

IZA DP No. 6016

## **Firm-Worker Matching in Industrial Clusters**

Octávio Figueiredo  
Paulo Guimarães  
Douglas Woodward

October 2011

# Firm-Worker Matching in Industrial Clusters

**Octávio Figueiredo**

*Universidade do Porto*

**Paulo Guimarães**

*Universidade do Porto,  
University of South Carolina and IZA*

**Douglas Woodward**

*University of South Carolina*

Discussion Paper No. 6016  
October 2011

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

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

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

## **ABSTRACT**

### **Firm-Worker Matching in Industrial Clusters<sup>\*</sup>**

In this paper we use a novel approach and a large Portuguese employer-employee panel data set to study the hypothesis that industrial agglomeration improves the quality of the firm-worker matching process. Our method makes use of recent developments in the estimation and analysis of models with high-dimensional fixed effects. Using wage regressions with controls for multiple sources of observed and unobserved heterogeneity we find little evidence that the quality of matching increases with firm's clustering within the same industry. This result supports Freedman's (2008) analysis using U.S. data.

JEL Classification: R12, R39, J31

Keywords: agglomeration, matching, fixed-effects

Corresponding author:

Douglas Woodward  
Division of Research and Department of Economics  
Moore School of Business  
University of South Carolina  
Columbia, SC 29208  
USA  
E-mail: [woodward@moore.sc.edu](mailto:woodward@moore.sc.edu)

---

<sup>\*</sup> The authors acknowledge the support of FCT, the Portuguese Foundation for Science and Technology.

# 1 Introduction

Urban economists have long proposed that three sources of external scale economies explain the benefits of industry agglomeration, which are external to the firm, but internal to an industry concentrated in a particular region [Marshall (1890)]. The first is the potential for more extensive interaction between suppliers and buyers, allowing for vertical disintegration and supplier specialization that leads to higher productivity within the area. The second is a firm's ability to capture industry-specific knowledge and information spillovers resulting from the close proximity of similar firms and other economic agents. The third is labor market pooling, where agglomeration improves each firm's productivity because it increases the quantity of available labor skills and the quality of the firm-worker matching process.

Alfred Marshall's concept of external scale economies underlies much theoretical work, including contributions to our understanding of long-run economic growth, international trade, and economic geography [for example, Rivera-Batiz (1988), Krugman (1991), Venables (1996), Rodríguez-Clare (1996) and Hanson (1996)]. Echoing Marshall (1890), the theoretical literature emphasizes increasing returns for firms stemming from some form of industry-specific external economies.

As for the empirical research, Rosenthal & Strange (2004) and Puga (2010) note that the majority of the evidence for the existence and extent of external economies is indirect and comes from studies showing excessive localization—spatial concentration over and above what would be expected—for a wide range of industries. Work along this line includes Ellison & Glaeser (1997), Maurel & Sedillot (1999), Devereux et al. (2004), Duranton & Overman (2005) and Guimarães et al. (2007). More direct evidence has been obtained in studies comparing wages across areas [Wheaton & Lewis (2002), Combes et al. (2008), Freedman (2008) and Mion & Naticchioni (2009)]. The

idea is that higher wages in clusters reflect higher firm productivity resulting from industry-specific external economies. Indeed, these higher wages would lead firms to relocate elsewhere unless there were some significant compensating productivity advantages in areas where industries agglomerate [Glaeser & Maré (2001)].<sup>1</sup>

Relying on wage data to uncover evidence for industry-specific external scale economies raises some considerable challenges. A first problem is that observed and non-observed abilities of workers may vary across areas. If better workers sort into clusters, then the wage premium may indicate workers' greater abilities, not any intrinsic externalities from clustering [Glaeser & Maré (2001); Combes et al. (2008)]. Likewise, observed and non-observed qualities of firms might differ across regions. Better firms may also sort into clusters, leading to higher mean wages in these areas [Mion & Naticchioni (2009)]. Thus, applied work using wage regressions should control for the possibility of spatial sorting of firms and workers. In turn, this approach requires matched employer-employee micro-level data.

Wage differences across areas can also be caused by local non-human endowments that boost firms' productivity and the marginal product of workers. Firms in some areas may exhibit higher productivity because of the natural features of a favorable location, such as a climate suited to a particular kind of economic activity or the presence of natural resources in the area [Ellison & Glaeser (1999) and Kim (1999)]. The built environment and human endowments, such as public infrastructure or local institutions and technology, may also matter. If these forces are at work, then larger

---

<sup>1</sup>Glaeser (1999) and Glaeser & Maré (2001) used wage regressions to look for evidence on urban (not industry-specific) external economies. They advanced the idea that the skills of firms and workers evolve more positively over time in urban areas, the so-called "learning hypothesis." A more direct approach to gauge the importance of external economies (urban or industry-specific) relied on productivity measures [e.g. Ciccone & Hall (1996), Ciccone (2002) and Henderson (2003)].

mean wages in clusters may not indicate the presence of industry-specific external scale economies. In addition, we should be aware that although empirical evidence on a wage premium may support the existence of productivity gains associated with external economies, the lack of evidence on such premium does not mean these economies are absent. Roback (1982) showed that the presence of local endowments, like natural or consumption amenities, will make workers more willing to accept lower wages. Thus, these localized amenities can offset the positive impact of industry-specific external economies, rendering the net effect on wages ambiguous.

In applied studies researchers must also take into account the possibility that the wage premium in clusters can be related to productivity gains associated with urban external economies [Jacobs (1969)]. These are economic benefits that accrue from the agglomeration of general economic activity, not from the spatial concentration of a particular industry.

Another empirical problem is disentangling the three sources of external economies originally proposed by Marshall (1890). The difficulty arises because all the three sources—vertical disintegration, knowledge spillovers, and better quality of the firm-worker matching process—share the prediction that productivity increases with the scale of an industry at a location, allowing firms to pay higher wages. This "observational equivalence" [Rosenthal & Strange (2004)] makes it complicated to distinguish the three main causes of industry-specific external economies using wage regressions. Thus, higher wages in industrial clusters, after controlling for the spatial sorting of both firms and workers, local human and non-human endowments, and urban external economies, can be seen as evidence that either one, two, or all three sources of external economies proposed by Marshall (1890) are at work.

In this paper we use wage regression analysis to test the Marshallian hypothesis that industrial clustering improves the quality of the firm-worker matching process. A

major advantage is the availability of a large Portuguese panel data base, with linked employer-employee information. Our work is in line with Andersson et al. (2007) and Mion & Naticchioni (2009), who also used wage regressions and micro-level data to examine the hypothesis that matching improves with agglomeration. Unlike us, however, these authors examined the relationship across urban agglomerations, not industrial clusters. Both studies computed match quality as the correlation between estimates of firm quality and its mean worker quality for each area. They then related this correlation with a measure of urban agglomeration (employment density across areas).<sup>2</sup> The two papers present conflicting evidence. Andersson et al. (2007) found a positive relation between match quality and urban agglomeration using data for California and Florida, while Mion & Naticchioni (2009) uncovered a negative relationship relying on an Italian data set.

The main problem with these two studies has to do with their estimates of worker and firm quality. Andersson et al.'s (2007) estimates are based on comprehensive data sets and on a wage regression that includes two high-dimensional fixed effects (firm and worker) following Abowd et al.'s (1999) model and econometric procedures.<sup>3</sup> On the other hand, the sampling procedure and relatively small size of Mion & Naticchioni's (2009) data base prevents them from using Abowd et al.'s (1999) specification. Thus, their estimates of worker quality are based on a regression with a single fixed effect for worker, while firm quality is proxied by a measure of firm size.

Recent work has convincingly argued that in the presence of unobserved worker,

---

<sup>2</sup>Andersson et al. (2007) also analyzed the relationship between matching and urban agglomeration using a productivity approach.

<sup>3</sup>See also Abowd, Lengeremann & McKinney (2002). The estimates of these individual fixed effects are used to measure the quality of each worker and firm.

firm, and match heterogeneity, wage regressions that do not control for all these unobservables may suffer from a considerable omitted variable bias [Woodcock (2007) and Woodcock (2008)]. Thus, the estimates of individual quality of workers and firms in Andersson et al. (2007) and Mion & Naticchioni (2009) may be plagued by this problem. As an alternative, Woodcock (2007) proposed the introduction of a worker-firm interaction term (the match-effect) in Abowd et al.'s (1999) model.

So far, the only study that uses micro-level data to test the hypothesis that the quality of match improves with firms' clustering within the same industry is Freedman (2008). Looking at data for a single manufacturing sector of a U.S. state, the author finds little evidence that the quality of matching increases with firm's clustering. His inference, however, is based on an *ad hoc* comparison of results from a wage regression *à la* Abowd et al. (1999) with another one that, in line with Woodcock's (2007) suggestion, also controls for unobservable worker-firm match effects.

Our paper improves on the existing literature, notably Freedman's (2008) research, by establishing a precise econometric framework to test the relationship between industrial clustering and matching. Moreover, our results are obtained with a more comprehensive data set that includes all manufacturing sectors in the economy. The paper proceeds as follows. We devote the next section to the discussion of our methodology. Then, in the third section we present the data and some descriptive statistics. Results are discussed in the fourth section, while section five concludes the paper.

## 2 Econometric Framework

Consider an augmented version of the traditional Mincerian wage equation for a single worker where we added a term to account for the impact of industry-specific external economies on wages. More formally, let

$$\ln w_{ijt} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{jt}\boldsymbol{\alpha} + L_{r(j)s(j)t}\gamma + \varphi_{s(j)} + \psi_t + \varepsilon_{ijt} , \quad (1)$$

where  $w_{ijt}$  is the wage of worker  $i$ , in firm  $j$ , at time  $t$ . The  $\mathbf{x}_{it}$  is a vector of observable worker level characteristics (such as age, education, gender or tenure), while the  $\mathbf{z}_{jt}$  is a set of observable firm level attributes (like its size, age or type of ownership). Other variables include a set of controls (dummies) for time-specific ( $\psi_t$ ) and inter-industry ( $\varphi_{s(j)}$ ) wage differentials.<sup>4</sup> Our variable of interest is  $L_{r(j)s(j)t}$ , a measure that is introduced in the regression to pick-up a potential wage premium linked to industry-specific externalities. As argued in the introduction, a proper specification should control for urbanization economies, as well as for regional human and non-human endowments that might affect individual wages. Thus, we add to the wage equation in (1) two new variables. The variable  $\tau_{r(j)}$  in equation (2) is a dummy variable for each region that accounts for time-invariant characteristics of the areas (including climate, natural amenities or other natural resources). This variable also controls for permanent inter-regional differences in variables such as local institutions and technology, infrastructures, or urbanization economies. The time-varying characteristics of the regions are picked up by the variable  $U_{r(j)t}$ . Thus, the wage equation becomes,

$$\ln w_{ijt} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{jt}\boldsymbol{\alpha} + L_{r(j)s(j)t}\gamma + U_{r(j)t}\eta + \tau_{r(j)} + \varphi_{s(j)} + \psi_t + \varepsilon_{ijt} . \quad (2)$$

---

<sup>4</sup>In equation (1)  $s$  stands for firms' sector of activity and  $r$  for region. We adopt the convention of using parentheses in the subscripts to specify the source of variation in each variable. For example, the ultimate source of variation in  $\varphi_{s(j)}$  are the firms ( $j$ ).

Estimation of the above specification may produce biased results. The problem is that non-observed abilities of workers and firms may be correlated with the regressors. At the same time, as already noted, if these unobserved abilities are positively correlated with the  $L_{r(j)s(j)t}$ , then higher wages in clusters may be a result of spatial sorting of workers and firms based on unobservables, not industry-specific external economies. As proposed by Abowd et al. (1999), with a large matched employer-employee panel data set it is possible to account for the non-observable characteristics of workers and firms. This can be done by adding two fixed effects, one specific to the worker and the other one specific to the firm. In this case, our specification becomes,

$$\ln w_{ijt} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{jt}\boldsymbol{\alpha} + L_{r(j)s(j)t}\gamma + U_{r(j)t}\eta + \psi_t + \theta_i + \delta_j + \varepsilon_{ijt} , \quad (3)$$

where we note that those variables that change only with  $j$ ,  $\tau_{r(j)}$  and  $\varphi_{s(j)}$ , are completely absorbed by the firm fixed effect. The introduction of these two fixed effects will also assimilate all other time-invariant observable characteristics of workers and firms that might affect wages. With a high-dimensional data set, estimation of a linear regression model with two fixed effects poses some computational challenges [see Abowd et al. (1999)]. However, the exact least-square solution to this problem can be found using an algorithm proposed by Guimarães & Portugal (2010).

As shown by Woodcock (2007), results obtained with specification (3) may be subject to substantial bias if unobservable firm-worker match characteristics are important determinants of wages. Following Woodcock (2007), we introduce an additional term in the regression ( $\lambda_{ij}$ ) that accounts for the specific firm-worker interaction. This leads to a model with three fixed effects accounting for unobservables. Thus, our specification becomes:

$$\ln w_{ijt} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{jt}\boldsymbol{\alpha} + L_{r(j)s(j)t}\gamma + U_{r(j)t}\eta + \psi_t + \theta_i + \delta_j + \lambda_{ij} + \varepsilon_{ijt} . \quad (4a)$$

Estimation of a model such as this poses some problems. As is, the model is overparameterized making it impossible to disentangle the three effects. In this model a good match may be indistinguishable from a good worker working in a good firm. In other words, without any restriction on the parameters,  $\lambda_{ij}$  absorbs the effect of  $\theta_i$  and  $\delta_j$ , meaning that a model such as

$$\ln w_{ijt} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{z}'_{jt}\boldsymbol{\alpha} + L_{r(j)s(j)t}\gamma + U_{r(j)t}\eta + \psi_t + \omega_{ij} + \varepsilon_{ijt} , \quad (4b)$$

that includes a single fixed effect for the interaction,  $\omega_{ij}$ , will capture the three effects and provide the same fit as (4a), i.e., identical estimates for  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha}$ ,  $\gamma$ ,  $\eta$  and  $\boldsymbol{\psi}$ .

However, we have to keep in mind that our main interest is in the relation between  $\lambda_{ij}$  and  $L_{r(j)s(j)t}$ , after controlling for all other explanatory variables in equation (4a). That is, we are interested in the  $\gamma^*$  in the relation

$$\lambda_{ij} = \mathbf{x}'_{it}\boldsymbol{\beta}^* + \mathbf{z}'_{jt}\boldsymbol{\alpha}^* + L_{r(j)s(j)t}\gamma^* + U_{r(j)t}\eta^* + \psi_t^* + \theta_i^* + \delta_j^* + \varepsilon_{ijt} . \quad (5)$$

One way to test for this relationship would be using a two-step procedure. In the first step we would obtain estimates of  $\lambda_{ij}$  from (4a). Then, in a second step, these estimates would be regressed in the other explanatory variables. Although intuitive, this approach faces a difficulty. Because the model is overparameterized, to separate the three effects it is necessary to impose restrictions on the parameters associated with the fixed effects in order to obtain estimates of  $\lambda_{ij}$ . Conceivably, there are many ways in which these restrictions can be imposed, and the estimates of  $\lambda_{ij}$  will depend on the restriction.<sup>5</sup> An interesting result we obtained when studying this problem is that the results of a regression between the estimated  $\lambda_{ij}$  and all other explanatory variables in (4a) will be invariant to the type of parameterization used for the fixed

---

<sup>5</sup>For example, Woodcock (2007) suggests a strategy for identification of  $\lambda_{ij}$  based on an assumption of orthogonality between the match-effect and the firm and worker effects.

effects (see Appendix A for a proof of this result). Thus, this two-step procedure is feasible without being affected by the type of parameterization we will use.

In Appendix B, we also show that the coefficients of a regression between the estimated  $\lambda_{ij}$  and all other explanatory variables in equation (4a) can be obtained directly by comparing the estimated coefficients of (3) and (4b). This result provides an alternative way to obtain our coefficient of interest, the  $\gamma^*$  in equation (5). More specifically, to obtain this coefficient we need only to subtract the estimated  $\gamma$  obtained in (4b) from that obtained in (3). To infer about the statistical significance of the difference in the coefficients we can then make use of a test proposed by Gelbach (2009).<sup>6</sup>

### 3 Data and Descriptive Statistics

We use a survey of workers, firms and establishments collected in October of every year during a reference week by the Portuguese Ministry of Employment, the *Quadros do Pessoal* data base. This is a mandatory survey for every firm operating in Portugal, except family businesses without wage-earners. Public administration is not covered and the coverage in agriculture is low, given the small share of businesses with wage-earners. For the other sectors, however, the mandatory nature of the survey leads to an extremely high response rate. The data set includes precise information on firm and establishment location, sector of activity, type of ownership, actual employment, and the characteristics of the workforce. For every single worker

---

<sup>6</sup>See section 7 in Gelbach (2009), in particular footnote 22. The second-step regression we described above can be interpreted as the auxiliary regression associated with Gelbach's (2009) decomposition. He shows that the asymptotic t-tests of this regression can be seen as an extended version of a Hausman test.

the reported data encompasses earnings and other personal information such as gender, age, tenure, and years of schooling. A unique worker identification code, based on a transformation of the social security number, allows for the tracking of workers over time. Similarly, unique identifiers for firms and establishments enables connecting data throughout the years. Matching of firms, establishments and workers' identifiers is also possible.

We constructed a panel of workers using data from 1995 through 2006.<sup>7</sup> We then restricted data to the manufacturing sector in the continental part of the Portuguese territory. Extensive checks on the consistency of the data were implemented following the methodology described in Cardoso & Portela (2009). Next, we trimmed the top and bottom one percent of the wages in each year to avoid problems with outliers and retained only wage-earners working full-time.<sup>8</sup> To ensure comparability of the estimates of the fixed effects, we restricted the data set to the largest connected group.<sup>9</sup> In our data, the largest connected group accounts for 95.8 percent of the observations.

Some basic descriptive statistics of our panel are shown in Table 1. We have a total of 5,245,296 worker-year observations. Columns 2 and 3 in this table show means

---

<sup>7</sup>Data was restricted to this period of time to avoid changes in the Portuguese Standard Industrial Classification system (*CAE*). Throughout this period the Portuguese *CAE Rev.2* classification remained in place. Worker-level data for 2001 is unavailable in the *Quadros do Pessoal*. Thus, our panel covers a period of eleven years.

<sup>8</sup>If in a single year the worker is found in more than one establishment we keep the observation where the highest number of worked hours is reported.

<sup>9</sup>Estimates of the fixed effects obtained for the regression model with two fixed-effects are only comparable within the same group. Groups are defined as the set of observations comprising all the workers that ever belonged to any firm in the group and all the firms that employed any worker in the group. Identification of the groups has been implemented using the algorithm in Abowd, Creecy & Kramarz (2002).

and standard deviations calculated across all observations, the "Weighted Sample." To furnish more meaningful statistics we also report these metrics calculated on the averages of time-values – the "Unweighted Sample" figures in columns 4 and 5. As shown in column 4, the 1,005,886 workers in our sample have an average real hourly wage of 4.4 euros – using a 2009 deflator.<sup>10</sup> Table 1 also shows statistics for the observable worker characteristics we will use in our regressions. *School1* to *School8* are dummy variables and *Tenure* is defined as the number of consecutive years in the same establishment.

In our regressions we included establishment fixed effects. However, we control for firm level observable characteristics such as size and type of ownership. *Size* is defined as the number of full-time workers in the firm and there are three types of ownership (*Private*, *Public* and *Foreign*), according to the majority in the firm's capital structure. These variables are relevant because the real hourly wage of workers may depend on the characteristics of the firms to which establishments belong.<sup>11</sup>

[insert Table 1 about here]

Our variable of interest, *Specialization*, is computed at the *concelho* (county) level and using a three-digit breakdown of the Portuguese Standard Industrial Classification system (105 industries).<sup>12</sup> We calculated this variable using two alternative ways,

---

<sup>10</sup>Wage is calculated as the sum of the base wage plus all other regularly paid components. To obtain hourly values, we divided by the number of normal working hours. Comparison of the figures for the real hourly wage in columns 2 and 4 indicates that workers that remain more years in the panel are remunerated above those who stay for shorter periods of time.

<sup>11</sup>Note, for example, that a worker in an establishment of a large firm is likely to be paid more than another that works in an equally sized establishment of a smaller firm.

<sup>12</sup>The *concelho* is a Portuguese administrative region roughly equivalent to the U.S. county, but with a smaller average area. Throughout our period of analysis the number of *concelhos* increased

employment or counts of establishments in the same industry and region. Typically, industry-specific externalities are captured with employment data [e.g. Wheaton & Lewis (2002), Combes et al. (2008), Freedman (2008) and Mion & Naticchioni (2009)]. In these studies, regional employment in each industry, is often introduced in the regressions as a density measure (divided by the area of the region) or as a share (with total regional employment in the denominator). We opted to include all these three variables individually in the regressions (*Specialization*, *Area* and *Urbanization*) in order to allow for a more flexible specification. The inclusion of total regional employment (*Urbanization*) serves an additional purpose. This variable picks up the effect of time-varying characteristics of the regions such as urbanization economies, infrastructure, and other local amenities that may affect productivity and wages.

As argued in Figueiredo et al. (2009) and Figueiredo et al. (2010), use of employment-based measures to account for industry-specific external economies has a drawback. These measures encompass both firm internal scale economies and external economies. Consider, for example, a region with a cluster of 1,000 small firms with one worker when compared with another with a single large firm with 1,000 workers. In the second case, we do not have any cluster of firms and the level of the employment variable is entirely explained by internal returns to scale. Hence, as an alternative to overcome this limitation of employment-based measures, we also compute *Specialization* as the count of establishments in each industry and region.

---

from 275 to 278. The three new *concelhos* were created by aggregating parts of five existing ones. We overcame this problem by grouping together the affected *concelhos* and ended up with 273 regions. Our panel only includes data for 272 of these regions because one was dropped for lack of observations.

## 4 Results

In Table 2 we show regression results using our preferred specification – with *Specialization* measured as a count of establishments by industry and region. All non-dummy variables are introduced in logs. In Columns (1) and (2), we present simple wage regressions that indicate a raw wage premium for clustering. Whether or not we adjust for the area of the regions, the elasticity of wages is around 0.01. Doubling the number of establishments in a region leads to an increase of wages in the same industry of about one percent. The last four columns follow the sequence of equations 1-4 presented in Section 2.<sup>13</sup> Column (3) shows the estimates for a traditional Mincerian equation, which includes observable characteristics of the worker and the firm. Goodness of fit improves and the coefficient on *Specialization* increases slightly. All other estimated coefficients are in line with expectations. Wages are higher for males, older workers (peaking around the age of 58), and increase with tenure and education. There is also a wage premium for working in larger firms, public companies, and especially foreign-owned firms.

As argued before, a proper specification should account for urbanization economies, as well as for regional endowments that might affect productivity and individual earnings. Thus, in column (4), we add two new controls: A set of individual dummies for regions and the *Urbanization* variable. The dummies are intended to account for time-invariant characteristics of the regions (e.g. climate, amenities or natural resources). These variables also pick permanent interregional differences in regional characteristics such as institutions, technology and infrastructures. The other vari-

---

<sup>13</sup>All models were estimated by ordinary least squares with a cluster-robust correction to the standard-errors. This correction accounts for possible unobserved correlation between repeated observations (i.e. the same worker in different years) and produces rather conservative *t*-statistics.

able, *Urbanization*, controls for urbanization economies and related time-varying attributes of the areas. It is interesting to note that the wage premium associated with the *Urbanization* variable is around three percent, in line with other studies [Ciccone & Hall (1996), Ciccone (2002), Combes et al. (2008) and Mion & Naticchioni (2009)]. Even though we now rely on variation over time within industries to identify the relation, we still find an elasticity for *Specialization* in line with the previous regressions. *Area* is absorbed by the location dummies. These dummies also serve to mitigate a potential problem of endogeneity due to regional omitted variables that might be correlated with the other explanatory variables.

Another potential problem, as already discussed, is that non-observed abilities of workers and establishments may be correlated with the regressors. If these unobserved abilities are positively associated with *Specialization*, then higher wages in clusters may be a result of spatial sorting. Thus, as explained in Section 2, we introduce two sets of fixed effects, one specific to the worker and the other to the establishment. This regression is found in column (5).<sup>14</sup> The coefficient on *Specialization* drops to less than half its previous values, showing that sorting based on unobservables matters.

Finally, in column (6), following Woodcock (2007), we introduce an establishment-worker specific fixed-effect that accounts for match heterogeneity. As indicated in Section 2, if no restriction is imposed on the coefficients, this match effect absorbs the worker and establishment fixed effects, rendering this a model with a single high-dimensional fixed effect. Note also that the estimates for the model in column (6) are equivalent to those that would be obtained in a specification where the three fixed effects (worker, establishment and interaction) were included with appropriate

---

<sup>14</sup>The regression was estimated with the *Stata* user-written routine *reg2hdfe* found in the *Statistical Software Components (SSC)* from the Boston College Department of Economics.

restrictions on the coefficients.

[insert Tables 2 and 3 about here]

We are interested in the relation between the establishment-worker match effect and *Specialization*. The coefficient of interest, the  $\gamma^*$  in equation (5), Section 2, can be obtained directly, as shown before, by subtracting the estimated coefficient of *Specialization* in column (6) from that on column (5). Thus, the estimate of  $\gamma^*$  is 0.00018 (=0.00411-0.00393), indicating that with the doubling of *Specialization* the component of wages associated with match-effects increases 0.018 percent – after controlling for multiple sources of observed and unobserved heterogeneity.

To check whether this result is statistically significant, we implemented the test described in Gelbach (2009). In a first step, we recovered the estimates of the  $\omega_{ij}$  using the model in equation (4b) (see also column (6) of Table 2) and decomposed the three effects that are included in these estimates based on the assumption of orthogonality between the match effect and the establishment and worker effects.<sup>15</sup> We then apply the second step regression shown in equation (5). Using this two-step procedure, we obtain the same 0.00018 for the  $\gamma^*$  (as expected) with an associated p-value of 3.1 percent. Hence, our estimate indicates little evidence that the quality of matching increases with establishment’s clustering within the same industry. The size of the coefficient is small and the p-value is not significant at the one percent level. Similar evidence can be found using our alternative measure of *Specialization* based on employment (see Table 3). Here, the difference in coefficients is 0.00003 and the p-value associated with this difference is 77.7 percent.

---

<sup>15</sup>To do this we followed the approach in Woodcock (2007).

## 5 Conclusion

In this paper we investigate the Marshallian hypothesis that localization of an industry improves the firm-worker matching process. To this end, we use a large Portuguese linked employer-employee data set and a novel approach that makes use of recent developments in the estimation and analysis of models with high-dimensional fixed effects. Relying on micro-level wage regressions with controls for multiple sources of observed and non-observed heterogeneity, we find little evidence that the quality of matching increases with firms' clustering within the same industry. This result extends on Freedman (2008), who reached the same conclusion looking at data for a single manufacturing sector of an undisclosed U.S. state. Our result is obtained with a more comprehensive panel data set that includes all manufacturing sectors in the economy. Moreover, we improve on Freedman's (2008) approach by establishing a precise econometric framework to test the relationship between industrial clustering and matching. Indeed, Freedman's (2008) conclusions are based on an *ad-hoc* comparison of results from the two-fixed effects model of Abowd et al. (1999) with another one that, in line with Woodcock's (2007) three fixed effects model, also controls for unobservable worker-firm match effects.

Despite having not found much evidence on Marshall's suggestion that localization of an industry improves matching, our regressions in columns (6) of Tables 2 and 3 still show a positive and significant wage premium for clustering. After controlling for a large variety of sources of heterogeneity, clustering within the same industry still matters. Thus, the other Marshallian sources of industry-specific external economies may be at work. At the same time, since in the regressions in columns (6) we rely solely in time variation to identify the relationship, the learning hypothesis advanced by Glaeser (1999) and Glaeser & Maré (2001) is also a plausible explanation. Unob-

served abilities of workers and firms may evolve over time more positively in clusters and this source of variation is not captured by the three-fixed effects model.

## References

- Abowd, J., Kramarz, F. & Margolis, D. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–333.
- Abowd, J., Lengermann, P. & McKinney, K. (2002), The measurement of human capital in the U.S. economy. Technical Working Paper 2002-09, U.S. Census Bureau, Washington DC.
- Abowd, J. M., Creecy, R. H. & Kramarz, F. (2002), Computing person and firm effects using linked longitudinal employer-employee data. unpublished manuscript.
- Andersson, F., Burgess, S. & Lane, J. (2007), 'Cities, matching and the productivity gains of agglomeration', *Journal of Urban Economics* **61**(1), 112–128.
- Cardoso, A. R. & Portela, M. (2009), 'Micro foundations for wage flexibility: Wage insurance at the firm level', *Scandinavian Journal of Economics* **111**(1), 29–50.
- Ciccone, A. (2002), 'Agglomeration effects in europe', *European Economic Review* **46**(2), 213–227.
- Ciccone, A. & Hall, R. E. (1996), 'Productivity and the density of economic activity', *American Economic Review* **86**(1), 54–70.
- Combes, P.-P., Duranton, G. & Gobillon, L. (2008), 'Spatial wage disparities: Sorting matters!', *Journal of Urban Economics* **63**(2), 723–742.

- Devereux, M. P., Griffith, R. & Simpson, H. (2004), ‘The geographic distribution of production activity in the UK’, *Regional Science and Urban Economics* **34**(5), 533–564.
- Duranton, G. & Overman, H. (2005), ‘Testing for localisation using micro-geographic data’, *Review of Economic Studies* **72**(4), 1077–1106.
- Ellison, G. & Glaeser, E. (1997), ‘Geographic concentration in U.S. manufacturing industries: A dartboard approach’, *Journal of Political Economy* **105**(5), 889–927.
- Ellison, G. & Glaeser, E. (1999), ‘The geographic concentration of an industry: Does natural advantage explain agglomeration’, *American Economic Review* **89**(2), 311–316.
- Figueiredo, O., Guimarães, P. & Woodward, D. (2009), ‘Localization economies and establishment size: Was Marshall right after all?’, *Journal of Economic Geography* **9**(6), 853–868.
- Figueiredo, O., Guimarães, P. & Woodward, D. (2010), ‘Vertical disintegration in marshallian industrial districts’, *Regional Science and Urban Economics* **40**(1), 73–78.
- Freedman, M. L. (2008), ‘Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry’, *Journal of Urban Economics* **64**(3), 590–600.
- Gelbach, J. B. (2009), When do covariates matter? and which ones, and how much? Department of Economics Working Paper 2009-07, University of Arizona, Tucson.

- Glaeser, E. L. (1999), ‘Learning in cities’, *Journal of Urban Economics* **46**(2), 254–277.
- Glaeser, E. L. & Maré, D. C. (2001), ‘Cities and skills’, *Journal of Labor Economics* **19**(2), 316–342.
- Guimarães, P., Figueiredo, O. & Woodward, D. (2007), ‘Measuring the localization of economic activity: A parametric approach’, *Journal of Regional Science* **47**(4), 753–774.
- Guimarães, P. & Portugal, P. (2010), ‘A simple feasible procedure to estimate models with high-dimensional fixed effects’, *Stata Journal* **10**(4), 628–649.
- Hanson, G. H. (1996), ‘Localization economies, vertical organization, and trade’, *American Economic Review* **86**(5), 1266–1278.
- Henderson, J. V. (2003), ‘Marshall’s scale economies’, *Journal of Urban Economics* **53**(1), 1–28.
- Jacobs, J. (1969), *The Economy of Cities*, Random House, New York.
- Kim, S. (1999), ‘Regions, resources and economic geography: Sources of U.S. regional comparative advantage 1888-1987’, *Regional Science and Urban Economics* **29**(1), 1–32.
- Krugman, P. (1991), ‘Increasing returns and economic geography’, *Journal of Political Economy* **99**(3), 483–499.
- Marshall, A. (1890), *Principles of Economics*, Macmillan. Eighth edition, 1920, London.

- Maurel, F. & Sedillot, B. (1999), ‘A measure of the geographic concentration in french manufacturing industries’, *Regional Science and Urban Economics* **29**(5), 575–604.
- Mion, G. & Naticchioni, P. (2009), ‘The spatial sorting and matching of skills and firms’, *Canadian Journal of Economics* **42**(1), 28–55.
- Puga, D. (2010), ‘The magnitude and causes of agglomeration economies’, *Journal of Regional Science* **50**(1), 203–219.
- Rivera-Batiz, F. L. (1988), ‘Increasing returns, monopolistic competition, and agglomeration in consumption and production’, *Regional Science and Urban Economics* **18**(1), 125–153.
- Roback, J. (1982), ‘Wages, rents, and the quality of life’, *Journal of Political Economy* **90**(6), 1257–1278.
- Rodríguez-Clare, A. (1996), ‘Multinationals, linkages, and economic development’, *American Economic Review* **88**(4), 852–873.
- Rosenthal, S. & Strange, W. (2004), Evidence on the nature and sources of agglomeration economies, in J. V. Henderson & J. F. Thisse, eds, ‘Handbook of Regional and Urban Economics’, Elsevier, Amsterdam, pp. 2119–2171.
- Venables, A. (1996), ‘Localization of industry and trade performance’, *Oxford Review of Economic Policy* **12**(3), 52–60.
- Wheaton, W. C. & Lewis, M. J. (2002), ‘Urban wages and labor market agglomeration’, *Journal of Urban Economics* **51**(3), 542–562.

Woodcock, S. (2007), Match effects. Discussion Paper 2007-13, Simon Fraser University, British Columbia.

Woodcock, S. (2008), 'Wage differentials in the presence of unobserved worker, firm, and match heterogeneity', *Labour Economics* **15**(4), 403–418.

## Appendix A: Invariance of Transformation in OLS

Consider the following regression model where the matrix of explanatory variables is partitioned into two sets of regressors,  $\mathbf{X}_1$  and  $\mathbf{X}_2$ :

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} .$$

Let  $\mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)}$  and  $\mathbf{b}_{2(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)}$  denote the least squares estimates for  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$ , respectively. If we replace  $\mathbf{X}_2$  by  $\mathbf{Z} = \mathbf{X}_2\mathbf{P}$  where  $\mathbf{P}$  is a regular matrix then

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\mathbf{P}\mathbf{P}^{-1}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}\boldsymbol{\varphi}_2 + \boldsymbol{\varepsilon} .$$

and it is easily seen that  $\mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} = \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{Z})}$  and  $\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2 = \mathbf{Z}\widehat{\boldsymbol{\varphi}}_2$ . Consider now a regression model that includes among its regressors two fixed effects and its interactions. The model can be represented in matrix terms as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}_1\boldsymbol{\alpha}_1 + \mathbf{D}_2\boldsymbol{\alpha}_2 + \mathbf{D}_3\boldsymbol{\gamma} + \boldsymbol{\varepsilon} ,$$

where the design matrices  $\mathbf{D}_1$  and  $\mathbf{D}_2$  account for the fixed effects and  $\mathbf{D}_3$  accounts for the interaction effect. We assume that superfluous columns have been removed from the design matrices to allow for the identification of all coefficients. Different parametrizations of the above model can be found by multiplying the design matrices by a regular transformation matrix. In the particular case of the interaction term we know that if  $\mathbf{D}_3$  is multiplied by any regular matrix  $\mathbf{P}$  then  $\mathbf{D}_3\widehat{\boldsymbol{\gamma}}$  [the estimate of  $\lambda_{ij}$  in equations (4a) and (5) in the main text] will remain the same.

## Appendix B: Equivalence of Coefficients

Consider again the regression containing two sets of regressors,  $\mathbf{X}_1$  and  $\mathbf{X}_2$ ,

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} , \tag{B.1}$$

and the alternative regression model

$$\begin{aligned} \mathbf{Y} - \mathbf{X}_2 \mathbf{b}_{2(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} \\ \mathbf{Y}^* &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} , \end{aligned} \tag{B.2}$$

where we replaced  $\boldsymbol{\beta}_2$  by its least-squares solution ( $\mathbf{b}_{2(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)}$ ) and rearranged the terms. Since  $\boldsymbol{\beta}_2$  is replaced by its optimal value, the least squares solution obtained for  $\boldsymbol{\beta}_1$  from (4) will be the same as that obtained from (B.1). This means that we can write

$$\begin{aligned} \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} &= (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{Y}^* \\ \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} &= (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 (\mathbf{Y} - \mathbf{X}_2 \mathbf{b}_{2(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)}) \\ \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} &= (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{Y} - (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \mathbf{b}_{2(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} \\ \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} &= \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1)} - (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \boldsymbol{\Gamma} \\ \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)} &= \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1)} - \mathbf{b}_{(\boldsymbol{\Gamma}, \mathbf{X}_1)} \\ \mathbf{b}_{(\boldsymbol{\Gamma}, \mathbf{X}_1)} &= \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1)} - \mathbf{b}_{1(\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2)}, \end{aligned}$$

what constitutes the well known formula for omitted-variable bias and shows the equivalence between the coefficients of the regressions. Recalling the regression model with two fixed effects and interaction,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}_1\boldsymbol{\alpha}_1 + \mathbf{D}_2\boldsymbol{\alpha}_2 + \mathbf{D}_3\boldsymbol{\gamma} + \boldsymbol{\varepsilon} .$$

we can let  $\mathbf{X}_1 = \begin{bmatrix} \mathbf{X} & \mathbf{D}_1 & \mathbf{D}_2 \end{bmatrix}$ ,  $\mathbf{X}_2 = \mathbf{D}_3$  and  $\boldsymbol{\Gamma} = \mathbf{D}_3\hat{\boldsymbol{\gamma}}$  to immediately conclude that the difference between the estimated coefficients of the regression with and without the interaction term is the least squares coefficient of a regression between the vector of estimated fixed effects ( $\mathbf{D}_3\hat{\boldsymbol{\gamma}}$ ) and the variables in  $\mathbf{X}_1$ . Gelbach (2009) shows that the asymptotic t-tests of this regression can be interpreted as an extended version of a Hausman test.

Table 1: **Descriptive Statistics**

	Weighted Sample		Unweighted Sample	
	Mean	S.D.	Mean	S.D.
Worker Characteristics	N=5,245,296		N=1,005,886	
Real Hourly Wage (euros)	4.579	2.939	4.443	2.792
Female	0.440	0.496	0.449	0.497
Age	37.514	10.978	36.978	11.277
Tenure	10.448	9.321	9.370	8.949
School1 (years=0)	0.004	0.067	0.006	0.075
School2 (0<years<=2)	0.010	0.097	0.011	0.101
School3 (2<years<=4)	0.416	0.493	0.406	0.480
School4 (4<years<=6)	0.292	0.455	0.286	0.438
School5 (6<years<=9)	0.139	0.346	0.146	0.341
School6 (9<years<=12)	0.099	0.299	0.101	0.294
School7 (12<years<=15)	0.012	0.110	0.013	0.107
School8 (years>=16)	0.028	0.165	0.032	0.174
Firm Characteristics	N=5,245,296		N=55,573	
Private	0.835	0.371	0.981	0.129
Public	0.016	0.126	0.001	0.023
Foreign	0.149	0.126	0.018	0.127
Size	249.502	502.691	15.716	57.483
Location Characteristics I	N=5,245,296		N=7,260	
Specialization (counts)	93.908	183.835	5.756	23.453
Specialization (employment)	1,651.609	2,731,070	93.563	345.252
Location Characteristics II	N=5,245,296		N=272	
Urbanization	28,321.860	37,686.640	5,788.414	16,342.230
Area (squared kms)	257.665	205.955	326.656	284.265

Table 2: **Regression Results with Counts**

	(1)	(2)	(3)	(4)	(5)	(6)
Specialisation (counts)	0.01002 (37.9)	0.01092 (41.2)	0.01404 (68.1)	0.00904 (29.7)	0.00411 (15.2)	0.00393 (13.8)
Area	-	-0.02752 (-57.9)	-0.01119 (-31.8)	-	-	-
Female	-	-	-0.23092 (-327.6)	-0.23449 (-336.8)	-	-
Age	-	-	0.02656 (154.0)	0.02588 (152.5)	-	-
Age Squared	-	-	-0.00023 (-99.1)	-0.00026 (-100.9)	-	-
Tenure	-	-	0.01354 (140.1)	0.01321 (139.3)	0.00892 (86.7)	-0.01632 (-0.4)
Tenure Squared	-	-	-0.00023 (-72.5)	-0.00026 (-70.1)	-0.00028 (-114.6)	-0.00029 (-104.4)
School1	-	-	-0.29675 (-90.5)	-0.29458 (-93.8)	0.01102 (1.7)	0.01484 (1.9)
School2	-	-	-0.21922 (-89.7)	-0.21322 (-91.6)	0.01480 (5.0)	0.01621 (4.5)
School3	-	-	-0.11792 (-165.2)	-0.11325 (-163.0)	0.00494 (4.8)	0.00828 (5.8)
School5	-	-	0.15083 (147.4)	0.13807 (137.4)	0.02217 (16.1)	0.01885 (10.1)
School6	-	-	0.29442 (222.1)	0.27743 (211.8)	0.06630 (29.5)	0.05876 (20.8)
School7	-	-	0.74844 (214.0)	0.73012 (211.4)	0.17218 (30.1)	0.14450 (21.6)
School8	-	-	0.89443 (373.9)	0.87192 (364.8)	0.29048 (49.2)	0.24234 (34.7)
Public	-	-	0.08966 (36.7)	0.09374 (39.7)	-0.04584 (-27.7)	-0.04734 (-27.5)
Foreign	-	-	0.13153 (136.6)	0.11548 (121.3)	0.00939 (11.5)	0.00895 (10.5)
Firm Size	-	-	0.00006 (79.0)	0.00006 (75.6)	0.00004 (52.5)	0.00004 (51.1)
Urbanization	-	-	-	0.03061 (19.2)	0.00783 (12.3)	0.00849 (12.5)
Fixed Effects						
Year	Yes [11]	Yes [11]	Yes [11]	Yes [11]	Yes [11]	Yes [11]
Industry	Yes [105]	Yes [105]	Yes [105]	Yes [105]	No	No
Location	No	No	No	Yes [272]	No	No
Worker	No	No	No	No	Yes [1,005,886]	No
Establishment	No	No	No	No	Yes [64,883]	No
Worker-Establishment	No	No	No	No	No	Yes [1,451,805]
$R^2$	34.1%	34.4%	61.2%	63.1%	93.7%	94.9%
$N$	5,245,296	5,245,296	5,245,296	5,245,296	5,245,296	5,245,296

Note: t-statistics associated with worker-level cluster-robust standard errors in round brackets; number of fixed effects in square brackets.

Table 3: Regression Results with Employment

	(1)	(2)	(3)	(4)	(5)	(6)
Specialisation (employment)	0.03548 (128.5)	0.03649 (132.2)	0.02463 (114.4)	0.02383 (84.9)	0.01150 (37.8)	0.01147 (35.5)
Area	-	-0.02962 (-62.5)	-0.01208 (-34.4)	-	-	-
Female	-	-	-0.23025 (-328.2)	-0.23452 (-337.6)	-	-
Age	-	-	0.02649 (154.1)	0.02579 (152.3)	-	-
Age Squared	-	-	-0.00022 (-98.7)	-0.00022 (-100.5)	-	-
Tenure	-	-	0.01331 (138.2)	0.01304 (137.9)	0.00885 (86.0)	-0.01699 (-0.4)
Tenure Squared	-	-	-0.00023 (-72.6)	-0.00022 (-70.1)	-0.00028 (-113.5)	-0.00028 (-103.2)
School1	-	-	-0.29822 (-91.4)	-0.29544 (-94.5)	0.01047 (1.6)	0.01411 (1.8)
School2	-	-	-0.22119 (-90.7)	-0.21463 (-92.4)	0.01460 (5.0)	0.01596 (4.4)
School3	-	-	-0.11853 (-166.7)	-0.11339 (-163.6)	0.00488 (4.7)	0.00818 (5.7)
School5	-	-	0.15040 (147.7)	0.13767 (137.4)	0.02202 (16.0)	0.01867 (10.0)
School6	-	-	0.29342 (222.5)	0.27679 (211.9)	0.06595 (29.4)	0.05834 (20.7)
School7	-	-	0.74532 (214.0)	0.72766 (211.4)	0.17165 (30.0)	0.14385 (21.6)
School8	-	-	0.89061 (373.8)	0.86911 (364.6)	0.28983 (49.1)	0.24156 (34.6)
Public	-	-	0.08524 (35.1)	0.08798 (37.5)	-0.04584 (-27.7)	-0.04734 (-27.5)
Foreign	-	-	0.12651 (133.1)	0.11093 (117.4)	0.00873 (10.7)	0.00833 (9.8)
Firm Size	-	-	0.00004 (53.4)	0.00004 (52.1)	0.00004 (46.5)	0.00004 (45.3)
Urbanization	-	-	-	0.01694 (10.6)	0.00456 (7.2)	0.00506 (7.5)
Fixed Effects						
Year	Yes [11]	Yes [11]	Yes [11]	Yes [11]	Yes [11]	Yes [11]
Industry	Yes [105]	Yes [105]	Yes [105]	Yes [105]	No	No
Location	No	No	No	Yes [272]	No	No
Worker	No	No	No	No	Yes [1,005,886]	No
Establishment	No	No	No	No	Yes [64,883]	No
Worker-Establishment	No	No	No	No	No	Yes [1,451,805]
$R^2$	35.0%	35.3%	61.5%	63.3%	93.7%	94.9%
$N$	5,245,296	5,245,296	5,245,296	5,245,296	5,245,296	5,245,296

Note: t-statistics associated with worker-level cluster-robust standard errors in round brackets; number of fixed effects in square brackets.