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

Sónia Torres,* Pedro Portugal,** John T. Addison,*** and Paulo Guimarães****

*Statistics Portugal
**Banco de Portugal and Universidade Nova de Lisboa
***University of South Carolina
****CEFUP, Universidade do Porto, and University of South Carolina

Juner 2012

Abstract
This paper estimates a wage equation with three high-dimensional fixed effects – worker, firm, and job title – 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 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 effects are less important overall (28.7 percent) and are due in roughly equal parts to the unobserved component (14.6 percent) and the observed component (14.0 percent). Job title effects emerge as the least important factor but they still explain 9.7 percent of wage variation, with the observed fixed effects component playing a more important role (7.9 percent versus 1.9 percent for the unobserved fixed component) while improving the precision of the estimates of the other sources of wage variation. Finally, and importantly given the lingering ambiguity in the wider empirical literature, there is material evidence of positive assortative matching.

JEL Classification: J2, J41
Keywords: worker fixed effects, firm fixed effects, job title fixed effects, decomposition of wage variability, high-dimension matrices, wage policies, firm performance, labor force quality

I. Introduction

An important research theme in labor economics is why similar workers receive different remuneration and why similar firms pay different wages. 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 (their 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 and not 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 – abstracting from unobserved intrinsic ability and ignoring related signaling models.

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 assume prime importance. 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 profitable 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 – and why the wages of workers from different groups of occupations, education, and seniority are higher in some firms or industries than in others.

The other explanation for wage differentials among workers with similar characteristics considered 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), Mortensen (1986), Pissarides (1985, 2000), and Mortensen and Pissarides (1994) the wage paid is set by the employer, but 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 (pay) different wages.[3]

The goal 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, we estimated a wage regression in which we introduced simultaneously worker and firm fixed effects.

However, besides worker and firm heterogeneity, a third important dimension of wage formation is job title heterogeneity. Job title heterogeneity may influence wage rates for a variety of reasons. First, it is well known that tasks that involve risks of fatal or otherwise serious accidents are better paid than safe tasks. One should therefore expect significant compensating differentials to attach to occupations such as deep sea divers or bullfighters. Second, jobs that need to be executed under difficult or stressful conditions are also expected to be more highly remunerated than jobs performed in pleasant environments. For example, one should observe higher wages for individuals that work on offshore oil platforms or in mines. Third, the complexity of some tasks may require heavy doses of specific training and/or unusually skilled workers. This is one important reason why, for example, brain surgeons and jet-fighter pilots earn higher wages. Fourth, some occupations are known to be chronically 'overcrowded' whereas others are thought to be in excess demand. For decades, it has been argued that there is an oversupply of teachers and a corresponding undersupply of nurses. Fifth, some jobs, by their nature, put workers in a position where they can inflict serious losses upon their employers and/or society. Unions organizing such workers will be powerful enough to extract significant rents in the form of higher wages. Industrial action by commercial airline pilots, by train drivers, and by flight controllers – or, more generally, by individuals employed by natural monopolies – often leads to substantial wage premia. Sixth, entry barriers to certain occupations that are erected by worker associations (e.g. closed-shop regulations and occupational licensing by associations of medical practitioners and lawyers) will also enhance the labor incomes of their incumbents. Finally, the kind of technology in use may also foster unionization of the workplace and favor rent seeking. Production activities that imply the concentration of a large number of workers in a single plant (say, in autos or ship building) facilitate industrial action, and thus improved conditions.

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 very disaggregated occupational count or listing would not be fit for purpose here because the wage policy regarding the same occupation (e.g. a secretary) may be governed by different collective agreements (e.g. by the banking industry agreement as opposed to the retail trade collective agreement). Fortunately, in our dataset we have access to an unusually rich set of information that enables us to identify the collective agreement that regulates the employment contract applicable to each worker and, within each collective agreement, one can further pinpoint the exact, detailed occupational category of each worker. The reason why this information is collected reflects the specificities of the Portuguese wage setting system (largely conformable to continental European practice). Each year, around 300 different collective agreements are negotiated. The collective agreement defines wage floors for each particular job title (so-called professional category or "categoria profissional"). On average, each collective agreement defines the wage floor for around 100 job titles. Overall, in a given year, one can classify each worker along some 30,000 collective agreement/job title combinations. The main use of the *Quadros de Pessoal* 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).

In this study, we are confident that, by incorporating job title fixed effects in the wage regression, we can make good progress in determining the contribution of job title heterogeneity. And by properly accounting for job title heterogeneity, one should be able to provide refined estimates filtered from the effects job title heterogeneity of worker and firm fixed effects. This should shed additional light on the current debate concerning the role of assortative matching, as measured through the association between worker and firm fixed effects. In the process, we should also be able to unambiguously disentangle the contribution of contract heterogeneity and occupation heterogeneity to wage formation.

The final objective of this estimation with three fixed effects is to calculate the contribution of each determinant of wages to overall wage variability, as described below and further elaborated upon in section VI. 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 – worker, firm, and job title 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 these 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. In particular, 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. We took this methodology a stage further, by including, as mentioned above, also a third fixed effect in our wage equation for the job title and sought explicitly to control for job heterogeneity.

To our knowledge, this exercise is performed for the first time under optimal conditions. To repeat, these are universal coverage of the employed population and the use of adequate econometric tools.

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), not least because the jury is out on the direction of matching. In the interests of completeness, we can also range further afield to consider whether the compensation policies followed by firms as revealed by all three types of heterogeneity are related to their *performance*. The general empirical framework necessary to estimate wage equations with worker, firm, and job title fixed effects is next established in Section III. A data description and barebones review of wage setting in Portugal is contained in Section IV. Wage variability is decomposed into its components in Section V, the determinants of worker, firm, and job title fixed effects investigated, and correlations between the components of compensation addressed. Section VI assesses the relationship between firms' wage policies and their performance as well as labor force quality, using productivity data. Section VII concludes.

II. Assortative matching

Also examined in the present paper is evidence on the sorting of heterogeneous workers across plants, and in particular the notion of positive assortative matching. 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 matching equilibrium 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 flawed method). The authors also report correlations between the two effects for a 1/10th sample of State of Washington Employment 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, Kramarz, Lengermann, and Pérez-Duarte (2004) test and discount the notion that the negative correlation between the fixed worker and employer effects – *vulgo*: good workers gravitate to bad firms – are 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 similar but somewhat more attenuated conclusion is reached by Andrews, Gill, Schank, and Upward (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. Nevertheless, although the biases are large, they do not in this study overturn the negative correlation between the worker and plant effects.

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 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, Mendes, van den Berg, and Lindeboom (2007) estimate a firm-specific productivity effect for each firm using a translog specification which 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.[5] 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.

Using a 10 percent random sample from the *Quadros de Pessoal*, Ferreira (2009) has examined the role of promotions and separations on wages. For this purpose, she deploys two regression models. The first is a conventional wage regression model with worker and firm fixed effects. In the second, in addition to the worker and firm fixed effects, the author includes a match fixed-effect – albeit at the cost of making the rather questionable assumption that this effect is orthogonal to either fixed effect. Ferreira concludes that worker and firm permanent unobserved heterogeneity account for more than one-half of the variation of wages.

Disappointingly, the match effect 'explains' just 2 percent of the variation. Strikingly, in Ferreira's study there is indication that assortative matching, where present, is negative.

Recently, a trenchant criticism of using unobserved worker and firm effects to conclude anything about assortative matching has been made by Eeckhout and Kircher (2010). 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 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. The authors speak of wages for a given worker having "an inverted U-shape around the optimal allocation that 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 (i.e. in the manner of Abowd, Kramarz, and Margolis, 1999) 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.[4] 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 (a proxy for) productivity data so as to address some of these authors' concerns.

III. The General Empirical Framework to Measure Wage Variation

The methodology applied in this paper expands that initially developed by Abowd and Kramarz (1999) and Abowd, Kramarz, and Margolis (1999), who presented a statistical framework permitting worker and firm fixed effects to be estimated simultaneously in wage regressions. However, as noted earlier, and as elaborated upon below, we shall use a different algorithm to obtain an exact solution for the estimation problem, and we also include a third fixed effect for the job title.

The linear wage equation to be estimated has the form:

$$ln\, w_{ifjt} = X_{ift}\beta + \theta_i + \varphi_f + \lambda_j + \varepsilon_{ifjt}\,, \tag{1}$$

which is related, in the statistical literature, with the "three-factor analysis of covariance." In this equation, $ln\, w_{ifjt}$ is the natural logarithm of the real hourly wage of individual *i* ($i = 1,...,N$) working at firm *f* ($f = 1,...,F$) and holding a job title *j* ($j = 1,...,J$) at year *t* ($t = 1,...,T_i$). There are T_i observations for each individual *i* and a total of N^* observations. X_{ift} is a vector of *k* observed (measured) time-varying exogenous characteristics of individual *i* and firm *f*. θ_i is the person or worker fixed effect (capturing observed and unobserved individual constant heterogeneity), φ_f is the firm fixed effect (capturing observed and unobserved firm constant heterogeneity), and λ_j is the job title fixed effect (capturing observed and unobserved job title constant heterogeneity). According to this equation, there are five components that explain the wage variability:

1. the observed time-varying characteristics of workers, firms, and the economy ($X_{ift}\beta$);
2. the workers' heterogeneity or worker fixed effects (θ_i);
3. the firms' heterogeneity or firm fixed effects (φ_f);
4. the job titles' heterogeneity or job title fixed effects (λ_j); and,
5. an error term component (ε_{ifjt}), assumed to follow the conventional assumptions.

In matrix notation, the stacked system has the form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\varphi + \mathbf{L}\boldsymbol{\lambda} + \boldsymbol{\varepsilon}\,. \tag{2}$$

In this equation $\mathbf{Y}$ is a $(N^* \times 1)$ vector of real hourly wage (in logs), $\mathbf{X}$ is a $(N^* \times k)$ matrix with *k* observed time-varying characteristics of individuals and firms, $\mathbf{D}$ is a $(N^* \times N)$ design matrix for the person effects, $\mathbf{F}$ is a $(N^* \times F)$ design matrix for the firm effects, $\mathbf{L}$ is a $(N^* \times J)$ design matrix for the job title effects, $\boldsymbol{\theta}$

is a $(N \times 1)$ vector of person effects, φ is a $(F \times 1)$ vector of firm effects, $\boldsymbol{\lambda}$ is a $(J \times 1)$ vector of job title effects, and $\boldsymbol{\varepsilon}$ is a $(N^* \times 1)$ vector of disturbances (we assume that mobility is exogenous, in order to make the design matrices orthogonal to the disturbances vector). All vectors/matrices ($\mathbf{Y}$, $\mathbf{X}$, $\mathbf{D}$, $\mathbf{F}$, and $\mathbf{L}$) have row dimensionality equal to the total number of observations (N^*).

Equations (1) and (2) can be interpreted as the conditional expectation of real hourly wages given the observable characteristics of workers and firms, the date of observation, and the identity of individuals, employing firms, and job titles. The total number of parameters to be estimated is therefore $k+N+F+J$. However, it will not be possible to identify all worker, firm, and job title fixed effects. Abowd, Kramarz, and Margolis (1999) show that in order to identify two of those effects (the firm and worker fixed effects) one needs to impose *G* restrictions on the parameters, where *G* is the number of "mobility groups," that is, the number of groups of connected firms and individuals.

The full least squares solution to estimate the parameters in (1) solves the following set of normal equations:

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

Application of the conventional least squares formula to estimate all parameters (worker fixed effects, firm fixed effects, job title fixed effects, and the coefficients of all observed time-varying worker and firm characteristics) requires the inversion of a high dimension matrix. This is impossible to achieve using standard software and present-day computers. Accordingly, special algorithms are required to estimate the full model parameters.

Abowd and Kramarz (1999) and Abowd, Kramarz, and Margolis (1999) proposed an approximate statistical solution that corresponds to using conditional estimation methods (based on a conditioning effects matrix, $\mathbf{Z}$) providing estimators that are as similar as possible to full least squares, but computationally tractable. More recently, Abowd, Creecy, and Kramarz (2002) have developed an algorithm that permits an exact solution of the least squares estimation of equations such as (1), for the two fixed effects case. The user-written command *a2reg* is the Stata implementation of this algorithm.

In the present treatment, we followed an alternative methodology that was able to provide estimates for the regression coefficients and for the three fixed effects. This procedure was developed by Guimarães and Portugal (2010) for the estimation of linear regression models with two high-dimensional fixed effects, and then updated to the three fixed effects case. In brief, this methodology is based on a partitioned algorithm strategy, follows an iterative procedure, and provides an exact solution to the least squares problem. While computationally intensive given its iterative nature, this approach nevertheless imposes minimum memory requirements. A detailed description of this methodology and how it can be implemented to estimate equation (1) is remitted to the Annex.

IV. Data, the Institutional Wage Setting, and Related Literature for Portugal

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, as noted earlier, the survey covers almost all Portuguese wage earners, outside of Public Administration and the 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 in the Annex. 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] The final dataset for the entire period (all 19 available years) comprises 27,020,044 observations drawn from 567,739 different firms, 5,492,332 individual workers, and 95.9 thousand 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). Descriptive statistics for the variables are provided in Tables A.2, A.3, and A.4 (in the Annex).

Institutional Wage Setting

Over our sample period, the wage bargaining system in Portugal is conventionally characterized as having displayed a high degree of centralization and a moderate degree of coordination (OECD, 1997). Insofar as wages are concerned, the greater centralization that occurred in the mid-1980s was via the agency of social pacts that involved indicative wage guidelines for the national average wage increase. Although such pacts were to shape subsequent collective bargaining, the latter still reflected the backdrop of decentralized employers' and workers' organization within their confederate bodies.

That said, collective bargaining in Portugal mainly takes place at sectoral level. Voluntary and mandatory extensions are commonplace. The former occur when one side subscribes to an agreement to which it was not a party (and gains the approval of the other side), or more typically when employers extend the coverage of an agreement they have signed with a particular union to the entire workforce. Mandatory extension by state fiat is also widespread and applies in circumstances where workers are unorganized or where bargaining for some reason fails. Note, however, that sectoral agreements may only have an occupational scope within the industry so that there can be more than one contract within a sector, reflecting occupation, region, trade union affiliation or some combinations of these alternatives.

In some instances, firms can negotiate their own collective agreements with either one or a number of unions or several companies can come together to bargain with the trade unions. But such formally decentralized wage bargaining is the exception rather than the rule. Such single-firm and multi-firm bargains as opposed to sectoral contracts are largely restricted to public enterprises. Note that the recent increase in multi-firm bargaining among joint stock companies is purely the result of a privatization/reorganization process occurring in such enterprises.

Sectoral bargaining in Portugal differs from that in other nations because Portuguese industrial relations are characterized by fragmentation and multiple unionism. The corollary is that contents of collective agreements at once extensive and general. They are extensive in covering a large number of categories of worker but general in setting only minimum conditions for each – in particular, base monthly wages – while dealing with few other terms and conditions. In a bargaining framework that sets wage floors and does not cover projected wage growth, employers have a margin to adjust their wage policies to the prevailing economic conditions (see Cardoso and Portugal, 2005, for a discussion of the ramifications of this *de facto* decentralization).

Related Literature for Portugal

Stimulated by the suitability and richness of information contained in the *Quadros de Pessoal* longitudinal dataset, a number of Portuguese studies seeking to estimate wage regression models with high-dimensional fixed effects have recently emerged. Thus, Monteiro, Portela, and Straume (2010) employ the Guimarães and Portugal (2010) algorithm to estimate a wage regression model with high-dimensional worker and firm fixed effects to investigate the impact of private versus public ownership of firms on the degree of rent sharing. Using the same dataset and technique, Martins and Opromola (2010) estimate a wage regression model with worker and firm fixed effects to study how imports and exports might generate wage premia. Also using the *Quadros de Pessoal*, Carneiro, Guimarães and Portugal (2011) estimate a three-way high-dimensional wage regression model (with worker, firm, and job title fixed effects) to study the cyclical sensitivity of real wages. But in none of these studies is information provided on the decomposition of the distinct sources of wage variation – covariate, worker, and firm effects, inter al. – or, for that matter, presentation of the correlations between different parcels of the extended wage equation permitting an examination of assortative matching.

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

In order to decompose wages variability into the components identified earlier, we first estimated equation (1), where our 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. 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) include both the firms' unobserved and observed but non-time-varying characteristics. Finally, the job title fixed effects (λ_j) include both the unobserved and observed but non-time-varying job titles' characteristics.

We next decomposed the three estimated effects ($\hat{\theta}_i$, $\hat{\varphi}_f$, and $\hat{\lambda}_j$) into their respective observed and unobserved components, by estimating the following three regression equations:

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

where W_i is a vector of non-time-varying worker characteristics (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 – which captures the worker unobserved characteristics effect and can be interpreted as the opportunity cost or the market valuation of worker heterogeneity – is obtained residually by $\hat{\alpha}_i = \hat{\theta}_i - W_i\hat{\eta}$;

$$\hat{\varphi}_f = const. + Z_f \gamma + \varepsilon_f , \tag{5}$$

where Z_f is a vector of non-time-varying firm characteristics (four region dummies, capital ownership [share of domestic capital and share of 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] ϕ_f, the firm-specific intercept (which captures the firm unobserved characteristics effect), is obtained residually, by $\hat{\phi}_f = \hat{\varphi}_f - Z_f \hat{\gamma}$;

and,

$$\hat{\lambda}_j = FE_{occup} + FE_{ca} + \varepsilon_j , \tag{6}$$

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

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

- $X_{ijt}\hat{\beta}$: observed firm, worker, and economy time-varying characteristics (comprising 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.
- $\hat{\varphi}_f$: 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.

(Table 2 near here)

Beginning with Table 2, we observe that the worker fixed effect for females is on average 14.7 percent $[=(exp(-0.15896)-1)\times 100]$ 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 a worker with pre-primary or without any level of completed education (the reference category). Note that these effects are *pure* effects. That is, they result from a regression in which the dependent variable (worker fixed effect) was estimated through a regression that controlled simultaneously for 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 3 near here)

From Table 3 we see that the geographic location of the firm, its capital ownership, its size (as measured by the number of employees), and the 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 (firm fixed effect) was estimated through a regression that controlled simultaneously for time-varying characteristics of workers and firms and for workers' heterogeneity.

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 chosen explanatory variables 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 total outcomes, for the entire period, respectively). 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 attributed to occupation, as the R^2 of an equation with only this explanatory variable explains 46.2 percent of the variability in job title fixed effects, whereas the R^2 of an equation with only the collective agreement explanatory variable 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-component of firm fixed effect; 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).

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), followed by the firm fixed effects component (0.67), by the individual and firm time-varying characteristics component (0.54), and by the job title fixed effects component (0.52). Both observed and unobserved components of worker fixed effects are highly correlated with the log of real total compensation (0.58 and 0.51, respectively). Concerning the components within the firm fixed effects, the observable part of the firm fixed effects is the 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 within 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 of the job title component 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.

(Table 5 near here)

For comparison purposes, 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, Creecy, and Kramarz (2002) found that 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 for France. And for the state of Washington the corresponding values were 0.51 and 0.52, respectively.

We also 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 but not very large (0.27). It is nonetheless considerably larger than that reported in the literature. For example, as we have already noted, Abowd, Creecy, 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. Our earlier review pointed to models in the matching and assignment literature that predict complementarity between worker and firm levels of productivity, suggesting that good workers tend to be found in high-paying firms. Our results are partly consistent with this literature. In terms of the observable characteristics, there is some evidence of positive assortative matching between workers and firms, the correlation coefficient between the corresponding components being 0.33. By the same token, we do not find any evidence of assortative matching in terms of the unobservable characteristics (the correlation is -0.02).[8]

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). (Note that the latter effect indicates there is positive matching with high paying jobs tending 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 but weak and that there are certainly factors other than wage policies that explain the distribution of high-ability workers across firms.

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

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

where C_{ifjt}^{p} is the pth component ($p = 1,...,10$; note thet X_{ift} comprises three components, as described above) that contributes to explaining wage variation. The contribution of each component, C_{ifjt}^{p}, can be calculated as:

$$Cov(ln\ w_{ifjt}\ , C_{ifjt}^{p}\ ) / Var(ln\ w_{ifjt}\ )\ , \qquad (8)$$

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

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 therefore a residual contribution of 8.1 percent. So, comparing worker and job title effects for example, it is evident that what workers 'are' is more important than what workers 'do'.

Among the worker fixed effects, the unobserved sub-component makes a larger contribution (21.0 percent) than the gender and education sub-component (15.0 percent). Among 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). Among the job title fixed effects, the unobserved component makes a much smaller contribution (1.9 percent) than the observed component (7.9 percent).

(Table 7 near here)

Finally, for purposes of comparison, in Table 7 we contrast the main results 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 main results of this comparison 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 remained true that what workers are is more important than what they do. Finally, the ranking of main components is preserved across specifications.

VI. The Relationship between Firms' Wage Policies, Labor Force Quality, and Performance

Firms differ not only in the wage policies they follow but also in the average quality of their workers and the job titles they hold. Accordingly, we next attempt to ascertain whether employing high-wage workers or high-wage job title holders or being a high-wage firm bears any relation to firm performance.

To these ends, we estimated an equation in which the dependent variable is the natural log of sales per employee – a rough measure of productivity that gives some indication of firm performance – and where the explanatory variables are wage components estimated in the previous section (comprising the compensation policy components followed by firms, the quality of their workforce components, and the job title composition components). The results of this exercise are shown in Tables 8 and 9 for two levels of aggregation.

(Tables 8 and 9 near here)

It appears that productivity is positively affected by all compensation components (except for the job title fixed effects), and mainly by the firm fixed effects (Table 8). Productivity is still positively affected by each of the more detailed components of compensation, principally by the firm observed characteristics sub-component (region, capital ownership, and industry), by the firm unobserved characteristics sub-component, and by the worker observed characteristics sub-component (gender and education) (Table 9). Accordingly, high-wage firms (those with above-average firm fixed effects) tend to be the most productive ones and high-wage workers (those with above-average worker fixed effects) tend to work in firms with higher productivity, as predicted by the rent-sharing model, inter al. Interestingly, the job title fixed effects component, however, has a negative relationship with productivity, which comes exclusively from its observed component (occupation and collective agreement). Additional regressions, whose results are not reported here, reveal that this negative sign comes from both variables included in the observed component, occupation and collective agreement.

Following a different procedure for France – first, using the results from a wage equation estimated with an approximate statistical procedure and including only worker and firm fixed effects and, second, estimating performance regressions at the firm level – Abowd, Kramarz, and Margolis (1999) have concluded that the major impact on firms' productivity stems from the time-varying observed characteristics of their workers. Next in importance is the unobserved component of the worker fixed effects followed by the firm fixed effects (selecting the results from the "persons first" method).

Our results, in turn, are in line with those obtained by Mendes, van den Berg, and Lindeboom (2007), who used Portuguese data from the same source and a statistical approach to assess the degree of assortative matching in the Portuguese economy based on the correlation between an estimate of a firm-specific productivity effect and the workforce skill. These authors found evidence of strong positive assortative matching in the Portuguese labor market, with firms and workers of similar productivities tending to match together.

VII. Conclusion

In this exercise we have used a large longitudinal matched employer-employee dataset with a little more than 27 million observations 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 less 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 do 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 overate the cyclicality of wages for incumbent or existing workers and to introduce a counter-cyclical bias in wage cyclicality.

We have also reported that firms hiring 'high-wage' workers and paying higher wages ('high-wage' firms) tend to be more productive firms. On the other hand, the connection between firms' compensation policies and the quality of their workforces was shown to be positive, in marked contrast with some previous evidence and some support for this contention came from job title effects.

Endnotes

1. The reader is directed towards implicit contract theory, principal-agent models, and efficiency wage theories.

2. The earliest rent-sharing studies used industry data (see, for example, 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 et al., 2009).

3. For treatments combining both approaches – equilibrium job search and matching – see Quercioli (1998); Robin and Roux (1998); Mortensen (2000). Recent extensions include Rosholm and Svarer (2004); Cahuc et al. (2005).

4. 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.

5. See also Haltiwanger, Lane, and Spetzer (1999), Andrews, Gill, Schank, and Upward (2008), and van den Berg and van Vuuren (2003) for the United States, Germany, and Denmark, respectively.

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 **Z** are structural characteristics of firms. Their changes over time are either nonexistent or too small to be considered time-varying and to be included as explanatory variables directly in equation (1). The same reasoning applies to the education variable for workers in equation (4) and to occupation and collective agreement in equation (6). Note further that the Portuguese industrial classification system changed in 1995. Because of this and given that the regression covers the entire period, we constructed an aggregated common classification comprising 29 different industries (see Table A.6).

8. In their Norwegian study, Barth and Dale-Olsen (2003, Table 1) report a positive and significant correlation between the observables in the case of low-skilled workers and a negative and significant correlation between the unobservables for both low- and high-skilled workers.

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 (March): 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 (March): 251-333.

Andrews, M. J., L. Gill, T. Schank, and R. 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.

Arai, Mahmood (2003), "Wages, Profitability and Capital Intensity: Evidence from Matched Worker-Firm Data," *Journal of Labor Economics* 32 (July): 593-618.

Barth, Erling and Harald Dale-Olsen (2003), "Assortative Matching in the Labor Market? Stylized Facts about Workers and Plants," Paper presented at the *Comparative Analysis of Enterprise Data Conference*, London.

Becker, Gary (1973), "A Theory of Marriage: Part 1," *Journal of Political Economy* 81 (July-August): 813-846.

Burdett, Kenneth and Dale T. Mortensen (1998), "Wage Differentials, Employer Size, and Unemployment," *International Economic Review* 39 (May): 257-273.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin (2005), "Wage Bargaining with On-the-job Search: Theory and Evidence," Unpublished Paper, CREST-INSEE.

Card, David, Devicienti, Francesco, and Agata Maida (2009), "Rent-Sharing, Holdup, and Wages: Evidence from Matched Panel Data," Unpublished Paper, Center for Labor Economics, University of California at Berkeley.

Cardoso, Ana Rute (1999), "Firms' Wage Policies and the Rise in Labor Market Inequality: The Case of Portugal," *Industrial and Labor Relations Review* 53 (October): 87-102.

Cardoso, Ana Rute (2000), "Wage Differentials across Firms: An Application of

Multilevel Modeling," *Journal of Applied Econometrics* 15 (May-June): 343-354.

Cardoso, Ana Rute and Pedro Portugal (2005), "Contractual Wages and the Wage Cushion under Different Bargaining Settings", *Journal of Labor Economics* 23 (October): 875-902.

Carneiro, Anabela and Pedro Portugal (2008), "Market Power, Dismissal Threat, and Rent Sharing. The Role of Insider and Outsider Forces in Wage Bargaining," *International Journal of Manpower* 29 (1): 30-47.

Carneiro, Anabela, Guimarães, Paulo, and Pedro Portugal (2011), "Real Wages and the Business Cycle: Accounting for Worker, Firm, and Job-Title Heterogeneity," Mimeo, Banco de Portugal. (Forthcoming in *American Economic Journal: Macroeconomics*.)

Diamond, Peter A. (1982), "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies* 49 (April): 217-227.

Dickens, W. T. and L. 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 (2010), "Identifying Sorting – In Theory," Mimeo, London School of Economics. (Forthcoming in *Review of Economic Studies*.)

Ferreira, Priscila (2009), "Returns to Job Mobility," Institute for Social and Economic Research Working Paper 2009-12.

Goux, Dominique and Eric Maurin (1999), "Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data," *Journal of Labor Economics* 17 (July): 492-533.

Gruetter, Max and Raphael Lalive (2009), "The Importance of Firms in Wage Determination," *Labour Economics* 16 (April): 149-160.

Guertzgen, Nicole (2009), "Rent-Sharing and Collective Bargaining Coverage – Evidence from Linked Employer-Employee Data," *Scandinavian Journal of Economics* 111 (June): 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. Spetzer (1999), "Productivity Differences Across Employers: The Roles of Employer Size, Age, and Human Capital," *American Economic Review*, *Papers and Proceedings* 89 (May): 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 (April): 318-337.

Kremer, Michael and Eric Maskin (1996), "Wage Inequality and Segregation by Skill," Working Paper 5718, Cambridge, MA: National Bureau of Economic Research.

Krueger, A. B. and L. H. Summers (1987), "Reflections on the Inter-Industry Wage Structure." In: Kevin Lang and Jonathan S. Leonard (eds.), *Unemployment and the Structure of the Labor Market*, New York: Basil Blackwell, pp. 17-47.

Machado, Josė A. F. and José Mata (2005), "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regressions," *Journal of Applied Econometrics* 20 (May-June): 445-465.

Martins, Pedro and L. Opromolla (2010), "Exports, Imports, and Wages," Mimeo, Banco de Portugal.

Melo de, Raphael Lopes (2008), "Sorting in the Labor Market: Theory and Measurement," Unpublished Paper, Yale University.

Mendes, Rute, Gerard J. van den Berg, and Maarten Lindeboom (2007), "An Empirical Assessment of Assortative Matching in the Labor Market," IZA Discussion Paper 3053, Bonn: Institute for the Study of Labor.

Mincer, Jacob (1974), *Schooling, Experience and Earnings*. New York, NY: Columbia University Press.

Monteiro, Natália, Portela, Miguel, and O. Straume (2010), "Firm Ownership and rent Sharing," Mimeo, Universidade do Minho.

Mortensen, Dale. T. (1986), "Job Search and Labor Market Analysis." In Orley C. Ashenfelter and Richard Layard (eds.), *Handbook of Labor Economics, Volume 2*, Amsterdam: North-Holland, pp. 849-919.

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.

Mortensen. Dale. T. (2003), *Wage Dispersion. Why Are Similar Workers Paid Differently*, Cambridge, MA: MIT Press.

Mortensen, Dale T. and Christopher A. Pissarides (1994), "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies* 61

(April): 397-415.

OECD (1997), *Employment Outlook*, Paris: Organization for Economic Co-operation and Development/OECD.

Pissarides, Christopher. A. (1985), "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review* 75 (September): 676-690.

Pissarides, Christopher A. (2000), *Equilibrium Unemployment Theory (*2nd ed.), Cambridge MA: MIT Press.

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 (November): 989-1016.

Quercioli, Elena (1998), *Firm Specific Training in a Search Equilibrium: Theory and Evidence*, PhD Dissertation, University of Essex.

Robin, Jean-Marc and S. Roux (1998), "Random Matching or Balanced Matching: An Equilibrium Search Model with Endogenous Capital and Matching," 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 (October): 623-645.

Roy, A. D. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers* 3 (June): 135-146.

Sattinger, Michael (1993), "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature* 31 (June): 831-880.

Shimer, Robert and Lones Smith (2000), "Assortative Matching and Search," *Econometrica* 68 (March): 343-369.

Shimer, Robert (2005), "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," *Journal of Political Economy* 113 (October): 996-1025.

Smyth, Gordon K. (1996), "Partitioned Algorithms for Maximum Likelihood and Other Non-Linear Estimation," *Statistics and Computing* 6 (September): 201-216.

Torres, Sónia, Pedro Portugal, John T. Addison, and Paulo Guimarães (2010), "The Sources of Wage Variation: An Analysis Using Matched Employer-Employee Data," Mimeo, Banco de Portugal.

Van den Berg, Gerard J. and Geert Ridder (1998), "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica* 66 (September): 1183-1221.

Van den Berg, Gerard J. and Aico van Vuuren (2003), “The Effect of Search Frictions on Wages,” Working Paper 3979, London: Centre for Economic Policy Research.

Woodcock, Simon D. (2010), “Heterogeneity and Learning in Labor Markets,” *The B.E. Journal of Economic Analysis & Policy* 10 (Issue 1).

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:
Other controls: 18 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] Time-varying observable characteristics of workers (age, age squared, seniority, and seniority squared), time-varying observable characteristics of firms (size), and 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] Time-varying observable characteristics of workers (age, age squared, seniority, and seniority squared), time-varying observable characteristics of firms (size), and 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] Time-varying observable characteristics of workers (age, age squared, seniority, and seniority squared), time-varying observable characteristics of firms (size), and 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: Performance equations, aggregated results

Variable	Dependent variable: productivity (ln sales per employee)	
	Coefficient	t-statistic
Constant	10.20000	6,353.38
Predicted effects of X variables[a]	0.83231	596.15
Worker fixed effects	0.52382	517.03
Firm fixed effects	1.95759	1,776.53
Job title fixed effects	-0.05520	-21.11
Observations	24,637,964	
R-squared	0.184	

Note:

[a] Time-varying observable characteristics of workers (age, age squared, seniority, and seniority squared), time-varying observable characteristics of firms (size), and year dummies.

Table 9: Performance equations, detailed results

Variable	Dependent variable: productivity (ln sales per employee)	
	Coefficient	t-statistic
Constant	10.68916	3,847.87
Time[a]	0.94851	589.58
Time-varying observable characteristics of workers[b]	0.97390	195.39
Time-varying observable characteristics of firms[c]	-0.68172	-190.19
Worker fixed effects: unobserved component	0.43367	382.48
Worker fixed effects: observed component[d]	0.95520	464.62
Firm fixed effects: unobserved component	1.86470	1,380.63
Firm fixed effects: observed component[e]	2.54113	1,270.21
Job title fixed effects: unobserved component	0.01660	4.26
Job title fixed effects: observed component[f]	-0.30816	-87.70
Observations	24,529,242	
R-squared	0.195	

Notes:

[a] Year dummies.

[b] Time-varying observable characteristics of workers (age, age squared, seniority, and seniority squared).

[c] Time-varying observable characteristics of firms (size).

[d] Gender and education.

[e] Region, ownership, and industry.

[f] Occupation and collective agreement.

Annex

Implementing Estimation of the Parameters of the Wage Equation

Here we describe how the algorithm developed by Guimarães and Portugal (2010) can be implemented to estimate the parameters of our wage equation, defined in Section III, which has the following specification:

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{D\theta} + \mathbf{F}\varphi + \mathbf{L\lambda} + \boldsymbol{\varepsilon}\,. \tag{A.1}$$

As stated previously, $\mathbf{Y}$ is a $(N^{*}\times 1)$ vector of real hourly wage, $\mathbf{X}$ 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 (vector $\boldsymbol{\beta}$), as well as the N worker fixed effects (vector $\boldsymbol{\theta}$), the F firm fixed effects (vector φ), and the J job title effects (vector $\boldsymbol{\lambda}$). However, as mentioned earlier, it is not possible to identify all the coefficients for the fixed effects. Abowd, Creecy, and Kramarz (2002) have shown that with two high-dimensional fixed effects one needs to impose one restriction on the coefficients for each mobility group. For three fixed effects a similar logic applies, but identification of the mobility groups is much more complex. In our applications we restricted the sample to the largest mobility group that could be identified by imposing only two restrictions on the coefficients of the fixed effects (therefore rendering them comparable). Next, we discuss the estimation strategy for three high-dimensional fixed effects.

The one high-dimension fixed effect case

As a starting point, consider equation (A.1) without firm fixed effects:

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{D\theta} + \boldsymbol{\varepsilon}\,. \tag{A.2}$$

The normal equations can be written as:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'D} \\ \mathbf{D'X} & \mathbf{D'D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{X'Y} \\ \mathbf{D'X} \end{bmatrix}, \tag{A.3}$$

which can be arranged into:

$$\begin{bmatrix} \mathbf{X'X\beta} + \mathbf{X'D\theta} = \mathbf{X'Y} \\ \mathbf{D'X\beta} + \mathbf{D'D\theta} = \mathbf{D'Y} \end{bmatrix}. \tag{A.4}$$

Solving each set of equations independently leads to the following solutions for $\boldsymbol{\beta}$ and for $\boldsymbol{\theta}$:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{Y} - \mathbf{D\theta}) \\ \hat{\boldsymbol{\theta}} = (\mathbf{D'D})^{-1}\mathbf{D'}(\mathbf{Y} - \mathbf{X\beta}) \end{bmatrix}. \tag{A.5}$$

This suggests an iterative estimation procedure. If $\boldsymbol{\theta}$ were known, the least squares estimates of $\boldsymbol{\beta}$ would be obtained simply by regressing the variable $\mathbf{Y} - \mathbf{D}\hat{\boldsymbol{\theta}}$ on $\mathbf{X}$. If, in turn, $\boldsymbol{\beta}$ were known, the least squares estimates of $\boldsymbol{\theta}$ would correspond to the group means (across workers) of the elements of $\mathbf{u} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. Accordingly, the strategy for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be implemented in the following steps:

1. Run a regression of $\mathbf{Y}$ on $\mathbf{X}$ to obtain starting values for $\boldsymbol{\beta}$;
2. Compute the residuals $\mathbf{u}$ using the last estimate of $\boldsymbol{\beta}$;
3. Estimate $\boldsymbol{\theta}$ as the group (worker) means of $\mathbf{u}$;
4. Estimate $\boldsymbol{\beta}$ by running a regression of $\mathbf{Y}$ on $\mathbf{X}$ and an additional variable, $\mathbf{D\theta}$, computed using the last estimates of $\boldsymbol{\theta}$; and,
5. Return to step 2 and iterate until convergence.

Following this approach all that is required is the estimation of successive linear regressions, by least squares, with $k+1$ explanatory variables, and the computation of group means of the elements of $\mathbf{u}$ in each iteration. We do not need to be concerned about the dimension of $\mathbf{D}$, since the transformation $(\mathbf{D'D})^{-1}\mathbf{D'}$ used to estimate $\boldsymbol{\theta}$ corresponds to a group average and the expression $\mathbf{D\theta}$ used to estimate $\boldsymbol{\beta}$ is a column vector containing all the elements of $\boldsymbol{\theta}$. With this strategy, we avoided the inversion of a large matrix that would be required if we had simply added $\mathbf{D}$ to the set of regressors.

The two high-dimension fixed effects case

Turn now to the situation where we include both worker and firm fixed effects. In this case, solving each set of normal equations independently yields:

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{Y}-\mathbf{D}\boldsymbol{\theta}-\mathbf{F}\varphi) \\ \hat{\boldsymbol{\theta}} = (\mathbf{D'D})^{-1}\mathbf{D'}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{F}\varphi) \\ \hat{\varphi} = (\mathbf{F'F})^{-1}\mathbf{F'}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{D}\boldsymbol{\theta}) \end{bmatrix}. \tag{A.6}$$

The partitioned algorithm can be easily extended to accommodate this case, by iterating between the estimation of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and φ. The algorithm will converge to the least squares solution. Notice that to estimate $\boldsymbol{\beta}$ we need to regress $\mathbf{Y}$ on $\mathbf{X}$ and two additional variables containing the estimates $\boldsymbol{\theta}$ and φ for each observation. At each step, we obtain estimates for $\boldsymbol{\theta}$ calculating the group averages of the residuals $\mathbf{u} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{F}\varphi$ with the estimates of φ being similarly obtained. This means that we can obtain the exact least squares solution without the need to invert a high dimensional matrix. The implementation of the algorithm requires the calculation of various regressions with $k+2$ explanatory variables and averages per group of estimation residuals.

The three high-dimension fixed effects case

If we wish to include a third high-dimensional fixed effect in the regression, we can implement the above described regressions using the within-groups estimator to avoid the direct estimation of third effect coefficients: that is, the only requirement is that we subtract from all variables the average calculated for the groups comprising the third fixed effect.

A familiar disadvantage of this method is the slow convergence rate. However, it is possible to accelerate the algorithm by retaining the estimates of $\boldsymbol{\theta}$ (or φ) produced in the last iterations and using them to adjust the convergence trajectory of the estimates of the fixed effect coefficients.

The standard error estimates associated with the estimation of β may also be obtained avoiding the inversion of a high dimensional matrix. They can be calculated through the application of the Frisch-Waugh-Lovell regression theorem. The strategy consists of firstly expurgating the three fixed effects (using the above described algorithm) of all variables ($\mathbf{Y}$ and $\mathbf{X}$). Next, we run the regression between the transformed $\mathbf{Y}$ and $\mathbf{X}$ variables. This regression, in addition to producing the correct estimates for $\boldsymbol{\beta}$, also produces the correct estimates of the standard errors (whether or not robust) provided that the degrees of freedom associated with the estimate of the variance of the perturbation term are adjusted to reflect the correct degrees of freedom. To calculate the degrees of freedom we need to subtract the number of estimable coefficients from the total number of observations. However, this procedure is complicated by the fact that some of the coefficients for the fixed effects are not identified. Calculating the number of non-identified coefficients requires a special-purpose algorithm. A simpler approach consists of restricting the number of observations on the data set in a way that requires the smallest possible number of non-identified coefficients for the fixed effects (viz. two for the three high-dimensional fixed effects case). An additional advantage of working with this restricted data set – which we shall refer to as “the largest connected group” – is the assurance that estimates of the fixed effects are comparable across all observations.

To identify the largest connected group we use the general approach of Guimarães and Portugal (2010), as described in the Annex. We start out by considering the set of all unique equations defined by the three fixed effects. Next, we select an initial equation and set two of its coefficients to zero, subsequently replacing by zero any occurrences of these two coefficients in other equations. Since the other two coefficients have been set to zero, the third coefficient in the initial equation is now identified and is dully replaced by zero in the system of equations. We keep searching for identified coefficients – namely, those in equations with a single coefficient – and continue to replace them by zero until no further changes are possible. All equations that were set to zero are those that are identified with the two initial restrictions on the coefficients and define a connected group. The procedure can be repeated to identify other connected groups.

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

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

	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)
inds29	Industry affiliation (29 sectors) – common classification from 1986 to 2006	
region	Firm NUTS II region	Norte
		Centro
		Lisboa
		Alentejo
		Algarve
sales	Firm sales (Euros)	
real_sales	Real firm sales (Euros). Deflator: Consumer Price Index (prices of 1986)	
real_sales_employee	Real firm sales (Euros) per employee	
share_n	Firm percentage of domestic capital (0 – 100)	
share_p	Firm percentage of public capital (0 – 100)	

Notes:

* ISCED: International Standard Classification of Education, 1997.

** ISCO: International Standard Classification of Occupations.

Table A.2: Means and standard deviations of continuous variables

Year	*real_hw*		*ln_real_hw*		*age*		*seniority*		*employees*		*ln_employees*		*share_n*		*share_p*		*real_sales*		*real_sales_employee*		*ln_real_sales_employee*	
	Euro; prices of 1986				**Years**		**Months**		**Number**				**%**				**Million Euro; prices of 1986**					
	Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
1986	1.1414	0.7204	0.0071	0.4606	35	11	113.4	95.0	1,686	4,583	5.1217	2.2447	65.9	46.1	15.4	35.1	31.9	124.0	0.0197	0.1886	-4.6634	1.2956
1987	1.1935	0.7590	0.0478	0.4676	35	11	112.9	97.0	1,605	4,458	5.0666	2.2361	69.5	44.8	13.9	33.7	39.1	144.0	0.0522	4.1702	-4.4135	1.3716
1988	1.1845	0.7705	0.0391	0.4685	35	11	108.6	98.6	1,502	4,351	4.9162	2.2336	69.6	44.9	12.5	32.2	39.0	145.0	0.0274	0.0752	-4.2128	1.1201
1989	1.1989	0.8308	0.0411	0.4819	35	11	104.3	99.8	1,119	3,591	4.7820	2.1371	72.2	43.6	9.8	28.8	27.0	120.0	0.0459	3.3378	-4.4146	1.5906
1991	1.4132	1.1025	0.1716	0.5349	35	11	101.8	103.0	1,307	4,045	4.8003	2.2117	70.7	44.3	11.3	30.9	39.0	142.0	0.0286	0.0650	-4.3845	1.6898
1992	1.4897	1.2020	0.2148	0.5489	35	11	98.9	101.6	1,202	3,796	4.7138	2.1931	75.0	42.3	8.4	27.3	44.0	163.0	0.0358	0.7519	-4.1875	1.4091
1993	1.5141	1.2643	0.2224	0.5595	35	11	98.6	100.7	951	2,784	4.6087	2.1621	72.9	43.3	7.6	25.9	49.4	212.0	0.0359	0.1679	-4.1186	1.4148
1994	1.5584	1.3385	0.2428	0.5720	35	11	99.1	100.3	870	2,637	4.4600	2.1857	71.5	43.8	8.0	26.6	27.0	114.0	0.0338	0.1829	-4.1547	1.3016
1995	1.5566	1.3193	0.2456	0.5648	35	11	100.6	101.7	845	2,573	4.4367	2.1822	73.5	42.8	5.5	22.0	35.7	126.0	0.0330	0.0804	-4.0572	1.1964
1996	1.5754	1.3306	0.2588	0.5636	35	11	101.0	102.3	800	2,348	4.4278	2.1863	72.5	43.4	6.2	23.4	34.1	126.0	0.0336	0.0851	-4.0302	1.2019
1997	1.5981	1.3146	0.2843	0.5462	36	11	96.5	101.1	762	2,281	4.3227	2.1822	71.3	44.1	5.2	21.4	32.2	105.0	0.0337	0.0819	-4.0558	1.2688
1998	1.6922	1.3817	0.3397	0.5497	36	11	97.2	102.9	802	2,339	4.3480	2.2211	71.5	43.9	5.0	20.8	48.8	188.0	0.0365	0.1405	-3.9936	1.2754
1999	1.7384	1.4277	0.3662	0.5489	36	10	96.1	102.2	777	2,290	4.2810	2.2160	71.5	44.0	4.6	20.2	36.2	133.0	0.0375	0.1752	-4.0015	1.2652
2000	1.7274	1.3993	0.3655	0.5406	36	10	91.6	100.7	793	2,338	4.1969	2.2233	71.2	44.1	4.1	19.2	32.5	122.0	0.0362	0.1352	-3.9944	1.2189
2002	1.7497	1.4174	0.3772	0.5430	36	10	86.8	98.0	728	2,204	4.0565	2.2293	71.9	44.0	4.0	19.0	40.1	178.0	0.0517	0.5550	-3.9774	1.3415
2003	1.7574	1.4505	0.3780	0.5471	37	10	88.2	96.6	656	1,998	4.0082	2.2141	72.8	43.6	3.5	17.8	39.7	171.0	0.0415	0.2879	-3.9838	1.3046
2004	1.8183	1.5302	0.4045	0.5575	37	10	90.0	96.2	626	1,873	4.0074	2.2089	73.2	43.4	3.3	17.2	36.6	149.0	0.0382	0.1307	-3.9813	1.2440
2005	1.8177	1.5430	0.4029	0.5589	37	10	89.0	95.3	623	1,905	3.9701	2.2130	72.3	43.9	3.1	16.7	32.0	130.0	0.0368	0.1505	-4.0153	1.2740
2006	1.8157	1.5261	0.4022	0.5593	37	10	89.7	95.3	690	2,094	4.0062	2.2400	71.7	44.1	3.5	17.7	50.8	253.0	0.0402	0.2499	-4.0027	1.3099
1986-2006	1.6012	1.3356	0.2806	0.5557	36	11	96.6	99.6	902	2,834	4.3733	2.2342	71.8	43.9	6.4	23.7	37.9	157.0	0.0372	1.0686	-4.1069	1.3294

Table A.3: Distribution across categories of categorical variables (%)

	1986	1987	1988	1989	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2002	2003	2004	2005	2006	1986-2006
gender	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Male	66.9	66.8	65.8	64.5	63.4	62.3	62.2	60.6	60.1	60.3	59.0	58.7	58.1	57.8	58.4	57.9	57.8	57.3	56.7	60.1
Female	33.1	33.2	34.2	35.5	36.6	37.7	37.8	39.4	39.9	39.7	41.0	41.3	41.9	42.2	41.6	42.1	42.2	42.7	43.3	39.9
educ	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Pre-primary or no education	7.7	7.1	6.2	5.6	4.3	4.0	3.6	2.8	2.4	2.3	2.0	1.8	1.6	1.5	1.7	1.6	1.6	1.3	1.2	2.8
First stage of basic education	71.7	71.9	72.7	71.7	68.3	67.1	66.2	64.6	64.0	62.1	61.2	58.8	57.8	55.9	50.7	48.7	46.7	44.6	42.0	58.3
Second stage of basic education	6.9	7.3	7.7	8.4	10.5	11.1	11.6	15.0	14.9	15.3	15.3	16.1	16.3	17.0	19.1	20.0	21.0	21.7	22.4	15.6
Secondary or post-secondary education	10.8	10.8	10.5	11.3	13.1	13.7	14.2	13.0	13.7	14.8	15.6	16.6	17.3	17.9	19.1	19.5	20.1	20.6	21.7	16.2
First stage of tertiary education	1.3	1.2	1.2	1.2	1.5	1.6	1.7	1.6	1.7	1.8	1.9	2.1	2.1	1.9	2.4	2.5	2.6	2.7	2.6	2.0
Second stage of tertiary education	1.6	1.6	1.7	1.8	2.3	2.6	2.8	3.1	3.3	3.6	3.9	4.6	4.9	5.7	7.0	7.6	8.1	9.1	10.2	5.0
inds6	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Mining and quarrying	0.9	0.8	0.7	0.8	0.8	0.7	0.7	0.7	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.5	0.5	0.5	0.4	0.6
Manufacturing	52.7	53.1	52.5	52.4	47.9	46.4	45.2	42.0	42.6	41.9	40.1	38.4	37.9	35.6	30.9	29.6	29.0	27.4	26.2	38.8
Electricity, gas, and water supply	1.8	1.7	1.7	0.7	1.3	1.3	1.2	1.1	1.1	1.1	1.0	1.0	0.9	0.8	0.6	0.6	0.6	0.6	0.6	1.0
Construction	7.7	7.9	8.0	8.6	8.8	8.8	9.0	9.0	9.1	9.4	9.7	9.6	9.6	10.6	12.1	11.5	11.5	11.6	11.3	9.9
Market services	31.3	31.2	31.5	31.6	34.8	35.9	36.8	38.9	40.7	41.1	41.2	42.8	42.9	43.8	45.6	46.9	47.1	46.9	47.5	41.1
Social services	5.6	5.4	5.5	5.9	6.3	6.8	7.2	8.2	5.8	5.9	7.3	7.6	8.0	8.6	10.3	10.9	11.4	13.0	13.9	8.6
size	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
< 5 employees	4.0	4.2	4.7	4.9	5.3	5.6	6.1	7.4	7.7	8.0	8.6	8.9	9.2	9.8	10.7	11.6	11.9	12.3	12.3	8.7
5-9 employees	6.3	6.4	7.1	7.4	7.5	7.8	8.3	9.4	9.6	9.7	10.3	10.4	10.7	11.2	12.5	12.7	12.6	12.6	12.3	10.2
10-49 employees	21.8	22.4	23.9	24.7	24.6	25.4	26.1	27.2	26.7	26.4	27.3	27.0	27.6	28.4	30.7	29.2	28.6	29.0	28.8	27.1
50-99 employees	10.9	11.2	11.6	11.9	12.1	12.1	12.3	12.0	12.1	11.7	12.1	11.5	11.5	11.7	10.0	10.8	10.8	10.7	10.6	11.4
100-249 employees	15.0	14.9	15.3	15.1	14.6	15.1	14.9	13.7	13.9	14.0	13.7	13.6	13.3	12.9	11.6	11.6	12.0	11.6	11.9	13.3
250-499 employees	11.0	10.5	10.1	10.7	9.9	9.7	9.4	9.1	9.1	8.5	8.2	7.9	7.7	7.4	6.8	6.8	6.9	6.8	6.7	8.3
500-999 employees	10.0	10.4	9.1	8.7	8.7	7.5	6.9	5.7	6.1	6.3	5.9	5.7	5.4	4.9	4.5	4.6	4.8	5.0	5.0	6.2
1,000-1,999 employees	6.9	6.4	5.7	6.0	4.8	5.3	4.9	5.1	5.1	4.8	3.9	4.3	4.9	4.7	4.6	4.6	4.5	4.0	4.1	4.8
2,000-4,999 employees	5.9	5.6	5.0	5.2	6.6	6.1	6.0	6.1	6.0	7.3	6.2	6.5	5.1	3.3	3.8	4.0	4.4	4.1	3.4	5.2
≥ 5,000 employees	8.2	7.9	7.6	5.3	5.9	5.4	5.1	4.3	3.8	3.2	3.8	4.3	4.6	5.6	4.9	4.1	3.6	3.9	4.8	4.9
region	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Norte	38.8	38.8	39.5	40.0	38.7	38.4	37.5	38.1	38.9	38.2	38.6	36.8	37.7	36.7	35.3	34.8	34.8	35.0	35.1	37.1
Centro	13.5	14.3	15.3	16.0	16.1	16.3	17.4	17.3	17.4	17.9	17.9	18.4	18.3	18.6	19.1	19.0	19.0	19.1	19.0	17.7
Lisboa	43.0	42.1	39.9	38.6	39.5	39.4	39.1	38.3	37.4	37.3	36.5	37.5	36.7	37.1	37.5	37.7	37.6	37.3	37.3	38.1
Alentejo	2.9	3.0	3.3	3.4	3.5	3.6	3.6	3.8	3.8	3.9	4.3	4.4	4.4	4.5	4.6	4.8	4.8	4.8	4.7	4.1
Algarve	1.8	1.8	2.0	2.1	2.3	2.3	2.4	2.6	2.5	2.7	2.8	2.9	2.9	3.1	3.5	3.8	3.8	3.9	3.9	2.9

Table A.4: Further descriptive statistics on real hourly wages (*real_hw*)

	1986	1987	1988	1989	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2002	2003	2004	2005	2006	1986-2006
Mean	1.1414	1.1935	1.1845	1.1989	1.4132	1.4897	1.5141	1.5584	1.5566	1.5754	1.5981	1.6922	1.7384	1.7274	1.7497	1.7574	1.8183	1.8177	1.8157	1.6012
Stdandard deviation	0.7204	0.7590	0.7705	0.8308	1.1025	1.2020	1.2643	1.3385	1.3193	1.3306	1.3146	1.3817	1.4277	1.3993	1.4174	1.4505	1.5302	1.5430	1.5261	1.3356
Variance	0.5189	0.5760	0.5937	0.6903	1.2155	1.4449	1.5985	1.7916	1.7406	1.7705	1.7282	1.9092	2.0384	1.9580	2.0089	2.1039	2.3414	2.3807	2.3289	1.7838
Coefficient of variation	0.6311	0.6359	0.6505	0.6930	0.7801	0.8069	0.8350	0.8589	0.8476	0.8446	0.8226	0.8165	0.8213	0.8101	0.8100	0.8254	0.8415	0.8488	0.8405	0.8341
Skewness	3.9	3.3	3.9	4.1	3.8	3.9	3.9	5.2	4.1	4.0	4.2	3.8	3.8	3.9	3.7	4.0	3.8	4.7	3.7	4.2
Kurtosis	53.4	28.6	41.1	38.8	28.2	30.4	27.8	120.3	34.0	32.7	38.8	29.6	29.2	29.8	25.0	30.2	25.8	148.4	24.5	51.1
Percentiles																				
1	0.5164	0.5475	0.5391	0.5401	0.5919	0.5805	0.5877	0.5960	0.6149	0.6155	0.6462	0.7040	0.7252	0.7289	0.7364	0.7329	0.7358	0.7348	0.7371	0.5925
5	0.5870	0.6065	0.5912	0.5887	0.6381	0.6522	0.6443	0.6458	0.6598	0.6626	0.7060	0.7313	0.7629	0.7575	0.7576	0.7542	0.7889	0.7844	0.7641	0.6788
10	0.6235	0.6438	0.6393	0.6333	0.6802	0.6962	0.6890	0.6938	0.7055	0.7125	0.7498	0.7910	0.8254	0.8288	0.8374	0.8297	0.8423	0.8384	0.8399	0.7358
25	0.7046	0.7326	0.7295	0.7213	0.7860	0.8141	0.8169	0.8269	0.8342	0.8472	0.8868	0.9357	0.9610	0.9649	0.9713	0.9728	0.9912	0.9846	0.9836	0.8840
50	0.9034	0.9331	0.9273	0.9296	1.0383	1.0824	1.0951	1.1096	1.1118	1.1297	1.1562	1.2058	1.2385	1.2453	1.2667	1.2648	1.2981	1.2934	1.2931	1.1619
75	1.2927	1.3561	1.3475	1.3422	1.5744	1.6560	1.6803	1.7245	1.7128	1.7352	1.7589	1.8660	1.8929	1.8852	1.9214	1.9170	1.9816	1.9739	1.9783	1.7560
90	1.9460	2.0676	2.0388	2.0743	2.6268	2.7810	2.7970	2.9481	2.9451	2.9545	2.9622	3.1732	3.2758	3.1991	3.2035	3.2356	3.3774	3.3765	3.3882	2.9549
95	2.5560	2.6992	2.6382	2.7455	3.4043	3.6363	3.7412	3.9152	3.9140	3.9651	3.9337	4.2292	4.3947	4.3393	4.3708	4.4387	4.6786	4.6931	4.6970	4.0104
99	3.9747	4.2277	4.2957	4.6106	6.0351	6.5558	6.9905	7.2697	7.1583	7.2361	7.1608	7.5327	7.7271	7.5760	7.7910	7.8670	8.3439	8.3873	8.2635	7.2615
Range ratios																				
95/5	4.3545	4.4505	4.4621	4.6641	5.3352	5.5756	5.8067	6.0629	5.9324	5.9838	5.5717	5.7828	5.7609	5.7285	5.7696	5.8851	5.9303	5.9832	6.1474	5.9079
90/10	3.1211	3.2115	3.1894	3.2751	3.8621	3.9945	4.0596	4.2492	4.1747	4.1465	3.9507	4.0117	3.9688	3.8597	3.8255	3.8996	4.0099	4.0272	4.0341	4.0159
75/25	1.8348	1.8512	1.8472	1.8608	2.0031	2.0342	2.0569	2.0855	2.0532	2.0482	1.9834	1.9943	1.9696	1.9538	1.9782	1.9705	1.9992	2.0046	2.0112	1.9865

Table A.5: Real hourly wage (*real_hw*) variability between firms and within firms, 1986-2006

	Mean sum of squares		**B/W**
	Between firms (B)	**Within firms (W)**	
Total	38.2165	1.0018	38.15
Workers' characteristics:			
gender			
Male	36.3965	1.2137	29.99
Female	16.1452	0.5894	27.39
educ			
Pre-primary or no education	1.5865	0.1705	9.30
First stage of basic education	7.7111	0.2791	27.63
Second stage of basic education	10.6261	0.7390	14.38
Secondary or post-secondary education	17.4963	1.2608	13.88
First stage of tertiary education	19.4688	2.6975	7.22
Second stage of tertiary education	41.6678	4.6934	8.88
Firms' characteristics:			
inds6			
Mining and quarrying	25.6579	0.7421	34.57
Manufacturing	36.7146	0.7019	52.31
Electricity, gas, and water supply	316.6787	1.5856	199.72
Construction	7.7563	0.7682	10.10
Market services	38.9309	1.2761	30.51
Social services	31.4632	1.2372	25.43
size			
< 5 employees	1.4303	0.2021	7.08
5-9 employees	4.5583	0.3251	14.02
10-49 employees	31.0635	0.6080	51.09
50-99 employees	137.1016	0.9308	147.29
100-249 employees	456.0715	1.2467	365.81
250-499 employees	995.9859	1.3978	712.52
500-999 employees	2,060.9589	1.3775	1,496.19
1,000-1,999 employees	5,437.2047	1.7777	3,058.65
2,000-4,999 employees	12,437.6790	1.7423	7,138.49
≥ 5,000 employees	35,673.6580	1.8885	18,889.65
region			
Norte	21.2746	0.6693	31.79
Centro	9.9644	0.5259	18.95
Lisboa	73.9201	1.6121	45.85
Alentejo	11.1877	0.6304	17.75
Algarve	5.5049	0.5255	10.48

Table A.6: Legend for the 29 common industries classification from 1986 to 2006

Industry	Description
Industry 1	Mining of coal and lignite; extraction of peat; extraction of crude petroleum and natural gas; service activities incidental to oil and gas extraction, excluding surveying; mining of uranium and thorium ores
Industry 2	Mining of metal ores; other mining and quarrying
Industry 3	Manufacture of food products and beverages; manufacture of tobacco products
Industry 4	Manufacture of textiles; manufacture of wearing apparel; dressing and dyeing of fur
Industry 5	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
Industry 6	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
Industry 7	Manufacture of pulp, paper and paper products; publishing printing and reproduction of recorded media
Industry 8	Manufacture of coke, refined petroleum products and nuclear fuel
Industry 9	Manufacture of chemicals and chemical products
Industry 10	Manufacture of rubber and plastic products
Industry 11	Manufacture of other non-metallic mineral products
Industry 12	Manufacture of basic metals; manufacture of fabricated metal products, except machinery and equipment
Industry 13	Manufacture of machinery and equipment n.e.c.
Industry 14	Manufacture of office machinery and computers; manufacture of electrical machinery and apparatus n.e.c.; manufacture of radio, television and communication equipment and apparatus; manufacture of medical, precision and optical instruments, watches and clocks
Industry 15	Manufacture of motor vehicles, trailers and semi-trailers; manufacture of other transport equipment
Industry 16	Manufacture of furniture; manufacturing n.e.c.; recycling
Industry 17	Electricity, gas, steam and hot water supply; collection, purification and distribution of water
Industry 18	Construction
Industry 19	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel; wholesale trade and commission trade, except of motor vehicles and motorcycles; retail trade, except of motor vehicles and motorcycles; repair of personal and household goods
Industry 20	Hotels and restaurants
Industry 21	Land transport; transport via pipelines; water transport; air transport; supporting and auxiliary transport activities; activities of travel agencies; post and telecommunications
Industry 22	Financial intermediation, except insurance and pension funding; insurance and pension funding, except compulsory social security; activities auxiliary to financial intermediation
Industry 23	Real estate activities; renting of machinery and equipment without operator and of personal and household goods; computer and related activities; research and development; other business activities
Industry 24	Public administration and defence; compulsory social security
Industry 25	Education
Industry 26	Health and social work
Industry 27	Sewage and refuse disposal, sanitation and similar activities; activities of membership organizations n.e.c.; recreational, cultural and sporting activities; other service activities
Industry 28	Private households with employed persons
Industry 29	Extra-territorial organizations and bodies