

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

Sónia Torres
Pedro Portugal*
John T. Addison
Paulo Guimarães

July 6, 2014

*Corresponding Author: Pedro Portugal, Banco de Portugal, Av. Almirante Reis 71, 1150-165
Lisbon, Portugal. Tel.: +351 21 312 8410. E-mail: jppdias@bportugal.pt

Abstract

This paper estimates a wage equation with three high-dimensional fixed effects, using a longitudinal matched employer-employee dataset covering virtually all Portuguese private sector wage earners. The variation in log real hourly wages is decomposed into different components related to worker, firm, and job title characteristics and a residual element. It is found that worker permanent heterogeneity is the most important source of wage variation (36.0 percent), followed by firm permanent effects (28.7 percent). Job title fixed effects contribute least, but they still explain around 10 percent of wage variation - roughly the same order of magnitude as has been reported for the education variable in conventional earnings equations. Having established that high-wage workers tend to match with high-paying firms, worker fixed effects from the wage equation are next correlated with firm fixed effects from sales and value-added production equations to provide evidence on the sign and strength of assortative matching. The correlations are positive and large, indicating that higher productivity workers thus tend to match with higher productivity firms.

JEL: J2, J41.

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

1 Introduction

This paper seeks to provide a better understanding of the sources of wage variation and the role of sorting. Its contribution is twofold. First, it disentangles in the explanation of wage variability the effects of demand-side determinants of wages from supply-side determinants in the manner of Abowd, Kramarz, and Margolis (1999) (AMK), while adding job title fixed effects to the mix of worker and firm fixed effects estimated by these authors. Job title effects reflect the distinct set of occupational tasks performed by workers that serve to define occupational boundaries. Second, it addresses the problem of non-monotonicity typically encountered when estimating the correlation between worker- and firm fixed effects (even if filtered from job title heterogeneity) obtained from wage equations because wages do not necessarily increase in firm productivity. It provides an alternative way of addressing the correlations between the contributions of worker and firm heterogeneity. It does so by observing sales of firms - and, for a restricted set of the data, information on firm value added - to estimate production equations identifying firm fixed effects. That is to say, worker (wage) fixed effects are correlated with firm (productivity) fixed effects to identify the sign and strength of assortative matching. The study furthermore draws on a unique matched worker-firm panel that has the advantages of covering virtually all Portuguese workers over a 21-year interval.

To anticipate the findings, it is first reported that worker permanent heterogeneity is the most important source of wage variation (36.0 percent), followed by firm permanent effects (28.7 percent) and by job title effects (9.7 percent). Note that the latter effect is roughly the same order of magnitude as has been reported for the contribution of the education variable in conventional earnings equations. Also the worker fixed effects contribution is reduced materially in the three fixed effects specification, even if it remains true that 'what workers are' is more important than 'what workers do.' The second and indeed major result is that the correlations between the

firm fixed effects derived from the production function and the worker fixed effects from the wage equation are positive in sign and large in magnitude, suggesting that higher productivity workers tend to match with higher productivity firms. Interestingly, the separate correlations between the worker (wage) fixed effect and the firm (wage) fixed effect are also positive - contrary to similar such correlations in the literature - and of much the same order of magnitude as the previous measure.

The plan of the paper is as follows. Section 2 contains a brief discussion of the nature of worker, firm, and job title fixed effects and a more detailed statement of the evolving literature on the complementarity between individual and firm productivity levels (i.e. assortative matching). The general empirical framework necessary to estimate wage equations with worker, firm, and job title fixed effects is next established in Section 3. A short data description and barebones review of Portuguese wage setting is provided in Section 4. Wage variability is decomposed into its various components in Section 5, where the determinants of worker, firm, and job title fixed effects are investigated and correlations between the components of compensation also addressed. Section 6 assesses the relationship between firms' wage policies and the quality of their labor forces under assortative matching using wage and productivity data. Section 7 concludes.

2 Wage Variation and Worker-Firm Complementarities

2.1 The Sources of Wage Variation

An important research theme in labor economics is why similar workers receive different remuneration and why similar firms pay different wages (Diamond, 1982a). 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 oper-

ating 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. The preceding narrative abstracts from issues of unobserved intrinsic ability (Griliches, 1977) and associated signaling considerations (Spence, 1973).

There is no shortage of models seeking additional or alternative explanations for wage variability, but it is now the characteristics of firms rather than those of workers (i.e. worker competence or productivity differences) that assume pole position. Given the plethora of such treatments (e.g. implicit contract theory, principal-agent models, and efficiency wages), we will consider just two such models 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.¹ These models explain why wages depend not only on external labor market conditions but also on those 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

¹The earliest rent-sharing studies using industry data (e.g. Dickens and Katz, 1987) were followed by firm studies (e.g. Hildreth and Oswald, 1997; Arai, 2003). Most treatments use matched employer-employee data to control for unobserved worker abilities (e.g. Guertzgen, 2009; Card et al., 2013).

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 (1982b) and Mortensen (1982), and Pissarides (2000), for example, while the wage is set by the employer, workers and firms bargain over the share of the matching rent *ex post*. Differences in match productivity, then, explain why similar workers (firms) may receive (offer) different wages.²

Our goal in the present exercise is to 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, of course, collective

²For treatments combining equilibrium job search and matching, see Quercioli (1998); Robin and Roux (1998); Mortensen (2000) and Rosholm and Svarer (2004).

bargaining (encompassing wage floors, promotion policies, and negotiated job titles). To properly incorporate these and other 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 employer's wage policy for 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 any 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 occupational heterogeneity to wage formation.

The objective of this (part of the) 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, AKM 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 dimensional 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.

2.2 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 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). 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 (all agents have higher productivity when they match with high-productivity agents), the transferability of utility, and the commitment to 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 AKM's pioneering study. Using matched employer-employee data for 1976-1987 for a 1/25th sample of the French labor force, these authors 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 et al. (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, as noted earlier, 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 (see also Woodcock, 2010), considerable effort has been expended in determining 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 (i.e. good workers gravitating 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. Note that in the present study since we are observing the whole population of wage earners in the Portuguese private sector, over a little more than two decades, concern over sampling mobility bias is attenuated.

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 AKM 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.

His 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 inferen-

tially. 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 et al. (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.³ 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.

More 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 here 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,

³See also Haltiwanger et al. (1999), Andrews et al. (2008), van den Berg and van Vuuren (2003) for the United States, Germany, and Denmark, respectively.

the highest compensation arises from correct matches and this process substitutes for a wage schedule that is increasing everywhere with type of firm. The authors speak of wages for a given worker having "an inverted U-shape around the optimal allocation, which corresponds to the frictionless wage" (Eeckhout and Kircher, 2011, 874). 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.

A recent application of the AKM model to the growth in wage inequality is provided by Card, Heining, and Kline (2013), who investigate German earnings dispersion over 1985-2009. The authors divide their sample period into four overlapping intervals and fit separate linear models to each with additive person and establishment fixed effects. They report that the model provides a good approximation of the wage structure and explains nearly all of the steep rise in wage inequality in that country. Specifically, increasing dispersion is attributed in large part to rising heterogeneity between workers and rising dispersion in the wage premiums of different establishments. Heightened assortativeness in the assignment of workers to establishments also plays a material role, so that those individuals expected to earn more at any job are increasingly located in establishments offering above-average wages to all employees.

Two aspects of this German study are of especial relevance to the AKM study. First, it is argued that a fully saturated model offers only a modest improvement

in fit over the additively separable AKM model. Second, and relatedly, it addresses the well-known point that exogenous mobility could lead to systematic biases in the AKM specification. If such bias is to be avoided, the (composite) error term must be orthogonal to the vector of establishment identifiers. That is, the assignment of workers to establishments must obey a strict exogeneity condition with respect to the error term. The composite error term comprises an idiosyncratic match component (workers receive some share of an idiosyncratic productivity component associated with each job match), a unit root component capturing drift in the portable component of an individual's earning power (as might occur via employer learning or the arrival of outside offers), and a transitory error term. In investigating sorting based on the idiosyncratic match component of wages, Card, Heining, and Kline (2013) first report that wage gains and losses are broadly symmetric for movers between higher- and lower wage establishments, as well as an absence of gains for moving between establishments with similar estimated fixed effects. Neither result would follow were workers to select jobs based on their match components. Second, in the presence of drift, workers who turn out to be more (less) productive should experience rising (falling) wages at their initial employer and also be more likely to move to higher (lower) wage establishments. Here again the German evidence suggests an absence of any systematic trend in wages prior to a move to better or worse jobs. Finally, there is also scant evidence that mobility patterns are related to transitory wage fluctuations. The net result is that dispersion in the worker specific job match component of wages is both small and stable through time and that rising workplace heterogeneity is not a chimera.

Another recent contribution to the problem of the allocation of workers to jobs eschews a fixed effects regression approach while yet accepting (contrary to Eeckhout and Kircher, 2011) that all parameters of the classic model of sorting based on absolute advantage (i.e. workers and firms can be ranked on their productivity) with search frictions can be identified using only matched employer-employee data

on wages and labor market transitions. Hagedorn et al. (2012) develop strategies for ranking worker and firms, as well as an implementation algorithm (not elaborated upon here). To date, their contribution remains theoretical.

The ranking of workers in this very different schema proceeds on the basis of deriving several (equivalent) statistics that are monotonically increasing in worker type, x : the lowest and highest accepted wages in firms, and the adjusted (for unemployment) average wage. For its part, firm ranking proceeds on the derivation of a statistic that is monotonically increasing in firm type, y . The value of a job vacancy is first established to be monotonically increasing in a firm's productivity and, since the surplus of a vacancy is also increasing in y , Nash bargaining implies that the average surplus of workers is also increasing in y . The authors then show that this surplus can be expressed as a function of wages; specifically, the worker's surplus is proportional to the difference between the wage and the reservation wage. The latter statistic is increasing in firm type and is the basis for ranking firms. The link is that, once workers are ranked, similarly ranked workers must have similar reservation wages. Having ranked workers and firms, the rank correlation between workers of type x and firms of type y denotes the direction and strength of sorting.⁴

Taken in the round, the literature critical of the use of worker and firm fixed effects derived from the wage equation cannot be said to contraindicate the use of worker fixed effects derived from the wage equation. Rather, the issue hinges on the derivation of the firm fixed effect given that firm productivity is not increasing in wages. Thus, in Melo (2008, 3) it is stated: "We propose an alternative measure of sorting, still based on the same fixed effects methodology, that captures the degree of sorting in the model remarkably well: the correlation between a worker fixed effect and the average fixed effects of his coworkers." Hence, the worker fixed effects are

⁴The actual value of the match is recouped from a production function obtained by inverting the wage equation. The determinants are actual wages and the two measured outside options, namely the value of a vacancy and the value of unemployment.

retained but the firm-fixed effects are discarded. Further, Melo (2008, 4) observes that non-monotonicities do not affect the ordering of wages across workers, noting that wages (and average worker wages) are always increasing in worker skill, that wages are unambiguously increasing in worker skill (but not firm productivity), and that in simulations worker fixed effects almost perfectly capture the relative ranking of workers. No less important, Eeckhout and Kircher (2011, 879-880) nowhere claim that wages $w(x, y)$ are non-monotone in x (only in y). Rather more positively, they confirm that the wage (unlike profits and output) is increasing in worker type (pp. 879-880) and alone permit identification of worker type (p. 900). Only the analysis of Hagedorn et al. (2012) resists the notion that wages are monotone in worker productivity. Rather, it is the value of the adjusted average wage in their treatment as opposed to the unadjusted average wage that is necessarily increasing in x . And even for these authors the "key problem underlying the fixed effect regression is the assumption that wages are monotone in firm's productivity (fixed effect)" (pp. 2-3). In the absence of empirical evidence, these authors provide little guidance on the limitations of the wage fixed effect measure, while they would presumably not contest the notion that productivity is increasing in the productivity of the firm.

In the present treatment, therefore, in recognition of potential wage non-monotonicities in firm type noted in the evolving literature on assortative matching, we shall refine our treatment of assortative matching to include firm level productivity filtered from the heterogeneity of labor inputs. That is, we retain the worker wage fixed effect but replace the standard firm effects with firm productivity estimates from a Cobb-Douglas production function where labor input is differentiated by job title. We shall also offer a number of robustness checks. We assume in line with developments in the literature that endogenous mobility bias is not a cause for concern, while contending that limited mobility bias is attenuated by our sampling procedure (see also sub-section 6.1).

3 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:

$$\ln w_{ifjt} = \mathbf{X}_{ifjt}\beta + \theta_i + \phi_f + \lambda_j + \varepsilon_{ifjt} . \quad (1)$$

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 \mathbf{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 ($\mathbf{X}_{ifjt}\beta$);
2. non-time-varying worker heterogeneity (θ_i);
3. non-time-varying firm heterogeneity (ϕ_f);
4. non-time-varying job title heterogeneity (λ_j); and,
5. unexplained random variation (ε_{ifjt}).

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 et al. (1999). In their seminal papers, 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 et al. (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 A we provide a more detailed description of this approach discussing its application to the estimation of equation (1).

4 Data and Institutional Context

4.1 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.⁵ We also dropped close to 2 percent of observations that did not belong to the largest connected set (see the discussion in Appendix A). 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).

4.2 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

⁵Individuals employed outside of mainland Portugal as well as those in agriculture, hunting, forestry and fishing (as well as misclassified industries) were also excluded.

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 maneuver 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 agreements 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.

5 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,

$$\widehat{\theta}_i = \text{const.} + \mathbf{W}_i\eta + \varepsilon_i \quad (2)$$

where \mathbf{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 $\mathbf{W}_i\eta$ 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 $\widehat{\alpha}_i = \widehat{\theta}_i - \mathbf{W}_i\widehat{\eta}$; second,

$$\widehat{\phi}_f = \text{const.} + \mathbf{Z}_f\gamma + \varepsilon_f \quad (3)$$

where \mathbf{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 $\mathbf{Z}_f\gamma$ is the firm non-time-varying observed characteristics component.⁶ As before, the firm-specific intercept, φ_f , capturing the firm unobserved characteristics effect, is obtained residually, by $\widehat{\varphi}_f = \widehat{\phi}_f - \mathbf{Z}_f\widehat{\gamma}$; and third,

$$\widehat{\lambda}_j = FE_{occup} + FE_{ca} + \varepsilon_j \quad (4)$$

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 $\widehat{\delta}_j = \widehat{\lambda}_j - \widehat{FE}_j$.

⁶We 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 (2) and to the occupation and collective agreement arguments in equation (4). 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.

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

- $\mathbf{X}_{ijt}\widehat{\beta}$: observed worker, firm, and economy time-varying characteristics that comprise three components: time dummies, time-varying characteristics of workers, and time-varying characteristics of firms.
- $\widehat{\theta}_i$: worker fixed effects.
 - $\mathbf{W}_i\widehat{\eta}$: observed worker non-time-varying characteristics.
 - $\widehat{\alpha}_i$: unobserved constant worker characteristics.
- $\widehat{\phi}_f$: firm fixed effects.
 - $\mathbf{Z}_f\widehat{\gamma}$: observed firm non-time-varying characteristics.
 - $\widehat{\varphi}_f$: unobserved constant firm characteristics.
- $\widehat{\lambda}_j$: job title fixed effects.
 - \widehat{FE}_j : observed job title non-time-varying characteristics.
 - $\widehat{\delta}_j$: 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 effects for females are on average 15.9 log points smaller than those 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 53.9 log points 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 effects) was estimated through a regression that controlled simultaneously for the time-varying characteristics of workers and firms and for firm

and job-title 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 the southern NUTS II (Nomenclature of Territorial Units for Statistics) regions (*Lisboa*, *Alentejo*, and *Algarve*) than in the north region (the reference category); the firm fixed effects tend to be higher among firms with larger shares of foreign 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 and job-title 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 (4) has a different specification from equations (2) and (3) above. This is due to the nature of the explanatory variables chosen for equation (4). 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 observed 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 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)

In Table 4, we report the correlations among the components of log 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 effects are highly correlated with the log of real total compensation (0.58 and 0.51, respectively). Concerning the components of the firm fixed effects, 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), we note that Abowd et al. (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 et al. (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 4 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. As discussed earlier, we should resist the temptation of interpreting this positive correlation 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 decomposition:

$$\ln w_{ifjt} = \mathbf{X}_{ifjt}\beta + \alpha_i + \mathbf{W}_i\eta + \varphi_f + \mathbf{Z}_f\gamma + \delta_j + FE_j + \varepsilon_{ifjt} = \sum_{p=1}^{10} C_{ifjt}^p \quad (5)$$

where the C_{ifjt}^p represent the individual summands (\mathbf{X}_{ifjt} comprises the three components described above: time; time-varying observed characteristics and time-varying observed firm characteristics) of the wage equation. The contribution of each component, C_{ifjt}^p , can be calculated as:

$$\frac{cov(\ln w_{ifjt}, C_{ifjt}^p)}{Var(\ln w_{ifjt})} \quad (6)$$

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

(Table 5 near here)

In Table 5, we report the contribution of each component to the real hourly wages variability. 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 fixed 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).

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

6.1 The Correlation between Firm Productivity and Worker Fixed Effects

As was noted in section 2.2, problems of limited mobility bias and the non-monotonicity of wages in firm productivity frustrate the attempt to assess the degree of assortative matching in the labor market from correlations between worker and firm fixed effects estimated from wage equations. It was argued that our results are unlikely to be materially affected by the former 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. This still leaves the second bias that we next tackle 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 Kircher, 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.

To test the hypothesis of assortative matching we rely on the estimation of a firm-specific measure of productivity that is filtered from the heterogeneity of job titles. Our approach can be viewed as a way of attenuating the attribution problem flagged by Eeckhout and Kircher (2011, 900). The estimating equation is specified as:

$$\ln Q_{ft} = \mu_f + \sum_{j=1}^{33491} \eta_j \ln jobs_{fjt} + \delta_t + \varepsilon_{ft} , \quad (7a)$$

where Q_{ft} denotes real sales ⁷ by firm f in year t , μ_f identifies the firm fixed effect, $jobs_{fjt}$ gives the number of workers filling job-title j at firm f in year t , η_j are output elasticities with respect to job-title j , δ_t represents year fixed effects, and ε_{ft} is an idiosyncratic random error term.

The main reason we specify this kind of (Cobb-Douglas type) production function is, as in Douglas (1976), to infer the size of the labor contribution to output. In our dataset we have to deal with firms with just one job title (24,677 firm/years) and with firms that employ up to 592 distinct job titles.

This multifactor production function requires the estimation of a high-dimensional linear regression function with 33,491 covariates, 18 year dummy coefficients, and the critical firm fixed effects. This is infeasible with standard econometric software. To proceed, therefore, we extended the iterative procedure of Guimarães and Portugal

⁷The use of value added provides a better measure of total factor productivity because it avoids the contamination of this estimate by the different uses of intermediate consumption.

(2010) in such a way that, by portioning the total set of covariates into smaller, manageable subsets, we were able to employ conventional software (e.g. Stata).⁸ Nevertheless, the procedure was slow and demanded large storage capabilities; specifically, in this exercise, the data storage required around 100 Gigabytes while estimation took almost six weeks to converge.

Two main restrictions were placed on the data so as to reduce the computational burden of the estimation. First, we focused on manufacturing alone, which sector better fits the notion of a production function. Second, we excised all years after 2004 because the coding of job-titles changed in 2005. With these restrictions, we downsized the number of job titles from 95,927 to a more manageable 33,491.

The bulk of this exercise is summarized in just one correlation coefficient. The number that we care about is the (employment-weighted) correlation coefficient between the firm averaged worker fixed effects extracted from the main wage equation (equation 1) and the firm fixed effects obtained from the production function. This value is 0.334 (see the first row entry in Table 6). We interpret this result as strong evidence that more productive workers tend to match with more productive firms. In short, we have found strong empirical support for the notion of positive assortative matching.

(Table 6 near here)

6.2 Robustness Checks

By way of robustness checks, we first estimated a production equation, where the dependent variable is the natural log of real sales per worker at firm f in year t ($\ln Q/L_{ft}$).

Specifically, using worker-level information and a two-way high-dimensional fixed

⁸The approach is explained in Appendix B.

effects procedure, we estimated the following equation:

$$\ln Q/L_{ifjt} = \mu_f + \eta_j + \delta_t + \varepsilon_{ifjt} \quad (7b)$$

where μ_f denotes the firm (productivity) fixed effect, η_j denotes the job title (productivity) fixed effects, δ_t denotes the year fixed effects, and ε_{ifjt} 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 allow for the estimation of job title specific regression coefficients, to estimate the following regression model:

$$\ln Q/L_{fjt} = \mu_f + \gamma_j share_{fjt} + \delta_t + \varepsilon_{fjt} \quad (7c)$$

where $share_{fjt}$ represents the fraction of workers with job title j at firm f at year t ($\sum_j share_{fjt} = 1$) and the γ_j are the respective coefficients. 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.

Two further restrictions were also imposed prior to estimating equations (7b) through (7c); 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 above equations were all estimated based on a dataset that comprised 25,518,858 worker/year pairs.

These first two robustness checks use all the data (all the job titles) but differ in terms of functional form. Our next specification considers a production function in which (the log of) sales is the dependent variable and where up to 2,459 distinct worker occupations are included as labor inputs in the following regression model:

$$\ln Q_{ft} = \mu_f + \sum_{k=1}^{2459} \xi_k \ln Occ_{fk} + \delta_t + \varepsilon_{ft} , \quad (7d)$$

where Occ_{fk} denotes the number of workers in firm f that fill occupation k .

Our final alternative specification uses firm-level data on value added, avoiding issues of intermediate inputs, and accounting for capital. We were able to merge a subset of firms (444,563) with the IES (*Inquérito às Empresas Simplificado*) dataset that provides extensive information on national accounts aggregates, including value added and capital. This allowed us to estimate a standard Cobb-Douglas production function of type:

$$\ln Q_{ft} = \mu_f + \alpha \ln K_{ft} + \beta \ln L_{ft} + \delta_t + \varepsilon_{ft} \quad (7e)$$

where Q now denotes value added, K stands for capital, and L measures labor input.

These robustness checks can be seen as a useful generalization of Mendes et al. (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 (7b) and (7c), one can think of our firm (productivity) fixed effects as a good proxy for the firm total factor productivity; one that takes into account the possible use of thousands of different labor inputs.⁹

The results for each measure of firm productivity appear below our preferred estimate in Table 6. As can be seen, the correlation between the worker fixed effects ($\widehat{\theta}_i$) estimated from equation (1) and the firm-averaged (productivity) fixed effects ($\widehat{\mu}_f$) estimated by (7b) and (7c) is 0.406 and 0.333, respectively. These findings

⁹Note that Mendes et al. (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 skill categories, measured in terms of their contributions to total hours worked), their interactions, 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.

accord fairly closely with the first row estimate and are supportive of the existence of positive assortative matching in the Portuguese labor market. More standard measures, estimated under conventional functional forms, provide identical results. Thus, the correlation between the firm-specific productivity estimate obtained from equation (7d) and the firm-averaged worker fixed effects is 0.345. And when we correlate our measure of total factor productivity obtained from equation (7e) with the firm-averaged worker fixed effects we obtain an even higher value of 0.511. These robustness checks offer close support for our preferred specification which we interpret as providing clear evidence in favor of the super-modularity or positive assortative matching hypothesis. Our secondary results are of course also broadly in line with those from Mendes et al. for Portugal, despite the very different methodologies applied.

7 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 four 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 to wage variation). 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 also found to be important (contributing 28.7 percent), due in roughly equal parts to the unobserved component (14.6 percent) and to observed non-time-varying characteristics such regional location, capital ownership, and industry (14.0 percent). Job title effects are less important than either of the worker or firm effects, but they still explain a nontrivial 9.7 percent of wage variation, much the same as the education argument in the standard Mincerian earnings function. The importance of job title effects in this treatment is that they are largely observed, having a basis in real world occupational diversity (implying compensating differentials and differential training needs stemming from complexity of tasks) and collective agreement impact. Note that job title effects serve to narrow the effect of unobserved worker heterogeneity, even if leaving its overall primacy unchallenged. We have also reported that high-wage workers tend to be matched to firms paying higher wages ('high-wage' firms). The finding that the connection between firms' compensation policies and the quality of their workforces is positive is in marked contrast with much of the previous evidence. We believe that this result is driven by the fact that we are largely avoiding sampling mobility bias because of our use of the entire population of Portuguese wage earners in the private sector. 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 heterogeneity. However, the generosity of the wage policy of firms, as indexed by the firm (wage) fixed effects, can not be taken as evidence that they are more productive. As has been noted in the literature, the correlation between the worker (wage) fixed effects and the firm (wage) fixed effects is not 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 job title shares, in the production function. The firm (productivity) fixed effects extracted from sales and value-added production functions exhibited a positive and large correlation with the

worker (wage) fixed effects. We interpret this latter finding as supportive of supermodularity, or positive assortative matching. In sum, higher productivity workers tend match with higher productivity firms.

Finally, there is the issue of external validity. In particular, might not Portuguese wage determination and the Portuguese wage distribution differ markedly from other nations? For its part, the sharp divergence between union density and union coverage is not confined to Portugal, whose collective bargaining system is not dissimilar from those in Spain, Italy, and France, while extension agreements (and informal ‘orientation processes’) are indeed common in many other continental European nations, including Germany. Without denying the need for more work on the wage distributions of different nations, we consider it unlikely that institutional idiosyncrasies lie at the heart of the findings on worker and firm heterogeneity and assortative matching reported here for Portugal.

Appendix A: 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 3. Rewriting equation (1) in matrix terms we have the following specification:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{D}\theta + \mathbf{F}\phi + \mathbf{L}\lambda + \epsilon \quad (7)$$

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 ϵ is a $(N^* \times 1)$ vector of disturbances.

Our goal is to estimate the k effects of the time-varying characteristics (namely, vector β), as well as the N worker fixed effects (vector θ), the F firm fixed effects (vector ϕ), and the J job title effects (vector λ).

Identification of all coefficients associated with the fixed effects is not possible and some restrictions need to be imposed. For the model with two high-dimensional fixed effects Abowd et al. (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:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} & \mathbf{X}'\mathbf{F} & \mathbf{X}'\mathbf{L} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} & \mathbf{D}'\mathbf{F} & \mathbf{D}'\mathbf{L} \\ \mathbf{F}'\mathbf{X} & \mathbf{F}'\mathbf{D} & \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{L} \\ \mathbf{L}'\mathbf{X} & \mathbf{L}'\mathbf{D} & \mathbf{L}'\mathbf{F} & \mathbf{L}'\mathbf{L} \end{bmatrix} \begin{bmatrix} \beta \\ \theta \\ \phi \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{D}'\mathbf{Y} \\ \mathbf{F}'\mathbf{Y} \\ \mathbf{L}'\mathbf{Y} \end{bmatrix}$$

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

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

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).

Appendix B: Estimation of the High-Dimensional Cobb-Douglas function

The Cobb-Douglas functions that were estimated in section 6 require the implementation of a high-dimensional linear regression with thousands of covariates and one fixed effect. With conventional software it is impossible to estimate these regressions.

However these models can be estimated using a variant of the algorithm presented in Guimarães and Portugal (2010). To illustrate, consider the general linear regression model given by:

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \dots + \mathbf{X}_k\beta_k + \varepsilon$$

where for simplicity we are partitioning the total set of covariates into smaller subsets that have a dimension with is amenable for estimation with conventional software. Now, to estimate this model we can employ a strategy which is similar to the one employed for estimation of our wage equation. Rewriting the normal equations as:

$$\begin{bmatrix} (\mathbf{X}'_1\mathbf{X}_1)\beta_1 \\ (\mathbf{X}'_2\mathbf{X}_2)\beta_2 \\ \dots \\ (\mathbf{X}'_k\mathbf{X}_k)\beta_k \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1(\mathbf{Y} - \mathbf{X}_2\beta_2 - \mathbf{X}_3\beta_3 - \dots - \mathbf{X}_k\beta_k) \\ \mathbf{X}'_2(\mathbf{Y} - \mathbf{X}_1\beta_1 - \mathbf{X}_3\beta_3 - \dots - \mathbf{X}_k\beta_k) \\ \dots \\ \mathbf{X}'_k(\mathbf{Y} - \mathbf{X}_1\beta_1 - \mathbf{X}_2\beta_2 - \dots - \mathbf{X}_{k-1}\beta_{k-1}) \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_1\mathbf{Y}_1^* \\ \mathbf{X}'_2\mathbf{Y}_2^* \\ \dots \\ \mathbf{X}'_k\mathbf{Y}_k^* \end{bmatrix}$$

we can iterate across the equations taking as known the β s that show up on the right-hand side (replacing them by the last known estimates) and estimating the β s on the left-hand side by implementing the linear regression implied by the above equations. Again, this is a process that will converge slowly but steadily. A fixed effect can be easily added. One option is to assume that one of the subsets consists of all the dummy variables that define the fixed effect. In this case the estimation step associated with the fixed effect consists on the calculation of simple group means.

References

- Abowd, J. M. and F. Kramarz. Econometric analyses of linked employer-employee data. *Labour Economics*, 6(1):53–74, 1999.
- Abowd, J. M., F. Kramarz, and D. N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–334, 1999.
- Abowd, J. M., R. H. Creecy, and F. Kramarz. Computing person and firm effects using linked longitudinal employer-employee data. Technical Report 2002-06, LEHD, U.S. Census Bureau, 2002.
- Abowd, J. M., F. Kramarz, P. Lengermann, and S. Pérez-Duarte. Are good workers employed by good firms? a test of a simple assortative matching model for France and the United States. Discussion paper, Cornell University, 2004.
- Andrews, M. J., L. Gill, T. Schank, and R. Upward. 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, 2008.
- Andrews, M. J., L. Gill, T. Schank, and R. Upward. High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, 117(3):824–827, 2012.
- Arai, M. Wages, profits, and capital intensity: Evidence from matched worker-firm data. *Journal of Labor Economics*, 21(3):593–618, 2003.
- Becker, G. S. A theory of marriage: Part I. *Journal of Political Economy*, 81(4): 813–46, 1973.
- Burdett, K. and D. T. Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–73, 1998.

- Card, D., J. Heining, and P. Kline. Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967–1015, 2013.
- Cardoso, A. R. and P. Portugal. Contractual wages and the wage cushion under different bargaining settings. *Journal of Labor Economics*, 23(4):875–902, 2005.
- Diamond, P. Wage determination and efficiency in search equilibrium. *Review of Economic Studies*, 49(2):217–227, 1982a.
- Diamond, P. Aggregate demand management in search equilibrium. *Journal of Political Economy*, 90(5):881–894, 1982b.
- Dickens, W. T. and L. F. Katz. Interindustry wage differences and industry characteristics. In Lang, K. and J. S. Leonard, editors, *Unemployment and the Structure of the Labor Market*, pages 48–89. Basil Blackwell, New York, 1987.
- Douglas, P. H. The Cobb-Douglas production function once again: Its history, its testing, and some new empirical values. *Journal of Political Economy*, 84(5):903–15, 1976.
- Eeckhout, J. and P. Kircher. Identifying sorting – in theory. *Review of Economic Studies*, 78(3):872–906, 2011.
- Goux, D. and E. Maurin. Persistence of interindustry wage differentials: A reexamination using matched worker-firm panel data. *Journal of Labor Economics*, 17(3):492–533, 1999.
- Griliches, Z. Estimating the returns to schooling: Some econometric problems. *Econometrica*, 45(1):1–22, 1977.
- Gruetter, M. and R. Lalive. The importance of firms in wage determination. *Labour Economics*, 16(2):149–160, 2009.

- Guertzen, N. Rent-sharing and collective bargaining coverage: Evidence from linked employer-employee data. *Scandinavian Journal of Economics*, 111(2):323–349, 2009.
- Guimarães, P. and P. Portugal. A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, 10(4):628–649, 2010.
- Hagedorn, M., T. H. Law, and I. Manovskii. Identifying equilibrium models of labor market sorting. NBER Working Paper 18661, National Bureau of Economic Research, Inc, December 2012. URL <http://www.nber.org/papers/w18661.pdf>.
- Haltiwanger, J. C., J. I. Lane, and J. Spletzer. Productivity differences across employers: The roles of employer size, age, and human capital. *American Economic Review, Papers and Proceedings*, 89(2):94–98, 1999.
- Hildreth, A. K. and A. J. Oswald. Rent-sharing and wages: Evidence from company and establishment panels. *Journal of Labor Economics*, 15(2):318–337, 1997.
- Kremer, M. and E. Maskin. Wage inequality and segregation by skill. NBER Working Paper 5718, National Bureau of Economic Research, Inc, August 1996. URL <http://ideas.repec.org/p/nbr/nberwo/5718.html>.
- Melo, R. L. d. Sorting in the labor market: Theory and measurement. Unpublished paper, Yale University, 2008.
- Mendes, R., G. J. van den Berg, and M. Lindeboom. An empirical assessment of assortative matching in the labor market. *Labour Economics*, 17(6):919–929, 2010.
- Mortensen, D. T. Property rights and efficiency in mating, racing, and related games. *American Economic Review*, 72(5):968–969, 1982.
- Mortensen, D. T. Equilibrium unemployment with wage posting: Burdett-Mortensen meets Pissarides. In Bunzel, H., B. Christensen, P. Jensen, N. M. Kiefer, and D. T.

- Mortensen, editors, *Panel Data and Structural Labor Market Models*, pages 281–292. Elsevier Press, Amsterdam, 2000.
- Pissarides, C. A. *Equilibrium Unemployment Theory*. Cambridge, MA: The MIT Press, 2000.
- Quercioli, E. *Firm Specific Training in a Search Equilibrium: Theory and Evidence*. University of Essex, 1998.
- Robin, J.-M. and S. Roux. Random or balanced matching : An equilibrium search model with endogenous capital and two-sided search. Working Papers 98-38, Centre de Recherche en Economie et Statistique, 1998. URL <http://ideas.repec.org/p/crs/wpaper/98-38.html>.
- Rosholm, M. and M. Svarer. Endogenous wage dispersion in a search-matching model. *Labour Economics*, 11(5):623–645, 2004.
- Sattinger, M. Assignment models of the distribution of earnings. *Journal of Economic Literature*, 31(2):831–80, 1993.
- Shimer, R. The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025, 2005.
- Shimer, R. and L. Smith. Assortative matching and search. *Econometrica*, 68(2):343–370, 2000.
- Spence, M. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–74, 1973.
- van den Berg, G. J. and A. van Vuuren. The effect of search frictions on wages. CEPR Discussion Papers 3979, C.E.P.R. Discussion Papers, July 2003. URL <http://ideas.repec.org/p/cpr/ceprdp/3979.html>.

Weeks, D. L. and D. R. Williams. A note on the determination of connectedness in an N-way cross classification. *Technometrics*, 6(3):319–324, 1964.

Woodcock, S. D. Heterogeneity and learning in labor markets. *The B.E. Journal of Economic Analysis & Policy*, 10(1):85, 2010.

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

Variable	Coefficient	t-statistic
Age (years)	0.02058	817.85
Age squared	-0.00023	-704.48
Seniority (years)	0.00619	477.65
Seniority squared	-0.00017	-399.93
Size (ln employees)	0.03460	2,068.81
Observations		26,777,404
R-squared		0.935

Notes: The remaining controls comprise eighteen year dummies. t-statistics are calculated with worker-cluster robust standard errors.

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

Variable	Coefficient	t-statistic
Constant	-0.07990	-197.40
Female	-0.15896	-640.82
First stage of basic education	0.06777	167.00
Second stage of basic education	0.16812	340.07
Secondary or post-secondary education	0.24255	467.93
First stage of tertiary education	0.48643	416.25
Second stage of tertiary education	0.53936	598.44
Observations		26, 777, 404
R-squared		0.279

Notes: t-statistics are calculated with worker-cluster robust standard errors.

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

Variable	Coefficient	t-statistic
Constant	-0.25251	-109.49
Centro	-0.00034	-1.41
Lisboa	0.09775	406.22
Alentejo	0.02684	56.83
Algarve	0.07141	127.46
Share of domestic capital	-0.00029	-112.84
Share of public capital	0.00047	92.84
Observations		26,662,583
R-square		0.369

Notes: Coefficients for industry dummy variables are not reported. t-statistics are calculated with worker-cluster robust standard errors.

Table 4: 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														
Predicted effects of X variables ^a	0.54	1													
Time	0.22	0.80	1												
Time-varying obs. char. of workers	0.31	0.42	0.03	1											
Time-varying obs. char. of firms	0.38	0.38	-0.15	0.19	1										
Worker fixed effects	0.74	0.16	-0.05	0.14	0.16	1									
Worker fixed effects: unobs. component	0.51	0.05	-0.15	0.18	0.12	0.85	1								
Worker fixed effects: obs. component ^b	0.58	0.23	0.14	-0.02	0.11	0.53	0.00	1							
Firm fixed effects	0.67	0.25	-0.02	0.15	0.38	0.27	0.10	0.35	1						
Firm fixed effects: unobs. component	0.43	0.12	-0.01	0.06	0.16	0.08	-0.02	0.19	0.79	1					
Firm fixed effects: obs. component ^c	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	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: unobs. component	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		
Job title fixed effects: obs. component ^d	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	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).^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: Contributions of compensation components to wage variation

	Contributions	
Total	1	100.0%
Predicted effects of X variables ^a	2	17.4%
Time	2.1	6.2%
Time-varying obs. char. of workers	2.2	2.9%
Time-varying obs. char. of firms	2.3	5.3%
Worker fixed effects	3	36.0%
Worker fixed effects: unobs. component	3.1	21.0%
Worker fixed effects: obs. component ^b	3.2	15.0%
Firm fixed effects	4	28.7%
Firm fixed effects: unobs. component	4.1	14.6%
Firm fixed effects: obs. component ^c	4.2	14.0%
Job title fixed effects	5	9.7%
Job title fixed effects: unobs. component	5.1	1.9%
Job title fixed effects: obs. component ^d	5.2	7.9%
Residual	6	8.1%

Notes: Same as for Table IV.

Table 6: Linear correlation between the workers fixed effects and distinct measures of firm productivity

Production Measure	Estimating Equation	Correlation Coefficient
(1) Total sales ^a	7a	0.334
(2) Sales per worker ^b	7b	0.406
(3) Sales per worker ^c	7c	0.333
(4) Total sales ^d	7d	0.345
(5) Value added ^e	7e	0.511

Notes: ^a Production function includes distinct labor inputs based on job titles. Period restricted to 1986-2003. Firm/year observations: 439,769. ^b Production function includes job title fixed effects. Firm/year observations: 2,551,858. ^c Production function includes shares of job titles. Firm/year observations: 2,551,858. ^d Production function includes distinct labor inputs based on occupations. Period restricted to 1995-2006. Firm/year observations: 1,871,467. ^e Production function includes measures of capital and labor (total number of workers). Sample restricted to manufacturing firms and period restricted to 1991-2006. Firm/year observations: 442,563. All linear correlation coefficients are weighted by the number of workers at the firm. All regressions include year dummies.

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</i> x <i>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*	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***	Mining and quarrying (NACE Rev.1 activities 10 to 14)
		Manufacturing (NACE Rev.1 activities 15 to 37)
		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)	
<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	

Variable	Description
<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.

*** common classification from 1986 to 2006