate it. We present the estimation method and discuss its assumptions.

We use longitudinal data on the population of firms and workers in the Portuguese economy. This enables us to observe the entire distribution of characteristics and outcomes of both the individual and all of his coworkers. As a result, we can set the analysis of knowledge spillovers at the workplace level, where interactions among workers are more intense. We can also exploit time-series variation in the composition of the peer groups, as we rely on two decades worth of information, tracking workers as they change jobs. We rely on unique coding of job titles, a finer classification than detailed occupations, which considers the complexity of tasks performed and the degree of responsibility, and is used to define wage floors in collective bargaining. Note that, in any case, the firm has margin for maneuver to set wages above those floors —Cardoso and Portugal (2005) indeed report evidence on the upward adjustment of wages according to firm level conditions, for selected groups of workers. Crucially, our results on the impact of peers are not restricted to a narrow set of occupations or industries and are likely to be representative of the economy at large.

Our linked employer-employee dataset is valuable for additional reasons. First, it reports the schooling of the worker. The combination of full coverage of the economy with data on schooling had never before been within the reach of researchers (Nix, 2016 is the exception, as she relies on Swedish data, though restricting her analysis to a sample of male workers). Secondly, we have accurate information on hours worked, as well as a control variable on whether the worker's earnings refer to full schedule and full earnings during the month. Therefore, we can undertake an analysis of hourly wages, which are less contaminated by measurement error than labor earnings, again unlike the earlier studies on peer effects at the firm level. Finally, our earnings data are not subject to any type of censoring.

Our main contribution lies at the intersection of the literature on education spillovers and that on employer wage policies, addressing the biases that could result from homophily in peer group formation and from worker assortative matching to firms and job titles. The remarkable quality of the dataset overcomes common sources of measurement error.

Section 2 provides an overview of the literature on wage heterogeneity across employers and education spillovers. Section 3 describes the institutional setting in the Portuguese labor market, followed by the section on the data used. Section 5 presents the methodology and results on the impact of worker sorting across firms and job titles structuring the returns to education. We adapt the procedure by Gelbach (2016) to decompose the returns to education into a firm, a job title, and an individual worker component. Section 6, in turn, presents the methodology and our estimation results on the returns to peers' and own education, duly accounting for worker and peers' unobservable quality, firm, and job quality. Section 7 concludes.

2. Returns to Education: Current Evidence on the Role of the Employer and the Peers

As early as the 1950s there was evidence that firms may find it profitable to deviate from a market-wide wage standard. Early case studies by Lester (1952) and Reynolds (1951) have shown that employers' pay standards vary widely, even within narrowly defined regions and industries. Later, Groshen (1991a) documented a large contribution of the employer to intra-industry wage differentials. Machin and Manning (2004) corroborated the idea that wages are far from competitive, as they documented high wage dispersion across firms within a narrowly defined occupation and geographic area, despite the operation of a large number of firms delivering a homogeneous good. AKM started a very prolific line of literature that explores large longitudinal linked employer-employee datasets to quantify firm effects on wages. This strand now includes Gruetter and Lalive (2009), Eeckhout and Kircher (2011), Card, Heining, and Kline (2013), Torres et al. (2013), and Lopes de Melo (2018), among several others'. Card et al. (2018) summarize the literature on the role of the firm in wage dispersion. Recent developments in this empirical line of research further evaluates the impact of peers' quality on a worker's wage (see Cornelissen, Dustmann, and Schönberg, 2017; Battisti, 2013; and Lopes de Melo, 2018, on the returns to peer unobserved quality at the firm level).[1] This empirical literature has remained silent on the impact of employer policies on the returns to schooling specifically (see the overviews by Card, 1999, 2001; Blundell, Dearden, and Sianesi, 2005; and Belzil, 2007). The exception is the recent article by Engbom and Moser (2017), who, however, relied on a dataset restricted to higher education graduates and accounted in their analysis for firm unobserved heterogeneity, but not workers'. Likewise, the role of the job has also been neglected when studying the returns to education, despite early concern about the impact that the introduction of controls for broad occupation might have on the estimates.

A separate strand of literature has dealt with spillovers of education. This line of research is characterized by a heated debate, so far unsettled. From a microeconomic perspective there are mainly two broad branches of literature explaining how and why positive external returns to education may arise. They are highlighted by authors like Acemoglu and Angrist (2000) and summarized by Moretti (2004a). The first is a theory of non-pecuniary external returns (technological spillovers), according to which the external returns arise from technological linkages across agents or firms. The second is a pecuniary model of external returns, in which spillovers arise from market interactions and changes in market prices resulting from the average education level of the workers. Furthermore, firms may deliberately cultivate a 'team dynamic', with information-sharing, co-training, monitoring, and support, in order to exploit these spillovers.

The key idea in technological spillovers is that the exchange of ideas among workers raises productivity. Marshall (1890) was the first to argue that social interactions among workers in the same industry and location create learning opportunities that enhance productivity. More recent literature on human capital externalities has built on Marshall's insight. An influential paper by Lucas (1988) suggests that human capital spillovers may help explain differences in long run economic performance of countries. In his work, the knowledge diffusion through formal and informal interaction is viewed as the

1. The impact of the employer determining the gender pay gap has received the attention of Groshen (1991b), Blau (1997), Meng and Meurs (2004), Card, Cardoso, and Kline (2016) and Cardoso, Guimarães, and Portugal (2016).

channel that generates positive spillovers across workers.[2] A recent paper by Martins and Jin (2010) presents a model of learning in which workers learn from the human capital of their colleagues in the same firm. Their model shows a stronger relationship between wages and education at the firm level than at the individual level, at least in non-monopsonistic labor markets.

In turn, the key idea in pecuniary spillovers is that human capital encourages more investment by firms and raises other workers' wages. In particular, human capital spillovers may arise if human and physical capital are complements even in the absence of learning or technological spillovers. The best example is provided by Acemoglu (1996), in which job search is costly and education spillovers are present due to the complementarity between physical and human capital. The complementary mechanism is discussed in the context of labor market imperfections, innovation investment by firms, and training by workers. Human capital externalities arise here because firms choose their physical capital in anticipation of the average human capital of the workers they will employ in the future.

In contrast, in theoretical terms, coworker education may also have negative spillovers. First, in signaling or screening models of education, education may be associated with negative externalities (Spence, 1973), as it enables some workers to access better firms, while excluding others, who are thus left unemployed or engaged in poorer firms (Moen, 1999). Furthermore, coworkers' education may also have negative spillovers if high and low-skill workers are imperfect substitutes (e.g., Moretti, 2004a; Ciccone and Peri, 2006), or if workers compete for promotions and do not share their human capital. In particular, if workers are in competition with each other for high-paying jobs within the firm, they may engage in sabotage activities to undermine their coworkers and promote themselves. If coworkers have different amounts of human capital, then there may be a "skills incompatibility" problem (Kremer, 1993). In this case, within an O-ring type of model, a firm with a uniform standard of education may have higher productivity than one where both average education levels and the spread of education are high.

Finally, in terms of policy implications, it should be emphasized that not every productivity spillover is necessarily a market failure requiring government intervention (Moretti, 2010). Spillovers that occur within a firm, for example, can in principle be internalized. For example, low productivity workers may benefit from the presence of more capable workers, while the productivity of high-skilled workers may not be hurt by the presence of low-skilled coworkers. This type of spillover could be internalized by the firm by raising the compensation of highly productive workers to reflect their external benefit on the productivity of less productive colleagues.

2. This notion of human capital externalities is also present in the works of Jovanovic and Rob (1989) and Glaeser (1999).

In the empirical literature there is no consensus on the presence of education spillovers. Most of the existing evidence for human capital externalities relies on US estimates of the effects of regionally aggregated human capital on individual wages. For example, Rauch (1993), Acemoglu and Angrist (2000), Moretti (2004b, 2004c), and Ciccone and Peri (2006) have taken the city or region as the main unit of analysis at which education spillovers could operate. Acemoglu and Angrist (2000) do not find significant external returns, Ciccone and Peri (2006) show evidence of negative spillovers, while Moretti (2004b, 2004c) reports significant positive impacts of graduates on the wages of workers in the same city. Rauch (1993) points to positive but small effects on wages.[3]

At the firm level, despite the importance of the topic, the empirical evidence is relatively scarce, especially regarding the specific effect of coworkers' education. There are, nevertheless, three noteworthy exceptions. Battu, Belfield, and Sloane (2003) find significantly positive effects in a cross-section study of British establishments, proxying firm average education from the distribution of workers across different occupations.[4] Martins and Jin (2010) estimate social (firm-wide) returns to education using Portuguese data. Wirz (2008) also finds positive and significant external returns for the Swiss economy. Her work is closest in spirit to ours, even though she relies on one cross-section of sample data and on a two-stage estimation process to identify the returns to own and peers' education, accounting for firm effects.[5]

3. Wage Setting

A national minimum wage is enforced in Portugal, defined as a monthly rate for full-time work. Currently, sub-minimum wage levels apply only to physically disabled workers and trainees, after all reductions based on age were abolished in 1999.

Collective bargaining plays a central role in the Portuguese labor market, as in several other continental European economies. Indeed, massive collective

3. Rauch (1993) does not take into account the endogeneity of location choices, arguably leading to an upward bias in the estimates.

4. Kirby and Riley (2008) look at this problem but at the more aggregated level of the industry.

5. A different strand of literature, on peer effects in the workplace, analyzes the contemporaneous impact of coworkers' behavior on an individual's productivity. These studies focus mainly on two specific channels (effort and the team dynamics), based on small datasets and very specific settings (sectors and tasks) or laboratory experiments. For example, Falk and Ichino (2006) report on an experiment over a task outting letters into envelopes, Mas and Moretti (2009) study workers in one large supermarket chain, and Bandiera, Barankay, and Rasul (2010) look at soft-fruit pickers in one large U.K. farm. For an interesting summary of the empirical results in the more general literature on the impact of peer effects on worker output, see Herbst and Mas (2015). However, these works do not examine the returns to education.

agreements, often covering an industry, are common in the economy. Firm level collective bargaining traditionally covers a low share of the workforce, less than 10%. Extension mechanisms are common, either by mandatory government regulation or on a voluntary basis, as employers automatically apply the contents of collective agreements to their non-unionized workforce.

Despite the relevance of collective bargaining, firms have always enjoyed some degree of freedom in wage setting. Cardoso and Portugal (2005) have documented that wage cushion (the difference between the actual wage level and the bargained wage level) promotes an alignment of wages with firm-level conditions. They show that once mandatory contract wages have been set, firm-specific arrangements stretch the returns to worker and firm attributes and shrink the returns to union power. The existence of wage cushion therefore leaves ample scope for firms to define distinct wage policies. It follows from such an institutional setting that it is of key interest to quantify the impact of the firm when estimating the returns to education.

Collective agreements set wage floors for very disaggregated job titles. In a typical year, around 300 collective agreements signed by trade unions and employers associations are enforced, determining minimum wages for around 30,000 job titles (see Carneiro et al. 2012). To take an illustrative example, in the ship building industry, there is a distinction between painters of the starboard and the port side of the ship.[6] Furthermore, note that under this definition of job title, two workers with the same job description (i.e. performing the same tasks and having the same responsibilities) covered by different bargaining agreements will often have different job titles. This level of detail is, of course, much more granular than the conventional occupation classification (on average, we have 15 job titles per occupation in our dataset). We take advantage of such an unusually fine accounting of the tasks to fill a job to determine the boundaries of the (highly homogeneous) peer group.

4. Data Source and Concepts Used

Quadros de Pessoal (QP) is an unusually rich and comprehensive linked employer-employee dataset, gathered annually by the Ministry of Employment. It covers all establishments having at least one wage earner. The wage information is collected with reference to the month of October. Civil servants, self-employed, and household employees are not covered; the share of wage-earners in agriculture is low and therefore the coverage of this sector is low. Instead, for manufacturing and the services private sector of the economy, the survey covers virtually the entire population of workers and firms.

6. It seems that the reasoning for the distinction relies upon the risk of falling in the water or on the ground.

The following variables are reported on each worker: gender, date of birth, schooling, occupation, date of hire into the firm, monthly earnings, hours of work, the collective bargaining agreement, and the worker's job title ("*categoria profissional*") in that agreement. The schooling information refers to the highest completed level of education. Information on the employer includes the industry and location. In the current exercise we use information stretching from 1994 to 2013. However, no worker data are available for 2001.

We have restricted the analysis to workers aged 16 to 64, reporting working full-time in the non-agricultural sectors, with at least 120 monthly hours of work, who are not apprentices and whose base wage does not fall below the national minimum wage, with non-missing schooling, and reported job duration between 0 and 600 months. To assure that our job title definition is meaningful we dropped observations that are not assigned to any collective agreement and job titles that are defined as residual categories. Furthermore, to assure that coworkers share the same workplace we dropped workers in industries that provide services to other firms mainly through outsourcing (e.g., cleaning and security industries).

Moreover, to separately identify firm/job-title and worker fixed effects, the analysis must be restricted to the set of firms that are connected by worker mobility (see the discussion in Abowd, Creecy, and Kramarz, 2002). We therefore limit our analysis to the largest connected set of observations defined as connected for two fixed effects. The largest dataset under analysis comprises 19.1 million observations on 3.7 million workers, 282.6 thousand firms, and 82.5 thousand job titles in collective bargaining.

Given the purpose of our analysis, we employ a rather strict definition of peers. The aim is to guarantee that workers share the same workplace and the same task. So workers belong to a given peer group if, in a given year, they have a common job title and establishment. Given our interest in quantifying the human capital spillovers, we of course restrict the analysis to peer groups with at least two workers. In total, we consider 3.9 million peer groups with an average of 4.9 workers per peer group; this also means that we have 14.0 peer groups by firm, and 47.8 by job title.

Hourly wages are computed as the actual overall monthly earnings (including base wage, tenure-related and other regularly paid components) over the number of normal hours of work. Wages were deflated using the consumer price index (base 2013), but this correction is inconsequential since we always include year dummies in the regression analysis. Table A.3.1 in the Appendix A.3 presents the descriptive statistics for the variables used in the estimation.

5. Sorting of Workers across Firms and Job Titles and the Returns to Education

5.1. Wage regressions with worker and firm/job-title effects

We begin by specifying the linear wage equation that serves as the basis for most of our analysis. As in AKM, we rely on a standard Mincerian wage equation with two high-dimensional fixed effects. However, our controls are tighter than AKM's given that in addition to a worker fixed effect we define a fixed effect for each unique combination of job title and firm (firm/job-title effect).[7] By controlling for firm/job-title and worker fixed effects we are able to control for unobservables that capture a substantial amount of wage variation, while at the same time mitigating potential endogeneity problems. More specifically, we consider an equation of the type,

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

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; $\theta_{\mathbf{F}\times\mathbf{J}(i,t)}$ is a unique firm/job-title specific time-invariant fixed effect; μ_t are time fixed effects; and ε_{it} is the disturbance term of the regression.[8] We assume strict exogeneity, $E(\varepsilon_{it}|\mathbf{x}_{it}, \alpha_i, \theta_{\mathbf{F}\times\mathbf{J}(i,t)}, \mu_t) = 0$, to ensure unbiasedness of all regression coefficients. 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 (log of number of employees). Both gender and worker education, our variable of interest, are time invariant and are explicitly accounted for only in specifications that omit the worker fixed effect.

Estimation of equation (1) by ordinary least squares (OLS) is complicated by the fact that it includes two high-dimensional fixed effects. As discussed in AKM, the large dimension of the design matrices for the fixed effects makes impractical the application of the conventional OLS formula. Fortunately, models of this type can be estimated using, for example, the algorithm proposed by Guimarães and Portugal (2010). This algorithm can easily be extended to deal with more than two high-dimensional fixed effects.[9] In its basic version, it

7. AKM's classic specification controls for firm and worker fixed effects. However, the introduction of a firm/job-title fixed effect produces the same fit as a model that separately adds a fixed effect for firm, another for job title, and a third for the interaction of job title and firm. Thus, our specification nests AKM's as a particular case.

8. The parentheses in the subscripts of the fixed effects coefficients are used to emphasize that the ultimate source of variation stems from the worker/time combination.

9. The *Stata* user-written package *reghdfe* coded by Sergio Correia and available on the Statistical Software Components (SSC) Boston Archive implements a modified version of the algorithm, which allows for efficient estimation of models with multiple high-dimensional fixed effects (Correia, 2014).

consists of an iterative procedure that alternates between the estimation of the fixed effects (taking as given the last estimates of the $\boldsymbol{\beta}$) and estimation of $\boldsymbol{\beta}$ (taking as given the last estimates of the fixed effects). The algorithm converges to the true OLS solution. There is, however, an additional complication that arises in models with more than one high-dimensional fixed effect. The likely existence of perfect multicollinearity between parameters associated with the fixed effects may introduce problems of identification. This may not be an issue if interest centers on the $\boldsymbol{\beta}$ coefficients, but in our case we also want to implement secondary analysis of the estimates of the αs and θs. Interpretation of the estimates of the fixed effects is meaningful only if the differences between coefficients (within each fixed effect) are estimable. As mentioned above, to guarantee identification and thus ensure comparability of the parameter estimates, we restrict our analysis to the largest subset of data in which all the fixed effects are connected.[10]

5.2. Gelbach's decomposition

To understand the contribution that the allocation of workers to firms and jobs has to the observed education pay differential we make use of Gelbach's (2016) decomposition method. His approach is based on the OLS formula for omitted variable bias and allows for a decomposition that unequivocally quantifies the portion of the variation attributed to each variable of interest. Gelbach's decomposition is easier to present if we resort to matrix notation. Consider a conventional Mincerian equation that includes the observable characteristics of firms and workers as well as time effects. For convenience we collect the observations for all variables but worker schooling, into the matrix $\mathbf{Z}$. Our variable of interest, schooling, is introduced separately and represented by the variable $\mathbf{S}$ (where $\mathbf{S} \equiv \mathbf{Ds}$ and $\mathbf{s}$ is a vector with dimension N containing each worker schooling level while $\mathbf{D}$ is the design matrix for workers). Thus, we have

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma_0} + \delta_0\mathbf{S} + \boldsymbol{\varepsilon} \quad . \tag{2}$$

By the Frisch-Waugh-Lovell theorem we know that the same OLS estimate of δ_0 may be obtained by running a simple regression of $\mathbf{Y}$ on $\mathbf{S}$ after partialing out the effect of $\mathbf{Z}$ from both variables. More specifically,

$$\widehat{\delta_0} = (\mathbf{S}'\mathbf{M_Z}\mathbf{S})^{-1}\mathbf{S}'\mathbf{M_Z}\mathbf{Y} = \mathbf{P_Z}\mathbf{Y} \quad , \tag{3}$$

10. We use the algorithm of Weeks and Williams (1964) to identify a connected set. This algorithm can be applied when dealing with two or more sets of fixed effects and will produce the same result as the algorithm described in Abowd, Creecy, and Kramarz (2002) if applied to a model with two high-dimensional fixed effects. The largest mobility group accounted for over 98% of our original data set, thus rendering negligible possible concerns about sample selection bias.

where $\mathbf{M_Z} \equiv \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the well-known symmetric and idempotent residual-maker matrix. Here $\widehat{\delta_0}$ is the conventional OLS estimator used to produce estimates for the returns to education. To show how Gelbach's decomposition can be used to tease out the contribution of the firm and job title fixed effects on the returns to education, consider now a full regression to which we have added these two sets of fixed effects: worker ($\boldsymbol{\alpha}$) and firm/job-title fixed effects ($\boldsymbol{\theta}$). This regression, written in terms of its fitted OLS expression, is:

$$\mathbf{Y} = \mathbf{Z}\widehat{\boldsymbol{\gamma}} + \mathbf{D}\widehat{\boldsymbol{\alpha}} + \mathbf{L}\widehat{\boldsymbol{\theta}} + \mathbf{e} \quad . \tag{4}$$

The education variable has to be dropped from this specification because the variable is time-invariant and thus its effect is fully absorbed by the worker fixed effect.[11] Note also that $\mathbf{D}\widehat{\boldsymbol{\alpha}}$ and $\mathbf{L}\widehat{\boldsymbol{\theta}}$ are column vectors containing the least-squares estimates for the worker and firm/job-title fixed effects in a regression that also controls for $\mathbf{Z}$. To obtain a decomposition of $\widehat{\delta_0}$ we multiply both terms of equation (4) by $\mathbf{P_Z}$. In other words, we regress each element of the above equation on education while controlling for the remaining observable variables ($\mathbf{Z}$). On the left-hand side we obtain $\widehat{\delta_0}$ directly and, given that $\mathbf{P_Z}\mathbf{Z}\widehat{\boldsymbol{\gamma}} = \mathbf{0}$ and $\mathbf{P_Z}\mathbf{e} = \mathbf{0}$, the right-hand side is simply:

$$\widehat{\delta_0} = \mathbf{P_Z}\mathbf{D}\widehat{\boldsymbol{\alpha}} + \mathbf{P_Z}\mathbf{L}\widehat{\boldsymbol{\theta}} = \widehat{\delta}_{\boldsymbol{\alpha}} + \widehat{\delta}_{\boldsymbol{\theta}} \quad . \tag{5}$$

This means that the conventional return on education, $\widehat{\delta_0}$, can be decomposed into two terms that reflect the impact of the worker and firm/job-title channel. If, conditional on all $\mathbf{Z}$ covariates, workers were randomly allocated to firm/job-title combinations, then the estimate for $\widehat{\delta}_{\boldsymbol{\theta}}$ would be zero. In this case the distribution of schooling levels within each firm/job-title cell would replicate the distribution of schooling levels in the economy, such that the matching of schooling levels to firm/job-titles of different pay standard would not be a source of returns to education. On the other hand, a positive value for $\widehat{\delta}_{\boldsymbol{\theta}}$ would be a clear indication that better educated workers were sorted to higher-paying firms and/or job titles. From the equation above we see that the estimate of $\widehat{\delta}_{\boldsymbol{\theta}}$ may be interpreted as the log point reduction/increase that occurs in the returns to schooling due to the allocation of workers to firms and job titles.

Is it possible to go further and decompose $\widehat{\delta}_{\boldsymbol{\theta}}$ on the contribution due to firms, job, and the matching effects? To do this we would need to separate $\mathbf{L}\widehat{\boldsymbol{\theta}}$ into three separate components, say:

$$\mathbf{L}\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\Phi}} + \widehat{\boldsymbol{\Lambda}} + \widehat{\boldsymbol{\zeta}} \tag{6}$$

where $\widehat{\boldsymbol{\Phi}}$ would reflect the contribution of firms, $\widehat{\boldsymbol{\Lambda}}$ that of jobs, and $\widehat{\boldsymbol{\zeta}}$ the firm/job-title matching effects. Now, if we multiply the above expression by

11. For ease of presentation we assume that schooling is the only time-invariant variable.

$\mathbf{P_Z}$ we obtain

$$\widehat{\delta}_{\boldsymbol{\theta}} = \widehat{\delta}_{\boldsymbol{\Phi}} + \widehat{\delta}_{\boldsymbol{\Lambda}} + \widehat{\delta}_{\boldsymbol{\zeta}} \quad . \tag{7}$$

Unfortunately there is no unique way to implement the decomposition in equation (6). However, we can follow Woodcock (2015) and assume that the matching effects are orthogonal to the firm and job title effects. In practical terms this amounts to running a linear regression of the fitted values $\mathbf{L}\widehat{\boldsymbol{\theta}}$ on a fixed effect for firm and another for job title. The estimates of these fixed effects give us the separate contribution of firms and job titles while the residual can only be attributed to matching effects. With this approach we are ascribing as much as possible of the variation on $\mathbf{L}\widehat{\boldsymbol{\theta}}$ to the additive effects of firms and jobs. Thus, the estimate we obtain for the firm/job-title matching effect ($\widehat{\delta}_{\boldsymbol{\zeta}}$) should be seen as a lower bound.

5.3. Benchmark regression

We start by estimating a conventional human capital wage function including as covariates a quadratic on age of the worker, a quadratic on tenure, a measure of firm size (log of number of employees), gender, and worker schooling. Table 1 reports the results of the benchmark specification in Column (1).

As expected, wages increase with age and tenure at a decreasing rate, reaching the maximum at 67 and 45 years, respectively. Also, not surprisingly, larger firms pay higher wages. Conditional on the workers' age, tenure, schooling, and firm size, the gender wage gap in Portugal over this period was around 27 log points.

According to our estimates in Column (1), in Portugal each additional year of education yields, on average, an 8.2% labor market return (7.9 log points). This return is in line with international evidence, even though it places Portugal among the countries with relatively high returns to schooling (see Harmon, Oosterbeek, and Walker, 2003; Card, 1999; the cross-country survey of estimates by Ashenfelter, Harmon, and Oosterbeek, 1999; Trostel, Walker, and Woolley, 2002; and Montenegro and Patrinos, 2014).

This figure, 7.9 log points, is our key number of interest. Despite the limitations that prevent its interpretation as a causal effect of education on wages (e.g. unobserved ability is correlated with schooling (ability bias)), it is a standard approach that is based on a formal model of investment in human capital. For this reason, it has been estimated on thousands of data sets for many countries and time periods, which clearly makes it one of the most widely used models in empirical economics. Therefore, we aim to analyze, decompose, and understand in more detail what lies behind the estimated return to education.

	(1)	(2)
Age	0.0403	
	(0.0006)	
Age squared	-0.0003	-0.0002
	(0.0000)	(0.0000)
Tenure	0.0181	0.0058
	(0.0005)	(0.0000)
Tenure squared	-0.0002	-0.0001
	(0.0000)	(0.0000)
Firm size (log)	0.0605	0.0263
	(0.0047)	(0.0000)
Gender (Female=1)	-0.2721	
	(0.0042)	
Schooling	0.0791	
	(0.0009)	
Time effects	✓	✓
Worker effects		✓
Firm/Job-title effects		✓
N	19,152,256	19,152,256
R Squared	0.5528	0.9572

Table 1. Conventional Wage Equation

Notes: The dependent variable is the logarithm of real hourly wages. Column (1) reports the OLS results of the benchmark specification including as covariates age, age squared, tenure, tenure squared, size of the firm, gender, and worker schooling. Column (2) shows the full specification, including worker and firm/job-title fixed effects. In the full specification age, gender, and schooling are absorbed by the worker fixed effects. Standard errors are clustered at the firm level in specification (1) and at the firm/job-title level in specification (2).

5.4. Regression accounting for firm/job-title and worker unobserved heterogeneity

We now extend our model by adding worker and firm/job-title fixed effects, as presented in Section 5.1. Column (2) in Table 1 presents the results. Gender and schooling are absorbed by worker fixed effects (time invariant worker heterogeneity). The linear term of the age polynomial is also absorbed in the worker fixed effects due to the well known age/year/cohort identification problem. The observed time-varying characteristics of the worker (age squared, tenure, and tenure squared) and firm (firm size) are affected by the allocation of the workers into firms and job titles, resulting in a smaller impact on wages when compared with the benchmark specification.

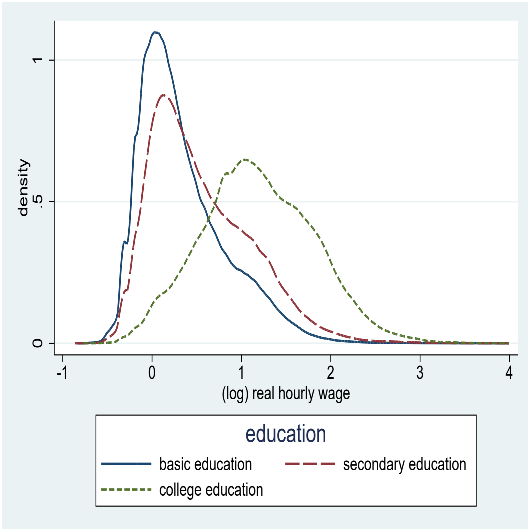

FIGURE 1: Distribution of (log) wages, separately by education level

Notes: Reports kernel densities of log hourly wages in the economy separately for three educational levels: basic education, secondary education, and college education.

Figure 1 shows the empirical distributions of the log hourly wages in the economy separately for three educational levels: basic education, secondary education, and college education. As expected, raw wages for the lowest education level are displaced to the left, and are less dispersed than for any other educational group. This relatively low dispersion of wages could reflect the operation of collective bargaining, setting binding wage floors for low-skilled workers, and in particular, the role of mandatory minimum wages. College education, instead, yields the most heterogeneous returns in the economy.

Figure 2 provides an overview of the different sets of fixed effects, separately for the three educational levels. The regression model includes two high-dimensional fixed effects: the firm/job-title and the worker fixed effects. However, the former comprises a firm fixed effect, a job title fixed effect, and the interaction between them (see Section 5.2).

The firm fixed effect reveals the heterogeneity of wage policies across firms. A high firm fixed effect (high-wage firm) is a firm with total compensation higher than expected once we control for the permanent heterogeneity of workers and job titles, and observable time-varying worker and firm attributes. In Figure 2, the panel displaying the distribution of the firm fixed effects reveals the existence of a wide range of pay standards across firms and the presence of mass points that correspond to large firms in the economy. It is clear that

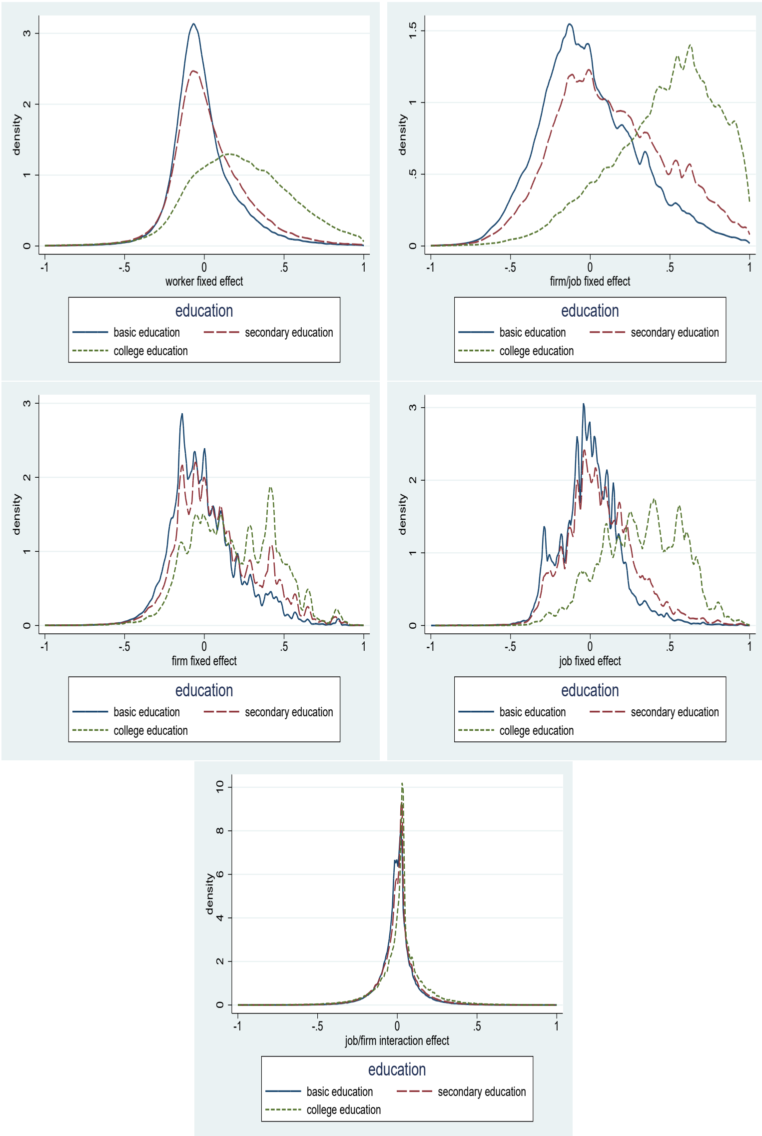

FIGURE 2: Distribution of firm, job title and worker fixed effects, separately by education level

Notes: At the top, the figure exhibits the kernel densities for the worker and the firm/job-title fixed effects separately for the three educational levels. The bottom of the figure depicts the kernel densities for firm, job title and firm/job-title interaction fixed effects. These figures follow from the estimation reported in Table 1 Column (2).

more educated workers are systematically overrepresented in firms with more generous wage policies. In other words, the better educated workers have better access to higher paying firms.

The heterogeneity of job title fixed effects is likely to be generated by variations across occupations and skills and by differences across collective wage agreements. A high job title fixed effect is a job title with total compensation higher that expected after controlling for the observed and unobserved characteristics of workers and firms. Figure 2 shows that the allocation of workers to job-titles is also clearly influenced by the levels of education. Therefore, education may be seen as a passport to higher paying job titles.

Overall, Figure 2 suggests that more educated workers tend to overpopulate matches characterized by high paying firms and high paying job titles.[12]

As opposed to the distribution of firm/job-title fixed effects, worker effects reveal very smooth distributions, presumably reflecting the existence of a continuum of worker abilities in the economy. The dispersion of worker abilities is considerably greater among college graduates than among the other schooling levels. The worker fixed effects represent the permanent worker heterogeneity, both observed (such as gender and schooling) and unobserved. A high worker fixed effect (high-wage worker) is an individual with total compensation higher than expected after controlling for observable time-varying worker and firm attributes, and for firm and job title permanent heterogeneity.

Given the evidence that education grants access to better paying firms and job titles, we next quantify precisely the relevance of the two channels determining the returns to education.

5.5. Decomposing returns to education

Starting from a traditional Mincer-type wage regression, we now distinguish between different sources of the returns to education. The first is the employer channel, which operates to the extent that education provides a "passport" to firms with more generous pay standards. In other words, if workers endowed with better schooling levels are matched to better-paying firms, that will result in an education wage premium that we capture as the "firm channel". This mechanism operates as long as workers with different schooling levels are not randomly allocated to firms of different pay standards. It thus reflects the existence of sorting of educational levels across firms.

A strictly parallel reasoning would apply to job titles. If workers endowed with better schooling levels are matched to better-paying job titles, that will result in an education wage premium that we capture as the "job title channel".

12. As discussed above, the graphical representation of this interaction term is not unambiguous, permitting a number of distinct parameterizations. For the illustration purposes only, here we assumed that the interaction term is orthogonal to the firm and job title fixed effect (Woodcock, 2015), as discussed in Section 5.2.

It thus reflects the existence of sorting of educational levels across job titles. One would expect, of course, that the level of education plays a key instrumental role facilitating access to different occupations or, more specifically, to distinct job titles.[13]

In some sense a firm can be seen as a collection of job titles. Different technologies and/or distinct human resources management strategies may result in combinations of high paying firms with high paying job titles. For example, it can be argued that technology sophisticated firms often organize highly complex tasks. The empirical relevance of this "sophistication technology channel" should manifest itself via the association of the levels of education with the sign and magnitude of the assortative match between high paying firms and high paying job titles.

In our setting, the remaining channel, after accounting for firm and job-title heterogeneity in pay standards, would be the individual component of the returns to education. Such component encompasses both a "pure" return on the worker's education and a return on other individual time-invariant attributes, whether observed or unobserved.

Table 2 (panel A) reports the results from the Gelbach decomposition discussed in Section 5.2. Column (1) shows the coefficient of the benchmark results on returns to education. Column (2) reports the coefficient of the full specification that includes worker and firm/job-title fixed effects, which of course is zero, because the regression coefficients of the time-invariant variables are absorbed into the fixed effects. The results of the decomposition are reported in Columns (3) and (4), which, by construction, sum up to the coefficient of the benchmark specification.

Having estimated two sets of (high-dimensional) fixed effects —one corresponding to the firm/job-title cell, and the other to the worker —we find, first of all, that only 3.0 log points out of the 7.9 overall return on education are immune to the allocation of individuals into firms and job titles. In other words, this decomposition shows that the economy's return to education would fall by 4.9 log points if workers of different schooling levels were randomly distributed across firms and job titles jointly.

With firm/job-title fixed effects estimates at hand, we now take advantage of the insights given by Figueiredo, Guimarães, and Woodward (2014), and Woodcock (2015) to disentangle the role of the firm fixed effect, job title fixed effect, and interaction between the two (see equation (6)). This decomposition shows that the economy's return on education would fall by 2.0 log points if workers were randomly distributed across firms (see Table 2 - panel B). Therefore, a remarkable one fourth of the returns to education operates via the allocation of workers to firms —"firm channel", thus reflecting the existence

13. This relates to an old debate discussing whether or not one should control for occupation in a Mincerian regression when estimating the returns to education.

Panel A - Gelbach Decomposition of the Return to Education

Benchmark Regression	Full Specification	Decomposition into: Worker FE	Firm/Job-title FE
(1)	(2)	(3)	(4)
0.0791			
	0.0000	0.0303	0.0488

Panel B - Decomposition of the Firm/Job-title FE

Firm/Job-title FE	Firm	Job Title	Interaction
(1)	(2)	(3)	(4)
0.0488	0.0205	0.0275	0.0007

Table 2. Conditional Decomposition of the Return to Education

Notes: Panel A: The conditional decomposition of the return to education is based on Gelbach (2016). Column (1) reports the coefficient of the benchmark result on return to education. Column (2) reports the coefficient of the full specification after including worker and firm/job-title fixed effects, which is zero by construction. The results of the decomposition are reported in Columns (3) and (4). Adding up the results of Columns (3) and (4) we obtain the benchmark coefficient in Column (1). Panel B: The conditional decomposition of the contribution of the firm/job-title FE to the return to education on a firm specific effect, job title specific effect, and the interaction effect between the two. Column (1) shows the coefficient of the firm/job-title FE contribution. The results of the decomposition are reported in Columns (2) to (4). Adding up the results of Columns (2) to (4) we obtain the result in Column (1).

of sorting of educational levels across firms (after controlling for job title heterogeneity).

The role of the firm's pay standards shaping wage differentials across education groups can be compared to its role shaping the gender pay gap. Cardoso, Guimarães, and Portugal (2016) and Card, Cardoso, and Kline (2016) report a firm contribution to the gender pay gap of around 20% or one fifth of the overall gap. We uncover that the role of the firm shaping the returns to education is more important than its role shaping the gender pay gap. To our knowledge, this is a novel fact that had until now attracted no discussion in the literature (except the comment by Card, Heining and Kline (2013) when dealing with Germany). Having come such a long way in recent decades, the literature on the returns to schooling had, nevertheless, not yet analyzed differences in firm wage effects across schooling levels.[14]

14. Engbom and Moser (2017) compare the returns to a bachelor, master, and PhD degree in a wage regression with and without firm fixed effects, showing that the returns to a bachelor or master degree fall by about one fourth as they add the firm fixed effects, whereas

Table 2 (panel B) reports that the returns to education would fall by 2.8 log points if workers were randomly distributed across job titles. Therefore, around one third of the returns to education operates via the allocation of workers to job titles —"job title channel". It is not surprising that more educated workers are allocated to more skill demanding tasks and thus into better paid job titles. Moreover, some occupations may have minimum education requirements. This counterfactual exercise of random allocation of workers with different levels of education to distinct job titles makes this effort somewhat artificial in nature.

There is some indication that high paying firms with high paying job titles tend to employ more highly educated workers. This effect is, however, rather modest, contributing a mere 0.1 log points to the overall return on education.

6. Accounting for Coworkers' Education and Human Capital Spillovers

6.1. Introducing the role of coworker education

To capture educational spillovers we add as an additional regressor the average education of the coworkers of worker i. The coworkers of worker i are defined as all workers that, in a given year, share the same establishment and job title with worker i.

Table 3 Column (1) reports the results of this extended regression. This specification suggests that the return to own education is reduced in a non-negligible way to 4.1 log points for an extra year of own education. More striking, an additional year of the coworkers' schooling with the same job title in a firm raises wages by 5.7 log points. This outcome should be interpreted with great caution, as it indicates that one additional year of coworkers' schooling would be more influential driving workers' wages than one additional year of their own education.[15] For its part, the gender wage gap is reduced to 14.2 log points, largely because the gender segregation impact on wages is estimated to be 21.0 log points.

A number of identification problems and specification pitfalls have been raised in the literature, in particular by Manski (1993) and more recently by Angrist (2014). Indeed, even in the absence of social interactions, individuals in the same firm and job title category will tend to have similar wages, which in general will lead to an upward bias in the estimation of the coworker education

---

the returns to a PhD hardly change. However, their analysis is restricted to holders of a Higher Education Degree and does not account for worker unobserved heterogeneity and, in any case, their results would vary depending on the order of introduction of the additional controls.

15. This result has some parallel with the studies on social returns to education at the firm level (Battu, Belfield, and Sloane, 2003; Wirz, 2008; and Martins and Jin, 2010).

	(1)	(2)	(3)	(4)
Age	0.0366	0.0157	0.0366	0.0150
	(0.0005)	(0.0001)	(0.0001)	(0.0000)
Age squared	-0.0003	-0.0002	-0.0003	-0.0002
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0198	0.0064	0.0198	0.0065
	(0.0005)	(0.0000)	(0.0001)	(0.0000)
Tenure squared	-0.0003	-0.0001	-0.0003	-0.0001
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Firm size (log)	0.0540	0.0299	0.0539	-
	(0.0041)	(0.0002)	(0.0003)	-
Gender (Female=1)	-0.1417	-0.0621	-0.1416	-0.0318
	(0.0017)	(0.0006)	(0.0003)	(0.0001)
Schooling	0.0411	0.0054	0.0411	0.0029
	(0.0004)	(0.0001)	(0.0001)	(0.0000)
Coworker schooling	0.0573	-	0.0574	-
	(0.0007)	-	(0.0001)	-
Coworker gender	-0.2095	-	-0.2097	-
	(0.0051)	-	(0.0006)	-
HC spillovers ($\overline{\alpha}_{-it}$)	-	0.5007	-	0.2050
	-	(0.0006)	-	(0.0006)
Time effects (μ_t)	✓	✓	✓	
Worker effects (α_i)		✓		✓
Firm/Job-title effects ($\theta_{\mathbf{F}\times\mathbf{J}(i,t)}$)		✓		
Establishment/Job-title/Year effects				✓
N	19,152,256	19,152,256	19,070,170	19,070,170
R Squared	0.6172	0.9584	0.6172	0.9766

TABLE 3. Wage Equation Accounting for Coworker Education and Human Capital Spillovers

Notes: Column (1) reports the results of the wage benchmark regression including as covariates age, age squared, tenure, tenure squared, size of the firm, gender, worker schooling, coworker schooling, and coworker gender. Column (2) reports a full specification that includes worker and firm/job-title fixed effects. Column (3) shows a benchmark regression identical to Column (1) but restricted to a smaller largest connected set, implied by the use of peer group fixed effect in the full specification. Column (4) reports an alternative full specification, including worker and peer group fixed effects. In both full specifications, age, gender, and schooling are absorbed by the worker fixed effects. In Columns (1) and (3), standard errors are clustered at firm level. Standard errors in Columns (2) and (4) are obtained as explained in Appendix A.2.

effect. Even without causal "peer" effects there are mechanical and statistical issues that may lead to similar outcomes between peers. We can distinguish three main problems in the estimation of these effects: reflection (or homophily), selection, and "mechanical" measurement error. The reflection problem states that it is very hard to disentangle whether or not the average behavior in one group is actually influencing that same behavior at the individual level of the members of that group. The selection problem arises if the group is formed endogenously, making it hard to distinguish peer effects from selection effects. The "mechanical" measurement error problem, discussed by Angrist (2014), states that even in settings where peers are assigned randomly there is a mechanical relationship between own and peer attributes that may bias the estimation of the peer effect.[16]

We are confident that our methodological approach can address the three above mentioned estimation hurdles. First of all, we explore a very rich and exhaustive longitudinal database that allows us to overcome the issue of homophily via the presence of individual fixed effects. Second, by controlling for highly disaggregated firm/job-title combinations, we circumvent the issues raised by sorting and peer group formation. Third, measurement error problems are attenuated in our administrative dataset because both wages and hours of work are obtained with unusual accuracy.

6.2. The identification problem

While the introduction of observed coworker education as a regressor presents no particular challenge, a more complicated problem arises if instead we believe that spillovers are linked to coworker characteristics. In the linear model, unobserved time-invariant ability is fully captured by the worker fixed effect and, as noted earlier, so are the other time-invariant observed characteristics of workers, such as education. Thus, to account for spillovers we follow the approach of Arcidiacono et al. (2012). To make matters simple, assume that there are only two time invariant factors, schooling (s_i) and ability (a_i). Further, we can assume that these factors generate spillovers. This means that we can rewrite equation (1) as

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta s_i + a_i + \eta_1 \overline{s}_{-it} + \eta_0 \overline{a}_{-it} + \theta_{\mathbf{F}\times\mathbf{J}(i,t)} + \varepsilon_{it} \quad , \tag{8}$$

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,

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + s_i(\delta - \omega) + (\eta_1 - \eta_0\omega)\overline{s}_{-it} + \alpha_i + \eta_0 \overline{\alpha}_{-it} + \theta_{\mathbf{F}\times\mathbf{J}(i,t)} + \varepsilon_{it} \quad . \tag{9}$$

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. The above equation remains overparameterized and some

16. Feld and Zoelitz (2017) build on Angrist (2014) and study the role of measurement error in the estimation of peer effects.

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. The model can still be estimated if only one of those conditions is imposed. In other words, the model can be estimated using either own schooling or coworker average schooling as a regressor. If s_i is included as a regressor then we have to set the coefficient on $\overline{s}_{-it}$ to zero and the coefficient on s_i becomes $\delta^* = \delta - \eta_1/\eta_0$. On the other hand, if we add $\overline{s}_{-it}$ as a regressor, we have to set the coefficient on s_i to zero, meaning that the coefficient on $\overline{s}_{-it}$ is now $\eta_1^* = \eta_1 - \eta_0\delta$. In the analysis that follows we report results for the following specification:

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta^* s_i + \alpha_i + \eta_0\overline{\alpha}_{-it} + \theta_{\mathbf{F}\times\mathbf{J}(i,t)} + \varepsilon_{it} \quad . \tag{10}$$

It may seem strange that a time invariant characteristic such as schooling is not absorbed by the worker fixed effect. But remember that this is a nonlinear model on the α_i. Under a set of assumptions regarding the error term, which are clearly identified in Arcidiacono et al. (2012), the least squares solution provides consistent estimates for the parameters of the model.[17] Arcidiacono et al. (2012) also provide an estimation algorithm for equation (10) that is similar in spirit to that proposed in Guimarães and Portugal (2010) for the solution of models with multiple fixed effects. In Appendix A.1 we discuss the algorithm for estimation of this model and propose some modifications that make it faster. Additionally, we also show in Appendix A.2 how to compute the standard errors for the nonlinear regression.

6.3. Extending Gelbach's decomposition

To show that Gelbach's decomposition can still be implemented in this setting we rewrite (10) in matrix form:

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma} + \delta^*\mathbf{S} + \mathbf{D}\boldsymbol{\alpha} + \eta_0\mathbf{W}\mathbf{D}\boldsymbol{\alpha} + \mathbf{L}\boldsymbol{\theta} + \varepsilon \ .^{18} \tag{11}$$

With this parameterization we know that $\mathbf{D}\boldsymbol{\alpha} = \eta_1/\eta_0\mathbf{S} + \mathbf{D}\mathbf{a}$ (where $\mathbf{a}$ is a vector of dimension N containing the unobserved value of ability for each worker) and thus the estimates for the fixed effects contain part of the effect of schooling. Since we cannot disentangle the effect of ability and education, we lump together these components into a single vector $\mathbf{D}\boldsymbol{\alpha}^*$ (i.e. $\mathbf{D}\boldsymbol{\alpha}^* \equiv$

17. However, as is common with panel data, the estimates for the fixed effects remain inconsistent if the time dimension is fixed.

18. Note that $\mathbf{W}\mathbf{D}\boldsymbol{\alpha}$ is a variable containing the average of the fixed effects of the coworkers. In Appendix A.1 we explain how we construct $\mathbf{W}$.

$\delta^*\mathbf{S} + \mathbf{D}\boldsymbol{\alpha}$).[19] Writing now equation (10) in terms of its fitted value

$$\mathbf{Y} = \mathbf{Z}\widehat{\gamma} + \mathbf{D}\widehat{\boldsymbol{\alpha}^*} + \widehat{\eta_0}\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}} + \mathbf{L}\widehat{\boldsymbol{\theta}} + \mathbf{e} \tag{12}$$

we can apply Gelbach's decomposition by left-multiplying both sides of the above expression by $\mathbf{P_Z}$. We know from the first-order conditions for the non-linear least squares problem (see Appendix A.1) that the residuals are orthogonal to $\mathbf{Z}$. Thus, when we multiply both sides of the above expression by $\mathbf{P_Z}$ we obtain

$$\widehat{\delta_0} = \mathbf{P_Z}\mathbf{D}\widehat{\boldsymbol{\alpha}^*} + \widehat{\eta_0}\mathbf{P_Z}\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}} + \mathbf{P_Z}\mathbf{L}\widehat{\boldsymbol{\theta}} = \widehat{\delta}_{\boldsymbol{\alpha}^*} + \widehat{\delta}_{\mathbf{W}} + \widehat{\delta}_{\boldsymbol{\theta}} \quad . \tag{13}$$

The conventional return on schooling, $\widehat{\delta_0}$, is now decomposed in three terms, reflecting the contribution of the different channels —the worker ($\widehat{\delta}_{\boldsymbol{\alpha}^*}$), the coworkers ($\widehat{\delta}_{\mathbf{W}}$), and the firm/job-title ($\widehat{\delta}_{\boldsymbol{\theta}}$). With the caveats already noted, we could use the approach discussed above to further break the firm/job-title contribution into a firm, job title, and firm/job-title matching effect. The procedure would be the same if instead we wished to obtain a decomposition of the baseline coefficient associated with coworker schooling. In that case, the $\mathbf{P_Z}$ matrix would need to be defined accordingly but implementation would be straightforward.

6.4. Empirical results on the returns to education and spillover effects

We now expand our exercise by specifying a model that includes a measure of human capital spillovers in addition to the worker and the firm/job-title fixed effects. As discussed above, we rely on an iterative estimation procedure to quantify the impact of coworkers' average individual fixed effect. Column (2) in Table 3 reports the results.

There is clear empirical support for the notion that peer quality has a strong impact on individual wages. The key parameter of interest (η_0) is estimated to be 0.5, meaning that if the quality of the peers as measured by ($\overline{\alpha}_{-it}$) increases by 1% wages will increase by 0.5%. Put differently, a one standard deviation increase in the measure of human capital spillover (0.1139) leads to a wage increase of about 5.7% (0.5007*0.1139).[20] This figure is not at odds with

19. Note that $\mathbf{D}\boldsymbol{\alpha}^* \equiv \delta^*\mathbf{S} + \mathbf{D}\boldsymbol{\alpha} = \delta\mathbf{S} + \mathbf{D}\mathbf{a}$. Had we adopted the alternative parameterization that included the average education of coworkers as a regressor, we would obtain the same results. In that case $\delta^*\mathbf{S}$ would be replaced by $\eta_1^*\mathbf{W}\mathbf{S}$ in equation (11). But note that $\mathbf{D}\boldsymbol{\alpha}$ would already equal $\delta\mathbf{S} + \mathbf{D}\mathbf{a}$ and to obtain a term equivalent to $\eta_0\mathbf{W}\mathbf{D}\boldsymbol{\alpha}$ in equation (11) we would need to combine $\eta_1^*\mathbf{W}\mathbf{S}$ and $\eta_0\mathbf{W}\mathbf{D}\boldsymbol{\alpha}$.

20. The standard deviation estimate (0.1139) corresponds to the average of the standard deviations of the measure of peer quality (as measured by the fixed effects of each peer). This and other statistics from the wage distribution corresponding to this specification are given in Table A.3.2 in the Appendix A.3.

those provided by Cornelissen et al. (2017) using data for Munich (3.6%), and that presented by Battisti (2013) using data from the Italian region of Veneto (3.9%).[21]

In this specification, the identification of human capital spillovers arises from changes in the coworker composition over time after isolating the endogenous sorting of workers into firm/job-title cells. In other words, the impact of peer quality is driven by the entry and exit of workers into particular job titles within the firm. At first glance, this seems to be a reasonable identification strategy. However, the occurrence of firm specific shocks may compromise this estimation strategy to the extent that they can influence both the level of the wages and the coworker composition. For example, a negative product demand shock may lead to lower wages and to a disproportionate decrease of low quality workers, engendering a spurious correlation between wages and peer quality. Nevertheless, if the firm specific shock affects wages solely through a change in the peer quality, one should resist the temptation to exclude this shock (e.g., via the inclusion of firm year specific effects). A similar argument can be advanced for the case of job title specific shocks. In this case, the presence of job title specific wage trends may confound the peer effect estimation. There is no obvious optimal level of disaggregation in the use of high-dimensional fixed effects. In the limit, if we were to use an establishment/job-title/year fixed effect then it would overlap with the definition of peer groups. As pointed out by Cornelissen et al (2017) in that case identification of η_0 would come strictly from changes on the size of the peer groups, eliminating any endogenous contamination from sorting into establishments and job titles, over time. [22]

In the specifications in Table 3, as discussed above, schooling is not fully absorbed by the worker fixed effects, because, by construction, time invariant covariates are no longer orthogonal to the worker fixed effects. In the current setting, the returns to own schooling cannot be directly extracted. Utmost we can provide an estimate of the returns to own schooling for any given return to coworker schooling. Thus, those two coefficients cannot be identified separately, because only a linear combination of the two can be estimated (see Section 6.2).

The idea that the returns to own and to peers' education is jointly identified can be thought of intuitively. The knowledge transmitted by the educational system can be acquired either directly at the origin —the school desks —or indirectly, as it trickles down from educated colleagues at the workplace. Therefore, the valuation in the labor market of the skills acquired through these two sources is most likely related. As an example, it is hard to conceive

21. In these comparisons we are using the closest sampling plan and econometric specification.

22. The value of α_{-it} in peer group j can be expressed as $(\alpha_{\bullet j} - \alpha_i)/(n_{jt} - 1)$ where $\alpha_{\bullet j}$ is the sum of the fixed effects for all workers in group j. Since α_i is absorbed by the worker fixed effect and $\alpha_{\bullet j}$ by the establishment/job-title/year fixed effect the only source of variation left to identify η_0 comes from changes in n_{jt}.

of a country where own education would provide few skills and thus yield a low return in the labor market, whereas, on the contrary, the knowledge acquired through educated peers, was highly valued; the symmetrical reasoning would naturally apply.

Therefore, we can progress by setting different scenarios. Assuming that external returns to education are null (see Column (2), under an assumption in line with Acemoglu, 2000) would imply a meager 0.5 log points return of an additional year of own education.[23] However, if we arbitrarily assume a 1 log point external return to education, our model would imply a 2.5 log points return to own education. Assuming, for example, a 6 log points return to own education, our model would imply a return to coworker education of 2.7 log points. Under similar reasoning, the gender wage gap would be around 6 log points in the absence of gender segregation.

The observed time-varying characteristics of the worker in the model (age and tenure) and firm (log size) are affected by the allocation of the workers into the different firms/job title combinations, presenting a smaller impact on wages when compared to the benchmark specification.

In our final specification, we now include establishment/job-title/year fixed effects, that is a peer group fixed effects (Column (4) in Table 3). The number of observations is slightly reduced due to a smaller largest connected set. Column (3) reports the benchmark specification on this slightly smaller sample. Proceeding in this way, we are adding the role of time varying changes in the wage policies of the firms (and within firms across establishments), the influence of the secular trends in the remuneration of job titles, and the interplay between establishment, job title, and year effects.

The presence of these additional fixed effects visibly reduced the impact of human capital spillovers on individual wages. Nevertheless, the role of peer quality is still sizeable. The peer regression coefficient is now estimated to be 0.21. Now, a one standard deviation increase in the measure of peer quality (0.10) leads to a wage increase of 2.1 log points. In our minds, this estimate is better interpreted as a lower bound for the impact of human capital spillovers. [24]

Under the assumption of no external returns, the implied own return to education is 0.3 log points. Correspondingly, assuming 6 log points own return

23. As mentioned in Section 6.2, if s_i is included as a regressor, then we have to set the coefficient on $\overline{s}_{-it}$ to zero and the coefficient on s_i becomes $\delta^* = \delta - \eta_1/\eta_0$.

24. We also examine the sensitivity of our results along two dimensions: Firstly, we exclude covariates from the full specification (à la Arcidiacono); Secondly, in addition to the covariates, we exclude also the peer groups whose size is larger than 10 individuals. Even after excluding covariates and reducing the peer group size, the main outcome is that the estimate of the human capital spillover is still quite sizeable. See Table A.3.3 in the Appendix A.3.

to education, our model would imply a 1.2 log points return to coworker education.

Panel A - Variance Decomposition	
α_i - worker	0.3052
$\eta_0\overline{\alpha}_{-it}$ - coworker	0.0503
$\theta_{\mathbf{P}(i,t)}$ - peer group fixed effect	0.5728
$Z_{it}\widehat{\gamma}$	0.0483
Panel B - Correlations	
$\rho(\alpha_i, \overline{\alpha}_{-it})$	0.7601
$\rho(\alpha_i, \theta_{\mathbf{P}(i,t)})$	0.1793
$\rho(\overline{\alpha}_{-it}, \theta_{\mathbf{P}(i,t)})$	0.1987
Panel C - Fixed Effect Heterogeneity	
σ_{α_i}	0.2497
$\sigma_{\overline{\alpha}_{-it}}$	0.2253
$\overline{\sigma}_{\alpha_{i\mathbf{P}(i,t)}}$	0.1031
$\sigma_{\theta_{\mathbf{P}(i,t)}}$	0.3802

Table 4. Statistical Moments from Wage Distribution

Note: The statistics are computed from the estimates given in Column (4) from Table 3. Panel A gives the variance decomposition according to the covariances between wages and the components of the wage equation (worker, coworker, peer group(establishment/job-title/year) and time variant covariates). Panel B shows the correlations between the worker, coworker, and peer group fixed effects. Panel C provides the standard deviations of worker, coworker, peer group fixed effects, and the average of the standard deviations of the measure of peer quality (as measured by the fixed effect of the peers).

Panel A in Table 4 shows the decomposition of the variance of the wages from our preferred specification, which includes besides the covariates, worker, establishment/job-title/year, plus the average peer quality. The worker time-invariant component accounts for 31% of the variance of individual wages, while the coworker's quality explains a sizeable 5% of the overall wage variation. The contribution of firm's and job title's heterogeneity to the variance of individual wages is 57 %. The contribution of the covariates component, including the time invariant variables accounts for 5%.

We now proceed to a twofold generalization of the decomposition exercise (Gelbach) discussed at length in Section 6.3. First, the complexity of the exercise is exacerbated by the need to decompose an additional regression coefficient. At this stage, the inclusion of this additional covariate (coworker schooling) will require, in the full specification, an adequate normalization. Second, we employ a methodology that enables us to assess the importance of human capital spillovers. The results are reported in Table 5.

For a given distribution of coworker schooling, the effect of one additional year of own schooling is 1.8 log points, after discounting the role of sorting

Gelbach Decomposition of the Return to Education (Own and Coworker)				
	Benchmark Regression	Decomposition into: Worker*	Establishment/Job-title/Year	Human Capital spillovers*
	(1)	(2)	(3)	(4)
Own Schooling	0.0411	0.0181	0.0203	0.0026
Coworker Schooling	0.0574	0.0198	0.0330	0.0045

Table 5. Decomposing Returns to Own and Coworkers' Education

Notes: The conditional decomposition of the return to education is based on Gelbach (2016). Column (1) reports the coefficient of the benchmark result on return to own and coworker education. Column (2) displays the worker contribution to the returns on own schooling (adding the coefficient associated with own education (0.0029) in the full specification with the worker FE contribution (0.0152), and to the returns to coworker education). Column (3) shows the contribution of sorting over establishments and job titles over time. Column (4) displays the contribution of Human Capital Spillovers to the returns to own and coworker education.

among establishment/job-title/year cells and the role of peers' quality. That is to say that if workers were randomly allocated into establishment/job-title/year cells the return to education would be reduced by 2.0 log points.[25]

The remaining component of the returns to education emerges because more educated workers tend to be matched with higher quality coworkers. This indication of positive assortative matching (in the spirit of Lopes de Melo, 2018) suggests that own education and peer quality are complements, generating a human capital spillover in the returns to education (of 0.3 log points). Put differently, if coworkers were allocated through a randomized experiment (holding constant the coworker education distribution) the returns to education would be reduced by 0.3 log points.

In sum, the estimated return to education (4.1 log points) can be decomposed into three parts: the individual return to education (contributing 44%), a sorting component corresponding to 49% (among establishment/job-title/year), and an assortative matching term responsible for 6%.

The naive regression coefficient estimate of the effect of coworker schooling on wages (5.7 log points) can also be decomposed into three different

25. It can be shown in an exercise similar to the one performed in Panel B of Table 3 that if workers were allocated randomly into firms the returns to education would fall by 0.9 log points, whereas if randomly allocated to job titles the return would fall by 1.1 log points. The evidence that more educated workers tend to be more represented in combinations of high paying firms and high paying job titles is rather muted (corresponding to a reduction in returns to schooling by 0.1 log points). This evidence confirms the decisive role of sorting driving the returns to education.

channels. The first component (2.0 log points) arises from the correlation between coworkers' education and the worker individual fixed effect. This component is engendered by homophily (or reflection in Manski's typology) or the resemblance between the worker and his coworker counterparts. More specifically, the return to coworker education is reduced by 2.0 log points when worker fixed effects are included in the regression (in the spirit of Arcidiacono et al., 2012). The second component (3.3 log points) arises from the allocation of more educated coworkers into higher paying establishment/job-title/year cells.[26] Finally, the impact of one additional year of coworkers' education is estimated to be 0.5 log points. This is the return to coworkers' education that would remain after dismissing the bias arising from homophily and the selection of more educated coworkers into better paying firm/job-title combinations.

Our results compare to those of Nix (2016), who also accounts for worker and employer fixed effects and several controls to tackle worker sorting and firm heterogeneity. She finds a 0.3% increase in a worker's wage as the share of college educated colleagues increases by 10 percentage points.[27]

A key point to retain from our analysis is therefore the quantification of the impact of coworkers' schooling on wages, net of reflection and sorting effects. Indeed, a naive model specification reported in the literature, which simply adds coworker education to a traditional wage regression, has led to implausibly large estimates. We show that sorting of education levels across firms and job titles can account for as much as 58% of that apparent return on coworker education; reflection can additionally account for 35%. We identify the remaining 0.5 log points as the impact of coworker schooling on a worker's wage.

7. Conclusion

We explore the sources of the returns to education, unveiling the impact of the firm, the job, and the coworkers channels. We thereby contribute to the intersection of three strands of the literature: the role of the firm shaping the wage distribution, the returns to education, and the spillovers of education. We combine longitudinal linked employer-employee data of remarkable quality with innovative empirical methods to address common problems in the estimation of the returns to peer attributes, namely: the reflection or homophily problem, selection issues, and common measurement errors and confounding factors.

Schooling grants access to better paying firms and jobs. The first part of our analysis concentrates on the returns to own education only. It reveals that

26. The sorting into firms accounts for 1.5 log points (firm channel), the allocation into job titles is responsible for 1.8 log points (job title channel), and the remaining 0.0 log points arises from clustering better educated coworkers into higher paying establishment/job-title/year cells.

27. The overall share of college-educated colleagues in her sample of Swedish males is 31%.

one fourth of the overall return on a year of education (7.9 log points) operates through the firm channel, whereas a third operates through the detailed job the worker performs. The worker component is responsible for 38% of the return to education.

In the second part of the analysis we show that peer quality has a sizeable impact driving wages. In our preferred specification, a 10% increase in the measure of peer quality leads to a wage increase of 2.1 log points.

In this setup, an additional year of average education of coworkers yields a 0.5 log points increase in a worker's wage, after we net out a 2.0 log point return due to homophily (similarity of own and peers' characteristics), and 3.3 log points due to worker sorting across firms and jobs over time. As such, in a naive specification of the wage regression that includes own and peers' average education, without tackling the reflection and sorting problems, all those effects would combine into a misleading overall return on peers' education of 5.7 log points.

Overall, our results show a discernible effect of coworkers' education on a worker's wage, consistent with the operation of spillover effects within the firm. They also stress the importance of access to firms, jobs, and coworkers, shaping the wage distribution along a dimension —returns to education —not previously explored in a comprehensive way in the literature.

References

. Abowd, John M., Francis Kramarz, and David Margolis (1999). High wage workers and high wage firms. *Econometrica*, 67(2): 251-333.

. Abowd, John M., Robert H. Creecy, and Francis Kramarz (2002). Computing person and firm effects using linked longitudinal employer-employee data. Mimeo Cornell University.

. Acemoglu, Daron (1996). A microfoundation for social increasing returns in human capital accumulation. *Quarterly Journal of Economics*, 111(3): 779-804.

. Acemoglu, Daron (2014). Labor market externalities. Labor Economics 14.661 Lectures 1 and 2 Slides, MIT.

. Acemoglu, Daron and Joshua Angrist (2000). How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, 15: 9-59.

. Angrist, Joshua (2014). The perils of peer effects. *Labour Economics*, 30: 98-108.

. Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler (2012). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics*, 3: 421-470.

. Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, 6(4): 453-470.

. Bandiera, Oriana, Iwan Barankay, and imran Rasul (2010). Social incentives in the workplace. *Review of Economic Studies*, 77(2): 417-458.

. Battisti, Michele (2013). High wage workers and high wage peers. Ifo Working Paper 168, IFO Institute for Economic Research at the University of Munich.

. Battu Harminder, Clive R. Belfield, and Peter J. Sloane (2003). Human capital spillovers within the workplace: evidence for Great Britain. *Oxford Bulletin of Economics and Statistics*, 65(5): 575-594.

. Belzil, Christian (2007). The returns to schooling in structural dynamic models: A survey. *European Economic Review*, 51(1): 1059-1105.

. Blau, Francine D. (1997). *Equal Pay in the Office*. Lexington Books.

. Blundell, Richard, Lorraine Dearden, and Barbara Sianesi (2005). Evaluating the effect of education on earnings: Models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society Series A*, 168(3): 473-512.

. Card, David (1999). "The causal effect of education on earnings", in O. Ashenfelter and D. Card (ed.), *Handbook of Labor Economics*, Elsevier, vol 3A, chapter 30.

. Card, David (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5): 1127-1160.

. Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1): S13-S70.

. Card, David, Ana Rute Cardoso, and Patrick Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*, 131(2): 633–686.

. Card, David, Jörg Heining, and Patrick Kline (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3): 967-1015.

. Cardoso, Ana Rute, Paulo Guimarães, and Pedro Portugal (2016). What drives the gender wage gap? A look at the role of firm and job-title heterogeneity. *Oxford Economic Papers*, 68(2): 506-524.

. Cardoso, Ana Rute and Pedro Portugal (2005). Contractual wages and the wage cushion under different bargaining settings. *Journal of Labor Economics*, 23(4): 875-902.

. Carneiro, Anabela, Paulo Guimarães, and Pedro Portugal (2012). Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job Title Heterogeneity. *American Economic Journal: Macroeconomics*, 4(2): 133-152.

. Correia, Sergio (2014). Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects. *Statistical Software Components S457874*, Boston College Department of Economics.

. Ciccone, Antonio, and Giovanni Peri (2006). Identifying human capital externalities: Theory with applications. *Review of Economic Studies*, 73(2): 381–412.

. Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg (2017). Peer effects in the workplace. *American Economic Review*, 107(2): 425-456.

. Eeckhout, Jan and Philipp Kircher (2011). Identifying sorting - in theory. *Review of Economic Studies*, 78(3): 872-906.

. Engbom, Niklas and Christian Moser (2017). Returns to education through access to higher-paying firms: Evidence from US matched employer-employee data. *American Economic Review: Papers and Proceedings*, 107(5): 374-378.

. Falk, Armin and Andrea Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1): 39-57.

. Feld, Jan and Ulf Zoelitz (2017). Understanding peer effects: On the nature, estimation and channels of peer effects. *Journal of Labor Economics*, 35(2): 387-428.

. Figueiredo, Octávio, Paulo Guimarães, and Douglas Woodward (2014). Firm–worker matching in industrial clusters. *Journal of Economic Geography, Oxford University Press*, vol. 14(1), pages 1-19, January.

. Gelbach, Jonah B. (2016). When Do covariates matter? And which ones, and how much? *Journal of Labor Economics*, 34(2): 509-543.

. Glaeser, Edward L. (1999). Learning in cities. *Journal of Urban Economics*, 46(2): 254-277.

. Groshen, Erica L. (1991a). Sources of intra-industry wage dispersion: how much do employers matter? *Quarterly Journal of Economics*, 106(3): 869-884.

. Groshen, Erica L. (1991b). The structure of the female/male wage differential: Is it who you are, what you do, or where you work? *Journal of Human Resources*, 26(3): 457-472.

. Gruetter, Max and Rafael Lalive (2009). The importance of firms in wage determination. *Labour Economics*, 19(2): 149-160.

. Guimarães, Paulo and Pedro Portugal (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, 10(4): 628-649.

. Harmon, Colm, Hessel Oosterbeek, and Ian Walker (2003). The returns to education: Microeconomics. *Journal of Economic Surveys*, 17(2): 115-155.

. Herbst, Daniel and Alexandre Mas (2015). Peer effects on worker output in the laboratory generalized to the field. *Science*, 350(6260): 545-549.

. Jovanovic, Boyan and Rafael Rob (1989). The growth and diffusion of knowledge. *Review of Economic Studies*, 56(4): 569-82.

. Kirby, Simon and Rebecca Riley (2008). The external returns to education: UK evidence using repeated cross-sections. *Labour Economics*, 15(4): 619-630.

. Kremer, Michael (1993). The O-ring theory of economic development. *Quarterly Journal of Economics*, 108(3): 551-75.

. Lester, Richard A. (1952). A range theory of wage differentials. *Industrial and Labor Relations Review*, 5: 483-500.

. Lopes de Melo, Rafael (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy*, forthcoming.

. Lucas, Robert Jr. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1): 3-42.

. Machin, Stephen and Alan Manning (2004). A test of competitive labor market theory: The wage structure among care assistants in the South of England. *Industrial and Labor Relations Review*, 57(3): 371-385.

. Manksi, Charles (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3): 531-542.

. Marshall, Alfred (1890). Principles of Economics (First Edition). London: Macmillan.

. Martins, Pedro S. and Jim Y. Jin (2010). Firm-level social returns to education. *Journal of Population Economics*, 23: 539-558.

. Mas, Alexandre and Enrico Moretti (2009). Peers at work. *American Economic Review*, 99(1): 112-145.

. Meng, Xin and Dominique Meurs (2004). The gender earnings gap: Effects of institutions and firms —a comparative study of French and Australian private firms. *Oxford Economic Papers*, 56: 189-208.

. Moen, Espen R. (1999). Education, ranking, and competition for jobs. *Journal of Labor Economics*, 17(4): 694-723.

. Montenegro, C. E., and H. A. Patrinos (2014). Comparable Estimates of Returns to Schooling around the World. World Bank Policy Research Working Paper No.7020.

. Moretti, Enrico (2004a). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data, *Journal of Econometrics*, 121: 175-212.

. Moretti, Enrico (2004b). Human capital externalities in cities, in J. V. Henderson and J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4. Elsevier, pages 2243-2291.

. Moretti, Enrico (2004c). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3): 656-690.

. Moretti, Enrico (2010). Local labor markets, in in O. Ashenfelter and D. Card (ed.), Handbook of Labor Economics, volume 4B. Elsevier, pages 1237-1313.

. Nix, Emily (2016). Learning spillovers in the firm. Mimeo, Yale University.

. Portugal, Ministério do Trabalho e da Segurança Social (1986 to 2013). Quadros de Pessoal. Data in magnetic medium.

. Rauch, James E. (1993). Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics*, 34(3): 380-400.

. Reynolds, Lloyd G. (1951). *The Structure of Labor Markets*. Westport, CT: Greenwood Press, p. 155-229.

. Spence, Michael (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3): 355-374.

. Torres, Sónia, Pedro Portugal, John T. Addison, and Paulo Guimarães (2013). The sources of wage variation: A three-way high-dimensional fixed effects regression model. IZA Discussion Paper No. 7276.

. Trostel, Philip, Ian Walker, and Paul Woolley (2002). Estimates of the economic return to schooling for 28 countries. *Labour Economics*, 9(1): 1-16.

. Weeks, D.L. and D.R. Williams (1964). A note on the determination of connectedness in an N-way cross classsification. *Technometrics*, 6: 319-324.

. Wirz, Aniela Maria (2008). Private returns to education versus education spill-over effects. *Empirical Economics*, 34(2): 315-342.

. Woodcock, Simon (2015). Match Effects. *Research in Economics*, 69(1): 100-121.

Appendix to "The returns to schooling unveiled"

Appendix A.1 - Estimation of Arcidiacono's *et al* regression model

Consider the specification in (10) that includes the worker fixed effect (α_i) and the peer average of the worker fixed effects ($\overline{\alpha}_{-it}$):

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta^* s_i + \alpha_i + \eta_0\overline{\alpha}_{-it} + \theta_{\mathbf{F}\times\mathbf{J}(i,t)} + \varepsilon_{it} \quad . \tag{.1}$$

Estimation of this model is better discussed if we resort to matrix algebra. To simplify notation we let $\mathbf{X}$ be a matrix that contains all but the variables involving the worker fixed effects. These include worker and firm observable characteristics, as well as other control variables such as additional sets of fixed effects. The number of linearly independent columns of $\mathbf{X}$ is given by k and the coefficients associated with the columns of $\mathbf{X}$ are represented by $\boldsymbol{\beta}$. The total number of observations is M while N stands for the total number of workers and P is the number of peer groups. In matrix terms worker fixed effects are given by the product of the worker design matrix $\mathbf{D}$ by the vector $\boldsymbol{\alpha}$ containing coefficients on worker fixed effects. Thus, $\mathbf{X}$ is $(M \times k)$, $\boldsymbol{\beta}$ is $(k \times 1)$, $\mathbf{D}$ is $(M \times N)$ and $\boldsymbol{\alpha}$ is $(N \times 1)$. The variable containing the peer average of the worker fixed effects can be represented by the vector $\mathbf{WD}\boldsymbol{\alpha}$ where $\mathbf{W}$ is an $N \times N$ mean computing matrix. Note that $\mathbf{W}$ is symmetric and block diagonal:

$$\mathbf{W} = diag(\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_P) \quad .$$

Each generic submatrix $\mathbf{w}_j$ identifies a peer group and is given by

$$\mathbf{w_j} = (n_j - 1)^{-1}[\mathbf{ii}' - \mathbf{I}]) \tag{.2}$$

where n_j stands for the number of elements on peer group j and $\mathbf{i}$ is a column vector of 1s with size n_j. Multiplication of $\mathbf{w}_j$ by any vector $[\alpha_1, \alpha_2, ..., \alpha_{n_j}]'$ will result in a vector with the same dimension, $[\overline{\alpha}_{-1}, \overline{\alpha}_{-2}, ..., \overline{\alpha}_{-n_j}]'$, containing the mean of all elements excluding the self. This means that we can write (.1) as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \eta_0\mathbf{WD}\boldsymbol{\alpha} + \varepsilon = \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} + \eta_0\mathbf{W}]\mathbf{D}\boldsymbol{\alpha} + \varepsilon \tag{.3}$$

The equation in (.3) is nonlinear on the $\boldsymbol{\alpha}$. However, this equation can be estimated using nonlinear least squares. To estimate $(\boldsymbol{\beta}, \eta_0, \boldsymbol{\alpha})$ using least-squares define the vector of residuals

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{D}\widehat{\boldsymbol{\alpha}} - \widehat{\eta}_0\mathbf{WD}\widehat{\boldsymbol{\alpha}}$$

and let $S(\widehat{\boldsymbol{\beta}}, \widehat{\eta}_0, \widehat{\boldsymbol{\alpha}}) = \mathbf{e}'\mathbf{e}$. Thus,

$$S(\widehat{\boldsymbol{\beta}}, \widehat{\eta}_0, \widehat{\boldsymbol{\alpha}}) = \left[\mathbf{Y}' - \widehat{\boldsymbol{\beta}}'\mathbf{X}' - \widehat{\boldsymbol{\alpha}}'\mathbf{D}' - \widehat{\eta}_0\widehat{\boldsymbol{\alpha}}'\mathbf{D}'\mathbf{W}\right]\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{D}\widehat{\boldsymbol{\alpha}} - \widehat{\eta}_0\mathbf{WD}\widehat{\boldsymbol{\alpha}}\right]$$

and from the first order conditions for minimization of $S(.)$ we get:

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

The above set of conditions makes clear that in Arcidiacono's *et al* peer effects model there is no requirement that $\mathbf{D}'\mathbf{e} = \mathbf{0}$ meaning that the coefficients of time-invariant variables associated with the worker may be identified.[28] These f.o.cs can be solved iteratively by alternating between the solution of each condition. But this approach is complicated by the high-dimensionality of $\mathbf{D}$ (and possibly that of other fixed effects included in $\mathbf{X}$). The main obstacle is solving the condition $[\mathbf{D}' + \widehat{\eta}_0\mathbf{D}'\mathbf{W}]\,\mathbf{e} = \mathbf{0}$. Conditional on $\widehat{\eta}_0$ we can solve this f.o.c iteratively. Rewriting

$$\left[\mathbf{D}' + \widehat{\eta}_0\mathbf{D}'\mathbf{W}\right]\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{D}\widehat{\boldsymbol{\alpha}} - \widehat{\eta}_0\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}}\right] = \mathbf{0}$$

and rearranging and solving for $\mathbf{D}'\mathbf{D}\widehat{\boldsymbol{\alpha}}$

$$\mathbf{D}'\mathbf{D}\widehat{\boldsymbol{\alpha}} = \mathbf{D}'\left[\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{Y} - \mathbf{D}'\left[\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{D}'\widehat{\eta}_0\left[2\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}}$$

and now premultiplying by $[\mathbf{D}'\mathbf{D}]^{-1}$ and letting $\mathbf{M}_\mathbf{D} \equiv [\mathbf{D}'\mathbf{D}]^{-1}\mathbf{D}'$ we obtain

$$\widehat{\boldsymbol{\alpha}} = \mathbf{M}_\mathbf{D}\left[\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right][\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}] - \widehat{\eta}_0\mathbf{M}_\mathbf{D}\left[2\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}}$$

The above expression provides a natural way to solve recursively for $\widehat{\boldsymbol{\alpha}}$ and this is basically the suggestion in Arcidiano et al (2012). Plug values for $\widehat{\boldsymbol{\alpha}}$ on the right hand side and solve for the $\widehat{\boldsymbol{\alpha}}$ on the left hand side. More specifically, letting h index iteration the updating equation becomes

$$\widehat{\boldsymbol{\alpha}}_{[h]} = \mathbf{M}_\mathbf{D}\left[\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right][\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}] - \widehat{\eta}_0\mathbf{M}_\mathbf{D}\left[2\mathbf{I} + \widehat{\eta}_0\mathbf{W}\right]\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}}_{[h-1]} \qquad (.4)$$

Computation of the above expression is simple because it involves mostly the calculation of group averages. The first two conditions can be solved by running an OLS regression. Thus, an algorithm for estimation would alternate between the following steps:

28. In the traditional fixed-effects model, $\mathbf{Y} = \mathbf{X}\beta + \mathbf{D}\alpha + \varepsilon$, the first order conditions are $\mathbf{X}'\mathbf{e} = \mathbf{0}$ and $\mathbf{D}'\mathbf{e} = \mathbf{0}$. This means that the coefficient on a time-invariant characteristic of the individual cannot be identified because the variable can be expressed as $\mathbf{Dz}$ (where $\mathbf{z}$ is a vector of length N) and since $\mathbf{D}'\mathbf{e} = \mathbf{0}$ the f.o.c. associated with that variable, $\mathbf{z}'\mathbf{D}'\mathbf{e} = \mathbf{0}$, is redundant.

- *Step 1* Given $\widehat{\boldsymbol{\alpha}}$ run an OLS regression on $\mathbf{X}$, $\mathbf{D}\widehat{\boldsymbol{\alpha}}$, and $\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}}$. The coefficients on $\mathbf{X}$ will provide an estimate of $\boldsymbol{\beta}$, while the coefficient on $\mathbf{W}\mathbf{D}\widehat{\boldsymbol{\alpha}}$ is an estimate of η_0. $\mathbf{D}\widehat{\boldsymbol{\alpha}}$ should have a coefficient of 1.
- *Step 2* Given $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta_0}$ estimate $\boldsymbol{\alpha}$ using the updating equation in (.4)

There is, however, a faster approch to solve the f.o.c. $[\mathbf{D}' + \widehat{\eta_0}\mathbf{D}'\mathbf{W}]\,\mathbf{e} = \mathbf{0}$. Rewrite the condition as

$$\mathbf{D}'\widetilde{\mathbf{W}}\left[\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \widetilde{\mathbf{W}}\mathbf{D}\widehat{\boldsymbol{\alpha}}\right] = \mathbf{0}$$

where $\widetilde{\mathbf{W}} = \mathbf{I} + \widehat{\eta_0}\mathbf{W}$. We can then rewrite the equation as

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

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

Appendix A.2 - Calculation of standard errors

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). The estimated variance-covariance matrix of this linear regression provides a valid estimate of the covariance matrix of the NLS estimates.Thus, for our case and after some simplication, the GNR becomes,

$$\mathbf{y} + \eta_0^o\mathbf{W}\mathbf{D}\boldsymbol{\alpha}^o = \mathbf{X}\boldsymbol{\beta} + [\mathbf{I} + \eta_0^o\mathbf{W}]\mathbf{D}\boldsymbol{\alpha} + \eta_0\mathbf{W}\mathbf{D}\boldsymbol{\alpha}^o + \varepsilon \qquad (.5)$$

Unfortunately, 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 which may be present in $\mathbf{X}$. But 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 $[\mathbf{I} + \eta_0^o\mathbf{W}]\mathbf{D}$ and calculate a matrix that contains only the estimates of the variance-covariances associated with the set of parameters we are interested in. To clarify let $\mathbf{X} = [\mathbf{X}_1\,\mathbf{X}_2]$ where $\mathbf{X}_1$ represents the regressors of interest and $\boldsymbol{\beta}_1$ the corresponding coefficients. Thus, to estimate the variance covariance matrix of the estimators of $\boldsymbol{\beta}_1$ we have to regress each element of $\mathbf{X_1}$ on $\mathbf{X_2}$ and $[\mathbf{I} + \eta_0^o\mathbf{W}]\mathbf{D}$ and calculate the residuals which we collect into a matrix denoted by $\mathbf{X_1^*}$. We do the same for the dependent variable $\mathbf{y} + \eta_0^o\mathbf{W}\mathbf{D}\boldsymbol{\alpha}^o$ and call the residual $\mathbf{y}^*$. Finally, we calculate the residual associated with the variable $\mathbf{W}\mathbf{D}\boldsymbol{\alpha}^o$ which we denote by $\mathbf{w}^*$. To implement these non-trivial regressions we use a similar strategy as detailed above for the calculation of the NLS estimates.

The estimated variance-covariance matrix of the linear regression shown below provides estimates for the NLS model:

$$\mathbf{y}^* = \mathbf{X}_1^* \boldsymbol{\beta}_1 + \eta_0 \mathbf{w}^* + \boldsymbol{\varepsilon}$$

With proper correction for degrees of freedom the (cluster) robust covariance matrix estimator implied by the above regression can also be used for the NLS regression. The *Stata* ado file *regpeer* coded by one of the authors implements the approach discussed above. This file relies heavily on Sergio Correia's *reghdfe* command for efficient estimation of linear regressions with high-dimensional fixed effects (Correia (2014)).

Appendix A.3 - Tables and Figures

	(1)
Log wages	0.3588
	(0.5482)
Age	37.4770
	(10.5699)
Tenure	8.6532
	(8.7654)
Firm size (log)	4.7029
	(2.0508)
Gender (Female=1)	0.4260
	-
Schooling	8.0009
	(4.0133)
N	19,152,256

Table A.3.1. Summary Statistics

Note: This table reports the summary statistics from Quadros de Pessoal (1994-2013).

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

TABLE A.3.2. Statistical Moments from Wage Distribution - firm/job-title specification

Note: The statistics are computed from the estimates given in Column (2) from Table 3. Panel A gives the variance decomposition according to the covariances between wages and the components of the wage equation. Panel B shows the correlations between the worker, coworker, and firm/job-title fixed effects. Panel C provides the standard deviations of worker, coworker, firm/job-title fixed effects, and the average of the standard deviations of the measure of peer quality (as measured by the fixed effect of the peers).

	Full Specification		Excluding Covariates		Excluding Covariates and Larger Peer Groups	
	(1)	(2)	(3)	(4)	(5)	(6)
HC spillovers ($\overline{\alpha}_{-it}$)	0.5007 (0.0006)	0.2050 (0.0006)	0.4540 (0.0006)	0.1989 (0.0005)	0.3619 (0.0008)	0.1027 (0.0011)
Time effects (μ_t)	✓		✓		✓	
Worker effects (α_i)	✓	✓	✓	✓	✓	✓
Firm/Job-title effects ($\theta_{\mathbf{F}\times\mathbf{J}(i,t)}$)	✓		✓		✓	
Establishment/job-title/year effects		✓		✓		✓
N	19,152,256	19,070,170	19,152,256	19,070,170	12,188,465	12,188,465

Table A.3.3. Sensitivity of the Human Capital Spillovers to the Presence of Covariates and the Size of Peer Groups

Notes: Columns (1) and (2) recover the full specification results inserted in Table 3. Columns (3) and (4) give the human capital spillover effect excluding all the covariates (à la Arcidiacono). Columns (5) and (6) report the same coefficient estimates excising, in addition, peer groups larger than 10 individuals.Standard errors are obtained as explained in Appendix A.2.