

IZA DP No. 7276

The Sources of Wage Variation: A Three-Way High-Dimensional Fixed Effects Regression Model

Sónia Torres
Pedro Portugal
John T. Addison
Paulo Guimarães

March 2013

The Sources of Wage Variation: A Three-Way High-Dimensional Fixed Effects Regression Model

Sónia Torres

Statistics Portugal

Pedro Portugal

Banco de Portugal, Universidade Nova de Lisboa and IZA

John T. Addison

University of South Carolina, TU Chemnitz and IZA

Paulo Guimarães

CEFUP Universidade do Porto and IZA

Discussion Paper No. 7276

March 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Sources of Wage Variation: A Three-Way High-Dimensional Fixed Effects Regression Model

This paper estimates a wage equation with three high-dimensional fixed effects, using a longitudinal matched employer-employee dataset covering virtually all Portuguese wage earners over a little more than two decades. The variation in log real hourly wages is decomposed into different components related to worker, firm, and job title characteristics (both observed and unobserved) and a residual component. It is found that worker permanent heterogeneity is the most important source of wage variation (36.0 percent) and that the unobserved component plays a more important role (21.0 percent) than the observed component (15.0 percent) in explaining wage differentials. Firm permanent effects are less important overall (28.7 percent) and are due in almost equal parts to the unobserved component and the observed component. Job title effects emerge as the least important dimension but they still explain close to 10 percent of wage variation. We found definitive evidence of positive assortative matching.

JEL Classification: J2, J41

Keywords: high dimensional fixed effects, wage decomposition, assortative matching

Corresponding author:

Pedro Portugal
Banco de Portugal
Av. Almirante Reis, 71
1150-165 Lisboa
Portugal
E-mail: jppdias@bportugal.pt

I. Introduction

An important research theme in labor economics is why similar workers receive different remuneration and why similar firms pay different wages (Diamond, 1982). There are two lines of reasoning to explain observed wage variability, one of which relies on the supply-side determinants of wages (workers' characteristics) and the other on demand-side factors (employers' characteristics).

In a labor market operating under perfect competition, each worker should receive a wage that equals his or her marginal (revenue) product. Wage differentials should reflect differences in worker productivity rather than depend on job or employer attributes (other than those affecting worker utility such as dangerous working conditions that will in normal circumstances attract a compensating differential). In turn, worker productivity has a basis in competence – whether observed or not – typically 'acquired' through investments in human capital. Here we are abstracting from issues of unobserved intrinsic ability (Griliches, 1977) and associated signaling models (Spence, 1973).

There is no shortage of models seeking additional or alternative explanations for wage variability, but in each case the characteristics of firms rather than those of workers (i.e. worker competence or productivity differences) now play a leading role. Given the plethora of such treatments,¹ we choose to focus here on just two of them that pose perhaps the sharpest contrast with the standard competitive model. The first approach has a basis in rent-sharing/insider-outsider considerations, while the second emphasizes labor market frictions.

Rent-sharing models predict that wages depend on the employer's ability to pay. In particular, wages are predicted to have a positive correlation with firm profits, since firms may find it beneficial to share their gains with their workers and pay above the going rate.² These models explain why wages depend not only on external labor market conditions but also on the conditions inside the firm, including its productivity, profits, degree of competition, turnover costs, and the bargaining strength of workers. They also explain why the wages of workers from different groups of occupations, educational categories, and seniority tiers are higher in some firms or industries than in others.

The other explanation for wage differentials among workers with similar characteristics targeted here derives from the job search and matching literature and emphasizes the role of labor market frictions in wage determination. Thus, the

equilibrium job search model of Burdett and Mortensen (1998) predicts that firms may have incentives to offer higher wages than their competitors in order to guarantee a low quit rate and attract a large number of workers in a market characterized by the existence of frictions – even in circumstances of homogeneous workers and firms *ex ante*. This model predicts that wages are increasing in firm size and workers' job seniority.

For their part, matching models that also take into account the existence of frictions in the labor market provide an explanation for wage dispersion. In the models of Diamond (1982) and Mortensen and Pissarides (1994) while the wage is set by the employer, workers and firms bargain over the share of the matching rent after they meet *ex post*. Differences in match productivity, then, explain why similar workers (firms) may receive (offer) different wages.³

Our goal in the present exercise is to appropriately disentangle the effects of employers' decisions (demand-side determinants of wages) from the effects of choices made by workers (supply-side determinants) in the explanation of wage variability. To this end, researchers have estimated wage regressions incorporating both worker and firm fixed effects. However, besides worker and firm heterogeneity, a third important dimension of wage formation is *job title heterogeneity*, reflecting the distinct tasks performed by workers that define the set of occupational boundaries. There are a variety of reasons why job title heterogeneity can be expected to influence wage rates. One is compensating advantages for riskier and/or less pleasant working environments. Another is the heavy doses of job specific training that some jobs may entail. Additional reasons include occupational crowding and active rent seeking.

To properly incorporate these and other such wage determinants one needs a very detailed accounting of the kind of jobs being undertaken by workers. Even a highly disaggregated occupational count would not be fit for purpose here because an employers' wage policy regarding the same occupation (e.g. a secretary) might be governed by different collective agreements (say the banking industry collective agreement as opposed to that for the retail trade sector). Fortunately, our dataset contains an unusually rich set of information enabling us to identify the collective agreement that regulates the employment contract applicable to each worker. Moreover, within each collective agreement, we can further pinpoint the exact, detailed occupational category of each worker. Each year, around 300 different collective agreements are negotiated in Portugal (see below) that define wage floors

for each particular job title (so-called *categoria profissional*). On average, each collective agreement defines the wage floor for around 100 job titles. Overall, in a given year, there are 30,000 collective agreement/job title combinations to which workers can be allocated. The main use of the dataset – the *Quadros de Pessoal* (see below) – is precisely to enable the officials of the Portuguese Ministry of Employment to ascertain whether employers are in compliance with what was actually agreed to at the bargaining table (i.e. wages, work schedules, and other conditions). This recording obligation also serves to underscore the accuracy of the Portuguese data.

By properly taking job title effects into account one should be able to provide refined estimates of worker and firm fixed effects, and shed additional light on the current debate concerning the role of assortative matching. In the process, we should also be able to unambiguously disentangle the joint contribution of contract heterogeneity and occupation heterogeneity to wage formation.

The objective of this estimation is to calculate the contribution of worker, firm, and job-title fixed effects to overall wage variability. The requirements of this decomposition exercise are daunting; specifically, the availability of longitudinal datasets combining information on firms and their employees (namely, matched employer-employee datasets with unique identifiers for firms, workers, and job titles) and the use of appropriate panel data econometric techniques to estimate three high-dimension fixed effects in wage equations. Fortunately, panel datasets have become available in recent years for many countries, while econometric tools (and computing capacity) have also improved greatly. Taken in conjunction, all three ingredients – data, econometric techniques, and computing facilities – have made it possible to bring new information to bear in the empirical debate on (many aspects of) wage determinants. Most notably, in their pioneering work using a French longitudinal matched employer-employee dataset, Abowd, Kramarz, and Margolis (1999) were the first to propose an empirical framework for estimating worker and firm effects in wage equations. They reported that worker characteristics explained the major part of wage differentials, of inter-industry wage differentials, and of firm-size wage differentials.

In the present treatment, we use a longitudinal matched employer-employee dataset covering virtually all employees in Portugal. Our dataset contains a total of a little more than 27 million observations, 1986-2006, drawn from 568 thousand firms and 5.5 million workers. In estimating a wage equation that includes worker and firm

effects, we use a routine that was especially developed in *Stata* providing an exact solution to the least squares problem that arises when dealing with very high dimension matrices. As noted earlier, we have taken this methodology a stage further by including a third fixed effect in our wage equation so as to control explicitly for job title heterogeneity.

The plan of the paper is as follows. A literature review on assortative matching theory (on the complementarity between individual and firm productivity levels) is next provided in section II. Since we can directly investigate the association between the two main forms of heterogeneity that have figured centrally in the matching literature, while introducing a third, we can provide new insights into a frankly unresolved literature. The general empirical framework necessary to estimate wage equations with worker, firm, and job title fixed effects is next established in Section III. A short data description and barebones review of wage setting in Portugal is contained in Section IV. Wage variability is decomposed into its various components in Section V, where the determinants of worker, firm, and job title fixed effects are investigated and correlations between the components of compensation also addressed. Section VI assesses the relationship between firms' wage policies and their labor force quality within the framework of assortative matching, using productivity data. Section VII concludes.

II. Assortative Matching

The sorting of heterogeneous workers across firms is the subject of heated debate. The idea behind *positive* assortative matching is the complementarity between individual and plant productivity levels, with good workers being teamed up with good firms. The theoretical basis for such matching is provided by assignment models. In his marriage market model, Becker (1973: 826) shows that if the production function is supermodular the unique equilibrium that occurs is both efficient and characterized by perfect sorting. In other words, the existence of sufficient complementarities in production generates positive assortative matching; here the union of the most (and least) desirable partners: the most desirable individuals get together, as do the least desirable. The early assignment models, however, were rooted in competitive equilibrium (e.g. Sattinger, 1993; Kremer and Maskin, 1996), thereby disregarding establishment-specific components in the wage equation. With the introduction of frictions, more recent developments have ensured a sorting of workers across plants

(Shimer and Smith, 2000; Shimer, 2005; Postel-Vinay and Robin, 2002). At issue in these models is the nature of the equilibrium matching pattern since different matching models predict different patterns (i.e. admitting of either positive or zero/negative assortative matching) according to the assumptions of the model such as strict supermodularity (i.e. all agents have higher productivity when they match with high-productivity agents), the transferability of utility, and the commitment for a wage schedule.

Empirical work – some of which is summarized below in presenting our own findings – has often failed to produce evidence of positive assortative matching in the wake of Abowd, Kramarz, and Margolis' (1999) pioneering study. Using matched employer-employee data for 1976-1987 for a 1/25th sample of the French labor force Abowd, Kramarz, and Margolis decomposed wages into fixed establishment and person effects and reported a positive albeit weak correlation between the two. However, these results were obtained on the basis of statistical approximations – limited by the capacity of the computers on which they were generated. In re-estimating the model using exact methods, Abowd, Creecy, and Kramarz (2002) report that the correlation between the person and firm effect is -0.283 (rather than 0.097 using the former methodology). The authors also report correlations between the two effects for a 1/10th sample of employees in the state of Washington, using matched data for 1984-1993. The corresponding coefficients were -0.025 and 0.050 for the exact and approximate estimates, respectively. And, to repeat, negative correlations have indeed figured largely in the literature using the wage data approach (e.g. Goux and Maurin, 1999; Gruetter and Lalive, 2009).

Although, as we have seen, negative assortative matching may have an economic explanation (see also Woodcock, 2010), considerable effort has been expended to determine whether this result might be an artifact of the use of standard econometric techniques. Abowd *et al.* (2004) test and discount the notion that the negative correlation between the fixed worker and employer effects – *vulgo*: good workers gravitate to bad firms – is caused by limited mobility bias in the estimation of each effect. They conclude that while sampling error does impart downward bias to the two effects, its magnitude is simply too small to modify the basic negative result for France or the absence of correlation for the United States (i.e. random assignment). A more attenuated conclusion is reached by Andrews *et al.* (2008), who show that the correlations between the two fixed effects will be downwardly biased if

there is true positive assortative matching and when any conditioning covariates are uncorrelated with the two fixed effects. The authors' simulations indicate that the extent of bias is a decreasing function of worker mobility which in turn reflects the propensity to move, the length of the panel, and the average size of firms. In applying formulae to correct the bias to West German matched employer employee data for 1993-1997, the authors find evidence of not inconsiderable bias: some 25 percent for the full sample, increasing to around 50 percent for the subsample of movers. Although in this study the biases are large, they do not overturn the negative correlation between the worker and plant effects. However, in their subsequent analysis of social security records for three German Länder, Andrews *et al.* (2012) report that low mobility bias does indeed obscure an estimated correlation that is strongly positive.

Melo (2008) also argues that the standard method to measure sorting (using worker and firm fixed effects in a log-linear wage regression as proxies for worker constant heterogeneity in the manner of Abowd, Kramarz, and Margolis, 1999) is biased against detecting it. Melo offers a model with four main components: worker and firm heterogeneity, complementarities in production (necessary to produce sorting in equilibrium), search frictions, and limitations on the ability of firms to post new vacancies. The frictions induce agents to accept suboptimal partners to avoid joblessness and the vacancy restriction creates *ex ante* rents for vacancies and provides a reason for firms to reject some workers in equilibrium. Although the model yields strong positive sorting with good workers teamed with good firms because of complementarities in production, this outcome is hidden because of non-monotonicities in the wage equation caused by the interaction between wage bargaining and the limited ability of the firms to post new vacancies. This in turn arises because high productivity firms have better outside options than their low productivity counterparts, which causes downward pressure on the wages of their workers; and in particular among low-wage workers. In other words, low skilled workers are then paid less when working for a more productive firm.

Melo's distinct solution is to examine the correlation between a worker's wage fixed effect and the average fixed effect of the coworkers in the same firm. His correction yields strong evidence of positive assortative matching, unlike the conventional measure which yields an absence of sorting when applied to Brazilian matched employer-employee data, 1995-2005. One problem with this approach – and

one admitted by the author – is that the positive association between a worker’s wage fixed effect and the average fixed effect of his/her coworkers does not in fact inform us as to the sign of sorting since good workers could be clustering in bad firms. Further, Melo’s preferred measure may not be sensitive to differences in firm characteristics such as average employee turnover and firm size.

The perception that one cannot distinguish positive from negative sorting using wage data – or the related concern that theoretical models can generate positive or negative correlations between firm and person effects from a wage equation – explains why some have advocated using a *productivity model* directly rather than inferentially. Unlike the more numerous studies employing wage data, those using output data point to positive assortative matching. As a case in point, using Portuguese matched employer-employee data from the *Quadros de Pessoal*, 1986-2000 and a translog specification, Mendes, van den Berg, and Lindeboom (2010) estimate a firm-specific productivity effect for each firm that they then relate to the skills of workers in the firm measured as the time average of the share of highly-educated workers in the firm.⁴ They report evidence of positive assortative matching, especially among longer-lived firms. They report that the results are not caused by heterogeneity in search frictions; for example, if all workers were attractive to firms but the high skilled types found it easier to locate high quality firms, one would still observe positive matching. The authors use data on job transitions to construct an index of search frictions for the various skill levels they examine within different submarkets. The test is to determine whether search frictions are high in those sectors and regions where positive matching is high. Although the correlation between search frictions and positive matching is positive, the incorporation of such frictions is to reduce the matching contribution by only 30 percent. That said, the authors’ definition of search friction is unconventional: the ratio between the probability of moving to another firm and leaving the labor force rather than the ratio of the job arrival rate and the separation rate.

Recently, a trenchant criticism of using worker and firm fixed effects to conclude anything about assortative matching has been made by Eeckhout and Kircher (2011). Their argument hinges upon non-monotonicity, which reflects the opportunity cost to the firm of a match with an inappropriate type of worker. The more productive firms run a risk (i.e. have to be compensated for) contracting with a ‘bad’ worker because it stops them contracting with a ‘good’ worker. So, a worker’s

wages are lower if he or she contracts with *either* a bad or a very good firm. What matters is the *proper match* – a worker coming together with the right firm. In other words, the highest compensation arises from correct matches and this process substitutes for a wage schedule that is increasing everywhere with type of firm. Eeckhout and Kircher (2011: 874) speak of wages for a given worker having “an inverted U-shape around the optimal allocation which corresponds to the frictionless wage.” The non-monotonic effect of firm type on wages translates into a wage that cannot then be decomposed into an additively separate worker and firm fixed effect. In this model, only the most productive firms make profits so that information on profits rather than wages is necessary to identify the sign of sorting.

Eeckhout and Kircher construct a model that allows for mismatched wages and show that if equilibrium wages are non-monotonic in firm type, the traditional method used in the literature is inappropriate in seeking to gauge the sign (and the intensity) of sorting precisely because firms pay wages based on the productivity gain from getting together with a higher type worker rather than because they themselves are productive.⁵ Although we would argue that the authors’ conclusion is sensitive to model parameterization – so that we should not throw out the decomposition exercise ‘baby’ with the bathwater – we shall further refine our treatment of assortative matching in section VII to include firm level productivity filtered from the heterogeneity of labor inputs. In this way we are able to address these authors concerns.

III. The General Empirical Framework to Decompose Wage Variation

Consider the problem of estimating a standard Mincerian wage equation to which we add three high-dimensional fixed effects to account for firm, worker and job-title heterogeneity:

$$\ln w_{ifjt} = X_{ifjt} \beta + \theta_i + \varphi_f + \lambda_j + \varepsilon_{ifjt} . \quad (1)$$

In the above equation, $\ln w_{ifjt}$ stands for the natural logarithm of the real hourly wage of individual i ($i = 1, \dots, N$) working at firm f ($f = 1, \dots, F$) and holding a job title j ($j = 1, \dots, J$) at year t ($t = 1, \dots, T_i$), whereas X_{ifjt} is a vector of k observed (measured) time-varying exogenous characteristics of individual i and firm f . There are T_i observations for each individual i and a total of N^* observations. All time-invariant

characteristics of the workers, firms and job titles are captured by the fixed effects which are, respectively, θ_i , φ_f and λ_j . According to this equation, there are five distinct sources of wage variability:

1. the observed time-varying characteristics of workers, firms, and the economy ($X_{it}\beta$);
2. time-invariant worker heterogeneity (θ_i);
3. time-invariant firm heterogeneity (φ_f);
4. time-invariant job title heterogeneity (λ_j); and,
5. unexplained random variation (ε_{it}).

Equation (1) includes three high-dimensional fixed effects. Estimation of linear regression models with more than one high-dimensional fixed effect poses some particular challenges. The problem was first tackled by Abowd and Kramarz (1999) and Abowd, Kramarz, and Margolis (1999). In their seminal paper, these authors proposed a computationally tractable solution that yielded an approximation to the full least squares solution of a linear regression model with two high-dimensional fixed effects. In a later paper, Abowd, Creedy, and Kramarz (2002) presented a conjugate gradient algorithm that led to the exact least squares solution of this problem. More recently, Guimarães and Portugal (2010) showed that with a full Gauss-Seidel iterative algorithm it is possible to obtain the exact least squares solution for linear regression models with two or more high-dimensional fixed effects. In the Appendix we provide a more detailed description of this approach discussing its application to the estimation of equation (1).

IV. Data and Institutional Context

Data

The Portuguese data used in this inquiry come from a longitudinal matched employer-employee dataset known as the Tables of Personnel (or *Quadros de Pessoal*) for the years 1986 to 2006 (excepting 1990 and 2001). This unique dataset was created by the Portuguese Ministry of Employment, and is taken from a mandatory annual survey addressed to firms with wage earners. The survey covers various firm and establishment characteristics, as well as a set of characteristics of the workforce (see below). Being compulsory, it does not suffer from the non-response

problems that often plague standard household and firm surveys. Further, the survey covers all Portuguese wage earners, with the exceptions of the Public Administration sector and domestic servants.

Turning to specifics, the dataset includes information on the establishment (establishment identifier, location, industry, and employment), the firm (firm identifier, location, industry, legal form, ownership, year of formation, employment, sales, and capital), and its workers (social security identifier, gender, age, education, skills, occupation, employment status, professional level, seniority, earnings [base wage, seniority-related earnings, other regular and irregular benefits, and overtime pay], normal and overtime hours, time elapsed since last promotion, professional category and the corresponding classification in a collective agreement).

For the purposes of this exercise, a subset of variables was selected, certain new variables created, and some observations removed. The final set of variables retained for analysis is given in Table A.1. Among the restrictions placed on the data were the exclusion of those individuals who were not working full time, who were aged less than 18 years or more than 60 years, who earned a nominal wage less than 80 percent of the legal minimum wage or above the 99.9 percent quantile in each year, who recorded errors in their admission/birth dates, and who had duplicate social security codes or other errors in those codes.⁶ We also dropped close to 2 percent of observations that did not belong to the largest connected set (see Appendix). The final dataset for all 19 available years comprises 27,020,044 observations drawn from 567,739 different firms, 5,492,332 individual workers, and 95,927 job titles (i.e. the code of the variable that results from the conflation of the professional category variable and the corresponding collective agreement variable).

Institutional Wage Setting

Wage setting in Portugal is dominated by the presence of mandatory minimum wages and by the widespread use of government extensions of sectoral agreements. There is a modicum of firm-level bargaining but formally decentralized bargaining of this nature is the exception rather than the rule – covering less than 10 percent of the workforce – and often taking place in large enterprises that were formerly part of the public sector. Sectoral agreements, conducted by employer and union confederations, may cover a wide range of industry-specific occupations but the system does not rule out *parallelism* or overlapping collective agreements, such that a single enterprise may be covered by two or more agreements depending on the union affiliation of its

workers. Indeed, the situation may be further stratified if the firm in question straddles more than one line of economic activity, thereby belonging to more than one employers' association. As a result of union fragmentation, therefore, several agreements may coexist for the same region, occupation, and firm. Horizontal agreements, covering a number of sectors, are also possible, but are not frequent. Overall, coverage of collective agreements in the Portuguese private sector is above 90 percent.

Collective bargaining in Portugal differs from that in other nations by virtue of its fragmentation and extent of multiunionism. The corollary is that the contents of collective agreements are at once extensive and general. They are extensive insofar as they cover many categories of workers. They are general in that they set only minimum conditions of which the most important is base level monthly wages – though others include normal working hours and overtime pay. The focus is upon wage floors rather than anticipated wage growth that in some centralized bargaining regimes (e.g. Sweden) is then incorporated into sectoral agreements. In consequence, employers have freedom of manoeuvre to tailor remuneration to their prevailing economic circumstances (on the determinants of the contractual wage and this 'wage cushion,' see Cardoso and Portugal, 2005).

The most relevant mechanism shaping the formation of wages is the systematic extension of industry-wide agreements by the Ministry of Employment. Even though by law the collective agreement only binds the trade union members and the employer associations' affiliated firms that are parties to the agreement, there is no legal mechanism that obliges the trade unions and the employers association to reveal their constituency. This legal conundrum is almost always circumvented by extending the agreement to the whole sector through the use of "*portarias de extensão*". This means that even wage agreement reached by trade unions and employers associations with very low representation have a strong impact in setting wage floors. Indeed, in any given year, collective bargaining sets around 30,000 minimum wages that correspond to 30,000 job-titles.

Finally, wage floors are also set under national minimum wage machinery, established in 1974. The minimum wage can exceed that set under sectoral bargaining. In this event of course the former dominates. Currently, the national minimum wage covers some 16 percent of full-time wage earners.

V. The Role of Individual, Firm, and Job Title Heterogeneity in Wage Differentials

In order to decompose wage variability into the components identified earlier, we first estimated equation (1). The explanatory variables (or observed time-varying characteristics) are age, age squared, seniority, seniority squared, firm size, and year dummies. The dependent variable is the natural logarithm of the real hourly wage.

(Table 1 near here)

The results are reported in Table 1. Observe that the R^2 of this equation is considerably higher than in standard wage regressions. The worker fixed effects, firm fixed effects, job title fixed effects, and worker and firm time-varying characteristics together explain 93.5 percent of the variability in real wages. As expected, wages increase with age and seniority at a decreasing rate. And, familiarly, larger firms pay higher wages.

In this framework, it will be recalled that the worker fixed effects (θ_i) include both the workers' unobserved and observed but non-time-varying characteristics. Similarly, the firm fixed effects (φ_f) and job title fixed effects (λ_j) include both the unobserved and observed but non-time-varying firm and job title characteristics, respectively. We decomposed the three estimated fixed effects into these components by estimating the following three regression equations: first,

$$\hat{\theta}_i = \text{const.} + W_i\eta + \varepsilon_i, \quad (4)$$

where W_i is a vector of non-time-varying worker characteristics (comprising gender and five education dummies), η is the associated vector of coefficients, and $W_i\eta$ is the worker non-time-varying observed characteristics component. Note that α_i , the worker specific intercept – capturing the effect of worker unobserved characteristics and that can be interpreted as the opportunity cost or the market valuation of worker heterogeneity – is obtained residually by $\hat{\alpha}_i = \hat{\theta}_i - W_i\hat{\eta}$; second,

$$\hat{\varphi}_f = \text{const.} + Z_f\gamma + \varepsilon_f, \quad (5)$$

where Z_f is a vector of non-time-varying firm characteristics (four regional dummies, capital ownership – specifically, the shares of domestic and public capital – and twenty-eight industry dummies), γ is the associated vector of coefficients, and

$Z_f \gamma$ is the firm non-time-varying observed characteristics component.⁷ As before, the firm-specific intercept, ϕ_f , capturing the firm unobserved characteristics effect, is obtained residually, by $\hat{\phi}_f = \hat{\phi}_f - Z_f \hat{\gamma}$; and third,

$$\hat{\lambda}_j = FE_{occup} + FE_{ca} + \varepsilon_j, \quad (6)$$

where the sum of the two fixed effects (FE_j), one for the occupation variable FE_{occup} and the other for the collective agreement variable FE_{ca} , corresponds to the non-time-varying observed characteristics component, and δ_j , the job title specific intercept capturing the job title unobserved characteristics effect, is obtained residually, by $\hat{\delta}_j = \hat{\lambda}_j - FE_j$.

We now have the following compensation components (plus the residual):

- $X_{ijt} \hat{\beta}$: observed firm, worker, and economy time-varying characteristics that comprise three components: time dummies, time-varying characteristics of workers, and time-varying characteristics of firms.
- $\hat{\theta}_i$: worker effects.
 - $W_i \hat{\eta}$: observed worker non-time-varying characteristics.
 - $\hat{\alpha}_i$: unobserved constant worker characteristics.
- $\hat{\phi}_f$: firm effects.
 - $Z_f \hat{\gamma}$: observed firm non-time-varying characteristics.
 - $\hat{\phi}_f$: unobserved constant firm characteristics.
- $\hat{\lambda}_j$: job title effects.
 - $F\hat{E}_j$: observed job title non-time-varying characteristics.
 - $\hat{\delta}_j$: unobserved constant job title characteristics.

Tables 2 and 3 report the estimation results for the worker fixed effects and the firm fixed effects regressions, respectively. Beginning with Table 2, we observe that the worker fixed effect for females is on average 14.7 percent smaller than that for men. Further, there is an increasing *premium* associated with the education level: a

worker who has completed the second stage of tertiary education shows a fixed effect that is on average 71.5 percent larger than that of a worker with pre-primary or no formal completed education (the reference category). Note that these results are *pure* effects; that is, they result from a regression in which the dependent variable (the worker fixed effect) was estimated through a regression that controlled simultaneously for the time-varying characteristics of workers and firms and for firms' heterogeneity. Overall, these non-time-varying worker characteristics explain 27.9 percent of the variability in worker fixed effects.

(Table 2 near here)

From Table 3 we see that the geographic location of the firm, its capital ownership and size (as measured by the number of employees) as well as industry affiliation play important roles in explaining the differences in the firm fixed effects. Specifically, the firm fixed effects are on average larger in all NUTS II regions than in *Norte* (the reference category); the firm fixed effects tend to be higher among firms with larger shares of non-domestic or public capital; and there is also strong evidence of material differences in firm fixed effects across different industries. Note again that these effects are *pure* effects, as they result from a regression in which the dependent variable (the firm fixed effect) was estimated through a regression that controlled simultaneously for time-varying characteristics of workers and firms and for workers' heterogeneity.

(Table 3 near here)

The estimation results for the job title fixed effects regression are not reported here as the explanatory variables are two high-dimension fixed effects. Note that equation (6) has a different specification from equations (4) and (5) above. This is due to the nature of the explanatory variables chosen for equation (6). Occupation and collective agreement are both categorical variables with too many outcomes to be included as dummy variables (4,328 and 943 different outcomes, respectively, for the entire period). Therefore, we decided to include them as two fixed effects. This is equivalent to the least square dummy variable approach (LSDV) of a fixed effects estimation.

We can summarize the estimation results as follows: the R^2 of this equation is 0.628, meaning that the two non-time-varying job title characteristics (occupation and collective agreement) explain 62.8 percent of the variability in job title fixed effects. The largest role is attributable to occupation, as the R^2 of an equation containing only

this explanatory variable explains 46.2 percent of the variability in job title fixed effects, whereas the R^2 of an equation with just the collective agreement argument explains 16.6 percent of that variability.

(Table 4 near here)

Descriptive statistics for the components of real compensation by gender are provided in Table 4. For all the components of real compensation, the averages for males are higher than those for females, other than the predicted effect of time. The gender differences are greater for the worker fixed effects component than for either the firm fixed effects or the job title fixed effects components (14.3 percent, 5.7 percent, and 3.5 percent, respectively). Within each of the three components, gender differences are greater for the observed sub-components: 14.3 percent for the gender and education sub-component of worker fixed effects; 4.3 percent for the region, ownership, and industry sub-components of firm fixed effects; and 3.0 percent for the occupation and collective agreement sub-components of job title fixed effects. In addition, the variability of worker fixed effects is greater than the variability of firm fixed effects and the variability of firm fixed effects is greater than the variability of job title fixed effects. Male workers exhibit higher variability in almost all wage components (except for the time-varying observable characteristics of firms and for the education and gender sub-component of worker fixed effects).

(Table 5 near here)

In Table 5, we report the correlations among the components of real hourly wages. Of the four main components – time-varying characteristics, worker fixed effects, firm fixed effects, and job title fixed effects – the worker fixed effects component shows the highest correlation with log real total compensation (0.74), next followed by the firm fixed effects component (0.67), then by the individual and firm time-varying characteristics component (0.54), and finally by the job title fixed effects component (0.52). Both the observed and unobserved components of the worker fixed effect are highly correlated with the log of real total compensation (0.58 and 0.51, respectively). Concerning the components of the firm fixed effect, the observable part is that most highly correlated with log real total compensation (0.54). The unobserved part of the firm component is less important in determining total compensation. As regards the components of the job title fixed effect, the observable part is also the most highly correlated with the log of real total compensation (0.53), while the unobserved part is practically irrelevant in determining total compensation. In sum,

the observable part of each component is more highly correlated with the log of real total compensation than the unobservable part. For purposes of comparison, and abstracting from differences in estimation method, explanatory variables, and the number of fixed effects included in equations (1) and (2), we note that Abowd, Creedy, and Kramarz (2002) found that for France the correlations between the log of real total compensation and the worker fixed effects and the firm fixed effects were 0.70 and 0.20, respectively. (Corresponding values for the state of Washington were 0.51 and 0.52.)

In addition, we find that the correlation between firms' wage policies (as proxied by the firm fixed effects) and the quality of their workforce (captured by the worker fixed effects) is positive (0.27). Although not large, this value is nonetheless much larger than that reported in the literature. For example, Abowd, Creedy, and Kramarz (2002) report a negative correlation for France and a correlation close to zero for the state of Washington. (See also the lower estimates in Goux and Maurin, 1999, using Labor Force Survey data.)

The correlations in Table 5 also suggest an interpretation in terms of sorting. In terms of observable characteristics, there is evidence of good workers tending to be found in high-paying firms: the correlation coefficient between the corresponding components of the firm and worker fixed effects is 0.33. These results are, then, partly consistent with this literature. We should however resist the notion, as discussed earlier, that this positive correlation can be interpreted as evidence of complementarity between worker and firm levels of productivity.

Finally, the correlation coefficient between worker fixed effects and job title fixed effects (0.42) is larger than the correlation coefficient between firm fixed effects and job title fixed effects (0.17). The latter effect indicates that high paying jobs tend to go hand in hand with high-paying firms. In both cases, the correlations are larger in terms of the *observable* characteristics of workers and firms (0.38 and 0.19, respectively).

On the whole, these results indicate that the relationship between firms' wage policies and the quality of the workers they select is positive and that there are certainly factors other than wage policies that explain the distribution of high-ability workers across firms.

Next, to measure the contributions of worker, firm, and job title characteristics – both observed and unobserved – to wage variation, we used the following equation:

$$\ln w_{ijt} = X_{ijt} \beta + \alpha_i + W_i \eta + \phi_f + Z_f \gamma + \delta_j + FE_j + \varepsilon_{ijt} = \sum_{p=1}^{10} C_{ijt}^p, \quad (7)$$

where C_{ijt}^p is the p^{th} component (where X_{ijt} now comprises three components, as described above) that contributes to explaining wage variation. The contribution of each component, C_{ijt}^p , can be calculated as:

$$\text{Cov}(\ln w_{ijt}, C_{ijt}^p) / \text{Var}(\ln w_{ijt}), \quad (8)$$

where $\sum_{p=1}^{10} \text{Cov}(\ln w_{ijt}, C_{ijt}^p) / \text{Var}(\ln w_{ijt}) = 1$.

In Table 6, we report the contribution of each component to the real hourly wages variability.

(Table 6 near here)

The largest contribution to wage variation comes from worker fixed effects (36.0 percent), followed by firm fixed effects (28.7 percent), by individual, firm, and economy time-varying effects (17.4 percent), and only then by job title effects (9.7 percent). There is a residual contribution of 8.1 percent. Accordingly, comparing worker and job title effects, for example, it is evident that what workers ‘are’ is more important than what workers ‘do.’

For the worker fixed effects, the unobserved sub-component makes a larger contribution (21.0 percent) than do the gender and education sub-components (15.0 percent). For the firm fixed effects, the two sub-components’ contributions are closely similar (at 14.6 percent and 14.0 percent for the unobserved and observed components, respectively). And for the job title fixed effects, the unobserved component makes a much smaller contribution (1.9 percent) than does the observed component (7.9 percent).

(Table 7 near here)

Finally, for purposes of comparison, we contrast the main findings of this section with those from the estimation of a wage equation similar to (1) but with only two fixed effects, namely, worker and firm fixed effects. The chief results of this comparison are given in Table 7 and are as follows. First, the R^2 of the three fixed effects equation is 2 percentage points (p.p.) larger. Second, the correlations between the compensation components and the real hourly wage are similar in both estimations. Third, the contribution of the predicted effects of the time-varying

arguments is modestly larger in the three fixed effects estimation. Fourth, the worker fixed effects are reduced significantly (by 10.2 p.p.) in the three fixed effects specification, mainly by virtue of the permanent observed sub-component. It would appear that the simpler model attributes to worker heterogeneity variation stemming from occupational heterogeneity and union rent seeking, even if it is still the case that what workers are is more important than what they do. Finally, the ranking of the main components is preserved across specifications.

VI. The Relationship between Firms' Productivity and their Labor Force Quality within the Framework of Assortative Matching

The literature reviewed earlier in section II pointed to difficulties in interpreting the correlation *between* worker and firm fixed effects estimated from wage equations that ultimately frustrate any attempt to measure the degree of assortative matching in the labor market. Two main explanations have been offered for this problem: first, economic mechanisms related to the non-monotonicity of wages, as elaborated by Melo (2008) and Eeckhout and Kirscher (2011); and, second, the problem of limited mobility bias identified by Andrews *et al.* (2012) associated with short samples where the number of inter-firm movers is very small.

Our results do not suffer from the second type of bias, as the data we are using correspond to the universe (rather than a sample) of Portuguese wage-earners in the private sector. Indeed, the correlation between the worker and firm fixed effects in Portugal – each estimated from wage equation (1) – is 0.27. The magnitude of this correlation lies in the interval estimated by Andrews *et al.* (2012) when the number of movers per establishment is sufficiently large (at least 25 percent), namely from 0.2 to 0.3. For this reason we need only tackle the first problem, which we can do using our existing estimation techniques and dataset.

There is a general consensus that good workers (i.e. more productive ones) tend to earn higher wages. Therefore, it is possible to rank workers' productivity based on the individual permanent component of their wages, namely the worker fixed effects estimated from wage equations. Similarly, good firms (i.e. more productive ones) tend to have higher profits. However, these firms may pay lower or higher wages due to the presence of non-monotonicities in the wage schedule. Indeed, high-productivity firms have better outside options than their low-productivity counterparts, which may exert downward pressure on their workers' wages. This can be particularly relevant

for low-skilled workers who may end up being paid less than if working for less productive firms (Melo, 2008). Non-monotonicities in the wage schedule also mean that wages reflect the marginal contribution to the value that the firm generates; and it can be either the more productive or the less productive firms that have a higher marginal benefit from employing a better worker (Eeckhout and Kirscher, 2011). As a result, wages do not necessarily increase with firms' productivity such that simply ranking firms according to the wages they pay will not identify the most productive ones. Minimally, without additional data on the productivity of firms, it will not be possible to determine whether sorting is positive or negative.

In place of the correlation between worker and firm fixed effects estimated from our wage equation, we instead correlated the worker fixed effects estimated from equation (1) with a measure of the unobserved component of firms' productivity taken from productivity data. To this end, we estimated a 'productivity' equation, where the dependent variable is the natural log of real sales per worker at firm f in year t ($\ln s_{ft}$).

Specifically, using worker-level information and a two-way high-dimensional fixed effects procedure, we estimated the following equation:

$$\ln s_{ijt} = \mu_f + \eta_j + \delta_t + \omega_{ijt}, \quad (9a)$$

where μ_f denotes the firm (productivity) fixed effect, η_j denotes the job title (productivity) fixed-effect, δ_t denotes the year fixed effects, and ω_{ijt} is a random error term. In essence, we are filtering the firm productivity variable from aggregate conditions and job title heterogeneity.

In addition, using information at the firm-job title level, we generalized our estimation procedure to allow for the estimation of job title specific regression coefficients, to estimate the following regression model:

$$\ln s_{jft} = \mu_f + \gamma_j \text{share}_{jft} + \delta_t + \omega_{jft}, \quad (9b)$$

where share_{jft} represents the fraction of workers with job title j at firm f at year t ($\sum_j \text{share}_{jft} = 100\%$). This latter specification is more faithful to the notion of a labor heterogeneous production function. As the productivity data are given at the firm level, each firm was assigned a weight corresponding to its size, as indexed by the number of workers.

Alternatively, using now firm level data, we considered a “production function” where we took (the log of) sales as the dependent variable, and included up to 41 distinct worker qualification levels as labor inputs to estimate the following regression model:

$$\ln sales_{ft} = \mu_f + \sum_{k=1}^{41} \xi_k \ln QL_{fk} + \delta_t + \varpi_{ft} \quad (9c)$$

where QL_K denotes the number of workers with a level of qualification k .

Two further restrictions were also imposed prior to estimating equations (9a) through 9(c); specifically, firms had to be in the dataset for at least five years and to employ, over the whole period, at least fifty workers. The final dataset used for the estimation of these two equations comprised 25,518,858 worker/year pairs.

Finally, for a subset of firms (444,563) for which we were able to merge with the IES (*Inquérito às Empresas Simplificado*) dataset that provides extensive information on for national accounts aggregates, including value added and capital, we specified a standard Cobb-Douglas production function:

$$\ln prod_{ft} = \mu_f + \alpha \ln K_{ft} + \beta \ln L_{ft} + \delta_t + \omega_{ft} \quad (9d)$$

where $prod$ denotes value added, K stands for capital, and L measures labor input.

We believe that our approach represents a useful generalization of Mendes, van den Berg, and Lindeboom (2010). Our measure of worker productivity, estimated from a three fixed effects wage equation (controlling in particular for the heterogeneity of the firm’s wage policies and the skill composition of its labor force) is better suited and more precise than the measure of workforce quality employed by these authors (viz. the proportion of hours worked by high-skilled workers in a firm as a share of total hours worked in that firm). Our measure of worker productivity is then correlated with alternative measures of firm-specific productivity that can also be estimated with great precision with our data. In the case of equations (9a) and (9b), one can think of our firm (productivity) fixed-effect as a good proxy for the firm total factor productivity; one that takes into account the possible use of thousands of different labor inputs.⁸

(Table 8 near here)

Our results, summarized in Table 8, indicate that the correlation between the worker fixed effects ($\bar{\phi}_f$) estimated from equation (1) and the productivity-firm

fixed effects (μ_f) estimated by (9a) and (9b) is 0.406 and 0.333, respectively. These findings provide strong evidence as to the existence of positive assortative matching in the Portuguese labor market, with firms and workers of similar productivities tending to form matches. More standard measures, estimated under conventional functional forms, provide identical results. The correlation between the firm-specific productivity estimate obtained from equation (9c) and the firm-average worker fixed effect is 0.366. And when we correlate our measure of total factor productivity obtained from equation (9d) with the (firm-averaged) worker fixed effect we found an even higher value – 0.511. We interpret these results as consistent and clear evidence in favor of the super-modular or positive assortative matching hypothesis. Our results are also broadly in line with those from Mendes, van den Berg, and Lindeboom for Portugal, despite the very different methodologies applied.

VII. Conclusion

In this exercise we have used a large longitudinal matched employer-employee dataset to estimate a wage equation with worker, firm, and job title fixed effects. We developed an econometric technique that provides an exact solution to the least squares estimation problem arising when estimating simultaneously high-dimension worker, firm, and job title fixed effects. We decomposed the (natural log of) real hourly wages into several components: observed worker and firm time-varying characteristics, worker heterogeneity (to include observed non-time-varying characteristics and unobserved characteristics), firm heterogeneity (again both observed and unobserved), job title heterogeneity (*idem*), and a residual component.

We have reported that worker heterogeneity is the most important source of wage variation in Portugal (contributing 36.0 percent). The unobserved component plays a more important role (21.0 percent) than the observed non-time-varying characteristics of workers such as gender and education (15.0 percent). Firm effects were found to be also important (contributing 28.7 percent), and are due in roughly equal parts to the unobserved component (14.6 percent) and to observed non-time-varying characteristics such region, capital ownership, and industry (14.0 percent). Job title effects are less important than worker or firm effects, but still explain 9.7 percent of wage variation. The real importance of job title effects in this treatment is they are largely observed, stemming from real world occupational diversity

(compensating differentials, complexity of task implying differential training needs, and so on) and collective agreement impact, and serve to narrow the effect of unobserved worker heterogeneity even if leaving the primacy of the latter unchanged. Their observed component even seems to detract from productivity. The role of job title heterogeneity may be more important in the analysis of wage dynamics. Failure to account for such heterogeneity has been shown to overstate the cyclicity of wages for incumbent or existing workers and to introduce a counter-cyclical bias in wage cyclicity (Carneiro, Guimarães, and Portugal, 2012).

We have also reported that high-wage workers tend to be matched to firms paying higher wages ('high-wage' firms). The evidence that the connection between firms' compensation policies and the quality of their workforces is positive is in marked contrast with most previous evidence. We believe that our result is largely driven by the fact that we are avoiding sampling mobility bias. This is because we observe the whole population of wage earners in the private sector for a period spanning 21 years. Furthermore, the strong correlation between the worker (wage) fixed effect and the firm (wage) fixed effect remains after the inclusion of a rather detailed control for job title fixed effects.

It has been argued that the generosity of the wage policy of the firms, as measured by the firm (wage) fixed effect, can not be taken as evidence that workers are more productive. If that is the case then the correlation between the worker (wage) fixed effect and the firm (wage) fixed effect is non-informative regarding the direction and the strength of assortative matching.

For this reason, we estimated firm-specific measures of productivity, carefully controlling for the heterogeneous composition of the workforce via the inclusion of close to 100,000 job title fixed effects or the job title shares in the production function. The firm (productivity) fixed effects extracted from those production functions exhibit a positive and large correlation with the worker (wage) fixed effect. We interpret this outcome as unambiguously strong evidence of supermodelarity or positive assortative matching. That is, higher productivity workers tend to match with higher productivity firms.

Endnotes

1. Witness implicit contract theory, principal-agent models, and efficiency wage theories.
2. The earliest rent-sharing studies used industry data (e.g. Dickens and Katz, 1987). Firm studies constituted the next phase (e.g. Hildreth and Oswald, 1997; Arai, 2003). The most recent treatments have used matched employer-employee data to control for unobserved worker abilities (e.g. Guertzgen, 2009; Card Devicienti, and Maida, 2010).
3. For treatments combining both approaches – equilibrium job search and matching – see Quercioli (1999); Robin and Roux (1998); Mortensen (2000). Recent extensions include Rosholm and Svarer (2004); Cahuc, Postal-Vinay, and Robin (2006).
4. See also Haltiwanger, Lane, and Spletzer (1999), Andrews, Gill, Schank, and Upward (2008), and van den Berg and van Vuuren (2003) for the United States, Germany, and Denmark, respectively.
5. Eeckout and Kircher argue that only the gain that is achieved from sorting workers into the right job can be gleaned from wage data. In this case, identification comes from the fraction of firms a worker is willing to match with as a proxy for the extent of complementarities.
6. Individuals employed outside of mainland Portugal as well as those in agriculture, hunting, forestry and fishing (as well as misclassified industries) were also excluded.
7. We assume that the variables included in \mathbf{Z} capture the structural characteristics of firms. Changes in them over time are either nonexistent or too small to be considered time-varying and warranting their direct incorporation as explanatory variables into equation (1). The same reasoning applies to the education variable for workers in equation (4) and to the occupation and collective agreement arguments in equation (6). Note further that the Portuguese industrial classification system changed in 1995. Because of this change, and given that the regression covers the entire period, we constructed an aggregated common classification comprising 29 different industries.
8. Note that from Mendes, van den Berg, and Lindeboom (2010) estimate a panel regression in which the dependent variable is the log of real sales per hour worked in firm f in year t (as in our case) and where the independent variables are the logs of three time-varying worker quality indicators (three skills categories, measured in terms of their contributions to total hours worked), their interactions, and two and two additional controls (the size of the workforce and an indicator for single-establishment

firms). The specification chosen was a translog approximation for a generalized production function.

References

- Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," Technical Paper 2002-06, Longitudinal Employer-Household Dynamics, Centre for Economic Studies, U.S. Census Bureau.
- Abowd, John M. and Francis Kramarz. 1999. "Econometric Analyses of Linked Employer-Employee Data," *Labour Economics* 6 (1): 53-74.
- Abowd, John M., Francis Kramarz, Paul Lengermann, and Sébastien Pérez-Duarte. 2004. "Are Good Workers Employed by Good Firms? A Test of a Simple Assortative Matching Model for France and the United States," Unpublished Paper, Cornell University.
- Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms," *Econometrica* 67 (2): 251-333.
- Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward. 2008. "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society, Series A*, 171 (3): 673-697.
- Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward. 2012. "High Wage Workers Matched with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias," *Economics Letters* 117 (3): 824-827.
- Arai, Mahmood. 2003. "Wages, Profitability and Capital Intensity: Evidence from Matched Worker-Firm Data," *Journal of Labor Economics* 21 (3): 593-618.
- Becker, Gary. 1973. "A Theory of Marriage: Part 1," *Journal of Political Economy* 81 (4): 813-846.
- Burdett, Kenneth and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment," *International Economic Review* 39 (2): 257-273.
- Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "Wage Bargaining with On-the-job Search: Theory and Evidence," *Econometrica* 74 (2): 323-364.
- Card, David, Francesco Devicienti, and Agata Maida. 2010. "Rent-Sharing, Holdup, and Wages: Evidence from Matched Panel Data," Working Paper 16192, Cambridge, MA, National Bureau of Economic Research.
- 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): 133-52.
- Diamond, Peter A. 1982. "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies* 49 (2): 217-227.
- Dickens, William T. and Lawrence F. Katz. 1987. "Inter-Industry Wage Differences and Industry Characteristics." In Kevin Lang and Jonathan S. Leonard (eds.), *Unemployment and the Structure of the Labor Market*, New York: Basil Blackwell, pp. 48-89.
- Eeckhout, Jan and Philipp Kircher. 2011. "Identifying Sorting – In Theory," *Review of Economic Studies* 78 (3): 872-906.
- Goux, Dominique and Eric Maurin. 1999. "Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data," *Journal of Labor Economics* 17 (3): 492-533.
- Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica* 45 (1): 1-22.
- Gruetter, Max and Raphael Lalive. 2009. "The Importance of Firms in Wage Determination," *Labour Economics* 16 (2): 149-160.
- Guertzgen, Nicole. 2009. "Rent-Sharing and Collective Bargaining Coverage – Evidence from Linked Employer-Employee Data," *Scandinavian Journal of Economics* 111 (2): 323-349.
- Guimarães, Paulo and Pedro Portugal. 2010. "A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects," *Stata Journal* 10 (4): 628-649.
- Haltiwanger, John C., Julia I. Lane, and James R. Spletzer. 1999. "Productivity Differences Across Employers: The Roles of Employer Size, Age, and Human Capital," *American Economic Review, Papers and Proceedings* 89 (2): 94-98.
- Hildreth, Andrew K. G., and Andrew J. Oswald. 1997. "Rent-Sharing and Wages: Evidence from Company and Establishment Panels," *Journal of Labor Economics* 15 (2): 318-337.
- Kremer, Michael and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill," Working Paper 5718, Cambridge, MA: National Bureau of Economic Research.
- Melo, Rafael Lopes de. 2008. "Sorting in the Labor Market: Theory and Measurement," Unpublished Paper, Yale University.

- Mendes, Rute, Gerard van den Berg, and Maarten Lindeboom. 2010. "An Empirical Assessment of Assortative Matching in the Labor Market," *Labour Economics* 17 (6): 919-929.
- Mortensen, Dale T. 2000. "Equilibrium Unemployment with Wage Posting: Burdett-Mortensen meets Pissarides." In H. Bunzel, Bent Christensen, P. Jensen, Nicholas M. Kiefer, and Dale T. Mortensen (eds.), *Panel Data and Structural Labor Market Models*, Amsterdam: Elsevier, pp. 281-292.
- Mortensen, Dale T. and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies* 61 (3): 397-415.
- Postel-Vinay, Fabien and Jean-Marc Robin. 2002. "The Distribution of Earnings in an Equilibrium Search Model with State-Dependent Offers and Counteroffers," *International Economic Review* 43 (4): 989-1016.
- Quercioli, Elena. 1999. *Firm Specific Training in a Search Equilibrium: Theory and Evidence*, University of Essex.
- Robin, Jean-Marc and Sébastien Roux. 1998. "Random or Balanced Matching: An Equilibrium Search Model with Endogenous Capital and Two-Sided Search," Working Paper 98-38, Centre de Recherche en Economie et Statistique.
- Rosholm, Michael and Michael Svarer. 2004. "Endogenous Wage Dispersion in a Search-Matching model," *Labour Economics* 11 (5): 623-645.
- Sattinger, Michael. 1993. "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature* 31 (2): 831-880.
- Shimer, Robert and Lones Smith. 2000. "Assortative Matching and Search," *Econometrica* 68 (2): 343-369.
- Shimer, Robert. 2005. "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," *Journal of Political Economy* 113 (5): 996-1025.
- Spence, Michael A. 1973. "Job Market Signaling," *Quarterly Journal of Economics*, 87 (3): 355-374.
- Van den Berg, Gerard J. and Aico van Vuuren. 2003. "The Effect of Search Frictions on Wages," Working Paper 3979, Centre for Economic Policy Research, London.
- Weeks, Donald L. and Donald R. Williams. 1964. "A Note on the Determination of Connectedness in an N-way Cross Classification." *Technometrics* 6 (3), 319-324.

Woodcock, Simon D. 2010. "Heterogeneity and Learning in Labor Markets," *The B.E. Journal of Economic Analysis and Policy*, Berkeley Electronic Press, 10 (1): pages 85.

Table 1: Fitted wage equation with worker, firm, and job title fixed effects

Variable	Coefficient	t-statistic
Age (years)	0.02058	1 841.53
Age squared	-0.00023	-1 481.81
Seniority (years)	0.00619	542.09
Seniority squared	-0.00017	- 434.29
Size (ln employees)	0.03460	2 267.00
Observations	26,777,404	
R-squared	0.935	

Note:

The remaining controls comprise eighteen year dummies

Table 2: Regression estimates of worker fixed effects on non-time-varying worker characteristics

Variable	Coefficient	t-statistic
Constant	-0.07990	-295.72
Female	-0.15896	-1,737.79
First stage of basic education	0.06777	246.98
Second stage of basic education	0.16812	577.92
Secondary or post-secondary education	0.24255	835.79
First stage of tertiary education	0.48643	1,170.18
Second stage of tertiary education	0.53936	1,614.69
Observations	26,777,404	
R-squared	0.279	

Table 3: Regression estimates of firm fixed effects on non-time-varying firm characteristics

Variable	Coefficient	t-statistic
Constant	-0.25251	-102.55
Centro	-0.00034	-3.11
Lisboa	0.09775	1,028.06
Alentejo	0.02684	138.73
Algarve	0.07141	313.41
Share of domestic capital	-0.00029	-294.70
Share of public capital	0.00047	229.01
Industry 2	-0.29899	-119.36
Industry 3	-0.43048	-174.40
Industry 4	-0.51620	-209.52
Industry 5	-0.48149	-194.82
Industry 6	-0.47182	-190.80
Industry 7	-0.30293	-122.47
Industry 8	0.18498	70.33
Industry 9	-0.23046	-92.96
Industry 10	-0.30871	-123.67
Industry 11	-0.33084	-133.88
Industry 12	-0.39881	-161.59
Industry 13	-0.34349	-138.80
Industry 14	-0.30150	-121.68
Industry 15	-0.33562	-135.61
Industry 16	-0.53059	-213.94
Industry 17	-0.10521	-42.32
Industry 18	-0.47216	-191.58
Industry 19	-0.42236	-171.45
Industry 20	-0.55907	-226.61
Industry 21	-0.30347	-123.21
Industry 22	-0.00352	-1.43
Industry 23	-0.39563	-160.38
Industry 24	-0.33056	-129.68
Industry 25	-0.39283	-158.43
Industry 26	-0.52982	-214.55
Industry 27	-0.41017	-165.84
Industry 28	-0.63512	-244.61
Industry 29	-0.26447	-19.43
Observations	26,662,583	
R-squared	0.369	

Table 4: Means and standard deviations of compensation components, by gender

	Male		Female		Total	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Ln of real hourly wage (1986 prices)	0.37011	0.56522	0.14066	0.51093	0.27808	0.55559
Predicted effects of X variables ^a	0.92071	0.18139	0.91537	0.17864	0.91857	0.18031
Time	0.33896	0.15781	0.35729	0.15054	0.34631	0.15520
Time-varying observable characteristics of workers	0.44068	0.05085	0.43499	0.05228	0.43839	0.05150
Time-varying observable characteristics of firms	0.15357	0.07804	0.14615	0.07470	0.15060	0.07680
Worker fixed effects	0.05574	0.27412	-0.08704	0.24611	-0.00153	0.27239
Worker fixed effects: unobserved component	0.00000	0.24287	0.00000	0.21302	0.00000	0.23136
Worker fixed effects: observed component ^b	0.05574	0.12148	-0.08704	0.13146	-0.00153	0.14376
Firm fixed effects	-0.61614	0.24095	-0.67316	0.22755	-0.63901	0.23732
Firm fixed effects: unobserved component	0.00556	0.19300	-0.00831	0.18203	0.00000	0.18880
Firm fixed effects: observed component ^c	-0.62177	0.14607	-0.66476	0.13776	-0.63900	0.14434
Job title fixed effects	0.01410	0.10529	-0.02134	0.09707	-0.00012	0.10354
Job title fixed effects: unobserved component	0.00237	0.06424	-0.00354	0.06129	0.00000	0.06314
Job title fixed effects: observed component ^d	0.01173	0.08336	-0.01780	0.07640	-0.00012	0.08193
Number of observations	16,036,759		10,740,645		26,777,404	

Notes:

^a The time-varying observable characteristics of workers (firms) are age, age squared, seniority, and seniority squared (firm size). There are also eighteen year dummies.

^b Gender and five education dummies.

^c Capital ownership (shares of domestic and public capital), four region dummies, and twenty-eight industry dummies.

^d Occupation and collective agreement.

Table 5: Correlations between compensation components

	1	2	2.1	2.2	2.3	3	3.1	3.2	4	4.1	4.2	5	5.1	5.2	6	
Ln of real hourly wage (1986 prices)	1	1														
Predicted effects of X variables ^a	2	0.54	1													
Time	2.1	0.22	0.80	1												
Time-varying observable characteristics of workers	2.2	0.31	0.42	0.03	1											
Time-varying observable characteristics of firms	2.3	0.38	0.38	-0.15	0.19	1										
Worker fixed effects	3	0.74	0.16	-0.05	0.14	0.16	1									
Worker fixed effects: unobserved component	3.1	0.51	0.05	-0.15	0.18	0.12	0.85	1								
Worker fixed effects: observed component ^b	3.2	0.58	0.23	0.14	-0.02	0.11	0.53	0.00	1							
Firm fixed effects	4	0.67	0.25	-0.02	0.15	0.38	0.27	0.10	0.35	1						
Firm fixed effects: unobserved component	4.1	0.43	0.12	-0.01	0.06	0.16	0.08	-0.02	0.19	0.79	1					
Firm fixed effects: observed component ^c	4.2	0.54	0.26	-0.03	0.16	0.42	0.33	0.19	0.33	0.61	0.00	1				
Job title fixed effects	5	0.52	0.17	-0.01	0.23	0.08	0.42	0.27	0.38	0.17	0.07	0.19	1			
Job title fixed effects: unobserved component	5.1	0.16	0.04	-0.04	0.14	0.02	0.10	0.08	0.05	0.00	0.00	0.01	0.61	1.00		
Job title fixed effects: observed component ^d	5.2	0.53	0.18	0.03	0.18	0.08	0.46	0.27	0.44	0.22	0.09	0.24	0.79	0.00	1	
Residual	6	0.28	-0.01	0.00	-0.01	0.00	0.00	-0.06	0.11	0.00	0.00	0.00	0.00	-0.05	0.04	1

Notes:

^a The time-varying observable characteristics of workers (firms) are age, age squared, seniority, and seniority squared (firm size).

There are also eighteen year dummies.

^b Gender and five education dummies.

^c Capital ownership (shares of domestic and public capital), four region dummies, and twenty-eight industry dummies.

^d Occupation and collective agreement.

Table 6: Contributions of compensation components to wage variation

		Contributions
Total		100.0%
Predicted effects of X variables ^a	2	17.4%
Time	2.1	6.2%
Time-varying observable characteristics of workers	2.2	2.9%
Time-varying observable characteristics of firms	2.3	5.3%
Worker fixed effects	3	36.0%
Worker fixed effects: unobserved component	3.1	21.0%
Worker fixed effects: observed component ^b	3.2	15.0%
Firm fixed effects	4	28.7%
Firm fixed effects: unobserved component	4.1	14.6%
Firm fixed effects: observed component ^c	4.2	14.0%
Job title fixed effects	5	9.7%
Job title fixed effects: unobserved component	5.1	1.9%
Job title fixed effects: observed component ^d	5.2	7.9%
Residual	6	8.1%

Notes:

^a The time-varying observable characteristics of workers (firms) are age, age squared, seniority, and seniority squared (firm size). There are also eighteen year dummies.

^b Gender and five education dummies.

^c Capital ownership (shares of domestic and public capital), four region dummies, and twenty-eight industry dummies.

^d Occupation and collective agreement.

Table 7: Comparisons between estimation results from a two fixed effects (worker and firm) wage equation and a three fixed effects (worker, firm, and job title) wage equation

	Worker and firm fixed effects	Worker, firm, and job title fixed effects
R-squared main equation	0.914	0.935
R-squared worker fixed effects equation	0.384	0.279
R-squared firm fixed effects equation	0.370	0.369
R-squared job title fixed effects equation	x	0.628
Correlations between Ln of real hourly wage (1986 prices) and:		
Predicted effects of X variables	0.48	0.54
Worker fixed effects	0.76	0.74
Worker fixed effects: unobserved component	0.51	0.51
Worker fixed effects: observed component	0.58	0.58
Firm fixed effects	0.67	0.67
Firm fixed effects: unobserved component	0.43	0.43
Firm fixed effects: observed component	0.54	0.54
Job title fixed effects	x	0.52
Job title fixed effects: unobserved component	x	0.16
Job title fixed effects: observed component	x	0.53
Correlations between worker fixed effects and firm fixed effects		
Total	0.27	0.27
Unobserved component	-0.04	-0.02
Observed component	0.32	0.33
Contributions of compensation components to wage variability		
Predicted effects of X variables	16.0%	17.4%
Worker fixed effects	46.2%	36.0%
Worker fixed effects: unobserved component	24.2%	21.0%
Worker fixed effects: observed component	22.0%	15.0%
Firm fixed effects	29.2%	28.7%
Firm fixed effects: unobserved component	14.8%	14.6%
Firm fixed effects: observed component	14.4%	14.0%
Job title fixed effects	x	9.7%
Job title fixed effects: unobserved component	x	1.9%
Job title fixed effects: observed component	x	7.9%
Residuals	8.6%	8.1%

Table 8: Linear correlation between the worker’s fixed effects and distinct measures of firm productivity

Firm Productivity	Correlation coefficient with the worker fixed effect
Sales per Worker	
firm productivity fixed effect (9a)	0.406
firm productivity fixed effect (9b)	0.333
Total Sales	
firm fixed efect (9c)	0.366
Total factor productivity (9d)	0.511

Note:

All linear correlation coefficients are weighted by the number of workers at the firm.

Appendix

Implementing Estimation of the Parameters of the Wage Equation

Here we discuss the implementation of the algorithm developed by Guimarães and Portugal (2010) to obtain the least squares solution of our wage equation defined in Section III. Rewriting equation (1) in matrix terms we have the following specification:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\varphi} + \mathbf{L}\boldsymbol{\lambda} + \boldsymbol{\varepsilon}. \quad (\text{A.1})$$

where \mathbf{Y} is a $(N^* \times 1)$ vector of real hourly wages, \mathbf{X} is a $(N^* \times k)$ matrix with k observed time-varying characteristics of individuals and firms, \mathbf{D} is a high-dimensional $(N^* \times N)$ design matrix for the worker effects, \mathbf{F} is a $(N^* \times F)$ high-dimensional design matrix for the firm effects, \mathbf{L} is a $(N^* \times J)$ design matrix for the job title effects and $\boldsymbol{\varepsilon}$ is a $(N^* \times 1)$ vector of disturbances.

Our goal is to estimate the k effects of the time-varying characteristics (namely, vector $\boldsymbol{\beta}$), as well as the N worker fixed effects (vector $\boldsymbol{\theta}$), the F firm fixed effects (vector $\boldsymbol{\varphi}$), and the J job title effects (vector $\boldsymbol{\lambda}$).

Identification of all coefficients is not possible and some restrictions need to be imposed on the coefficients. For the model with two high-dimensional fixed effects Abowd, Creedy, and Kramarz (2002) have shown that one needs to impose one restriction on the coefficients for each “mobility group” in the data (a “mobility group” contains all workers and firms that are connected, that is, all workers who ever worked for any of the firms in the group and all the firms at which any of the workers were ever employed). With several mobility groups (and thus several restrictions) the estimated coefficients of the fixed effects are not comparable across groups. If these coefficients are of interest then a simple solution is to work only with the largest mobility group which usually comprises the majority of the observations. With three fixed effects a similar logic applies. Since we want to use the estimates of the fixed effects for posterior analysis, we restrict the data set to connected observations for which we are assured comparability of the estimates of the fixed effects. This is accomplished by using an algorithm proposed by Weeks and Williams (1964) that identifies all connected observations. In practical applications this largest connected group comprises most observations.

Consider now the estimation problem. The least squares estimates of equation (1) are the solution to the following set of normal equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} & \mathbf{X}'\mathbf{F} & \mathbf{X}'\mathbf{L} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} & \mathbf{D}'\mathbf{F} & \mathbf{D}'\mathbf{L} \\ \mathbf{F}'\mathbf{X} & \mathbf{F}'\mathbf{D} & \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{L} \\ \mathbf{L}'\mathbf{X} & \mathbf{L}'\mathbf{D} & \mathbf{L}'\mathbf{F} & \mathbf{L}'\mathbf{L} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta} \\ \varphi \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{D}'\mathbf{Y} \\ \mathbf{F}'\mathbf{Y} \\ \mathbf{L}'\mathbf{Y} \end{bmatrix} \quad (3)$$

The high-dimensionality of the matrices prevents the use of the conventional least-squares formula. However, if we rearrange the above equation as follows:

$$\begin{bmatrix} (\mathbf{X}'\mathbf{X})\boldsymbol{\beta} \\ (\mathbf{D}'\mathbf{D})\boldsymbol{\theta} \\ (\mathbf{F}'\mathbf{F})\varphi \\ (\mathbf{L}'\mathbf{L})\lambda \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{D}\boldsymbol{\theta} - \mathbf{X}'\mathbf{F}\varphi - \mathbf{X}'\mathbf{L}\lambda \\ \mathbf{D}'\mathbf{Y} - \mathbf{D}'\mathbf{X}\boldsymbol{\beta} - \mathbf{D}'\mathbf{F}\varphi - \mathbf{D}'\mathbf{L}\lambda \\ \mathbf{F}'\mathbf{Y} - \mathbf{F}'\mathbf{X}\boldsymbol{\beta} - \mathbf{F}'\mathbf{D}\boldsymbol{\theta} - \mathbf{F}'\mathbf{L}\lambda \\ \mathbf{L}'\mathbf{Y} - \mathbf{L}'\mathbf{X}\boldsymbol{\beta} - \mathbf{L}'\mathbf{D}\boldsymbol{\theta} - \mathbf{L}'\mathbf{F}\varphi \end{bmatrix}$$

then all square matrices on the left-hand side are easily inverted. The Guimarães and Portugal (2010) procedure consists on iteratively reestimating each set of parameters assuming in each step that the parameters on the right-hand side are known (are set to the latest estimates). This procedure is computationally intensive but converges steadily albeit at a slow rate. For more details on ways to accelerate the algorithm and how to obtain the standard errors see Guimarães and Portugal (2010).

Table A.1: Variables used and their definition/construction

Variable	Description	
<i>year</i>	Year of reference (from 1986 to 2006, except 1990 and 2001)	
<i>firm</i>	Firm identification number	
<i>ss</i>	Worker identification number (Social Security code)	
<i>job title</i>	Job title (or contract) agreed between worker and firm: corresponds to <i>categ x ca</i> (see description below)	
Worker characteristics:		
<i>gender</i>	Gender (male and female)	
<i>age</i>	Age in years	
<i>educ</i>	Education level (ISCED)*	No formal education or below ISCED 1
		Primary education or first stage of basic education (ISCED 1)
		Lower secondary education or second stage of basic education (ISCED 2)
		(Upper) secondary education and post-secondary non-tertiary education (ISCED 3 and 4)
		Tertiary level of education 1 (ISCED 5b)
		Tertiary level of education 2 (ISCED 5a and 6)
<i>tenure</i>	Tenure or seniority (number of months since admission)	
<i>occup</i>	Occupation (ISCO)**	
<i>ca</i>	Collective agreement	
<i>categ</i>	Professional category, defined for each collective agreement	
Compensation and hours:		
<i>w1</i>	Base wage (Euros per month)	
<i>w2</i>	Seniority payments (Euros per month)	
<i>w3</i>	Regular benefits (Euros per month)	
<i>w4</i>	Irregular benefits (Euros per month)	
<i>w5</i>	Overtime pay (Euros per month)	
<i>hours1</i>	Number of normal hours per month	
<i>hours2</i>	Number of extra hours per month	
<i>hw</i>	Hourly wage (Euros). Computed as $(w1+w2+w3+w5)/(hours1+hours2)$	
<i>real_hw</i>	Real hourly wage (Euros). Deflator: Consumer Price Index (prices of 1986)	
<i>ln_real_hw</i>	Logarithm of real hourly wage	
Firm characteristics:		
<i>employees</i>	Number of employees in the firm	
<i>ln_employees</i>	Logarithm of the number of employees in the firm	
<i>Inds</i>	Industry affiliation	
<i>inds6</i>	Industry affiliation (6 sectors) –	Mining and quarrying (NACE Rev.1 activities 10 to 14)
		Manufacturing (NACE Rev.1 activities 15 to 37)

	common classification from 1986 to 2006	Electricity, gas, and water supply (NACE Rev.1 activities 40 to 41)
		Construction (NACE Rev.1 activities 45)
		Market services (NACE Rev.1 activities 50 to 74)
		Social services (NACE Rev.1 activities 80 to 99)
<i>inds29</i>	Industry affiliation (29 sectors) – common classification from 1986 to 2006	
<i>region</i>	Firm NUTS II region	Norte
		Centro
		Lisboa
		Alentejo
		Algarve
<i>sales</i>	Firm sales (Euros)	
<i>real_sales</i>	Real firm sales (Euros). Deflator: Consumer Price Index (prices of 1986)	
<i>real_sales_employee</i>	Real firm sales (Euros) per employee	
<i>share_n</i>	Firm percentage of domestic capital (0 – 100)	
<i>share_p</i>	Firm percentage of public capital (0 – 100)	

Notes:

* ISCED: International Standard Classification of Education, 1997.

** ISCO: International Standard Classification of Occupations.