An Empirical Assessment of Assortative Matching in the Labor Market

Rute Mendes
Gerard J. van den Berg
Maarten Lindeboom

September 2007

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

An Empirical Assessment of Assortative Matching in the Labor Market

Rute Mendes
Collegio Carlo Alberto
and Free University Amsterdam

Gerard J. van den Berg
Free University Amsterdam, IFAU-Uppsala,
Netspar, CEPR, INSEE-CREST, IFS and IZA

Maarten Lindeboom
Free University Amsterdam,
Tinbergen Institute, HEB-Bergen and IZA

ABSTRACT

An Empirical Assessment of
Assortative Matching in the Labor Market[*]

In labor markets with worker and firm heterogeneity, the matching between firms and workers may be assortative, meaning that the most productive workers and firms team up. We investigate this with longitudinal population-wide matched employer-employee data from Portugal. Using dynamic panel data methods, we quantify a firm-specific productivity term for each firm, and we relate this to the skill distribution of workers in the firm. We find that there is positive assortative matching, in particular among long-lived firms. Using skill-specific estimates of an index of search frictions, we find that the results can only to a small extent be explained by heterogeneity of search frictions across worker skill groups.

JEL Classification: J21, J24, D24, J63

Keywords: positive assortative matching, matched employer-employee data, productivity, skill, unobserved heterogeneity, sorting, fixed effects

Corresponding author:

Gerard J. van den Berg
Department of Economics
Free University of Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
E-mail: gjvdberg@xs4all.nl

1 Introduction

Recent research based on matched employer-employee databases has revealed that wages of similar workers within firms and productivities of similar firms within industries can be substantially dispersed. Moreover, a given worker may earn different wages at different firms. The classical framework under which all workers and firms are alike and can freely enter and exit the marketplace has difficulties to accommodate such findings. This has spurred the development of models that allow for search frictions and that consider wages and productivity as involving worker-specific, firm-specific, and match-specific components. Burdett and Coles (1999), Shimer and Smith (2000), and Shi (2001) are some examples in the theoretical literature. They present search models with two-sided heterogeneity and search frictions. A major aim of these studies is to investigate the way firms and workers sort themselves out when they form a match, and, in particular, whether the matching is positive assortative, meaning that the best firms team up with the best workers. Different equilibrium matching patterns are possible, depending on key assumptions like the transferability of utility, the logsupermodularity of the production function, and the commitment for a wage schedule. Which matching pattern actually arises in the labor markets is in the end an empirical question.

There are several reasons why it is important to know the actual matching pattern in the market. First, it allows us to test different economic models that predict distinct matching equilibrium patterns, and this gives insights into the realism of the assumptions on which the models rely. Second, it help us to understand the agents' behavior and the allocation in the labor market. Third, it is relevant for policy. In particular, labor market policies that target welfare of workers may have different effects depending on the existence and extent of assortative matching, because the latter determines the range of jobs available to a worker.

There are a few examples of empirical studies that provide evidence on assortative matching (Haltiwanger *et al.*, 1999, Van den Berg and Van Vuuren, 2003, and Abowd *et al.*, 2004). The latter study focuses on wages as the main outcome variable, while the other two focus on total firm output. Specifically, Abowd *et al.* (2004) work with a model that predicts that indices of worker and firm productivity are measurable from wages, and so assortative matching is directly manifested in a wage equation from which firm-specific and worker-specific effects can be deduced. They find no evidence for positive assortative matching (PAM), i.e., a positive correlation between indices of worker and firm productiv-

ity. The wage data approach is very flexible and takes account of a wide range of endogeneity issues. However, assignment models with coordination frictions and PAM, like the Shimer (2005) model, can generate positive or negative correlation between the firm and person effects as obtained from a wage equation. This calls for the analysis of productivity data.

Haltiwanger *et al.* (1999) focus on output data. They use a panel of long-lived firms and estimate the relationships between output, as measured by the natural logarithm of sales per worker, and workers' and firms' characteristics. The empirical specification imposes an additive linear production function, and hence assumes that workers' and firms' productivity inputs are perfectly substitutable. The results indicate that PAM may be an important phenomenon in the US labor market.[1] Van den Berg and Van Vuuren (2003) investigate the sign and the extent of assortative matching in the Danish labor market. As Haltiwanger *et al.* (1999), they take additive linear production functions and estimate these on a cross-section. They use the model estimates to construct a measure of firm-specific heterogeneity and relate this measure to workforce characteristics.[2] They conclude that PAM is a common phenomenon.

In the present paper, we empirically investigate the matching pattern between firms and workers using firm productivity data. We exploit that the data are longitudinal in both the worker and the firm dimension. The data cover the population of establishments in manufacturing and services in the private sector in Portugal. They contain annual detailed information on the workforce composition and firms' output over a fifteen year period. Using dynamic panel data methods, we quantify a firm-specific productivity term for each firm, and we relate this to the skill distribution of workers in the firm. Specifically, we estimate a general relationship between the output of the firm and its workforce composition, which we approximate using a translog specification. Estimating this translog approximation by fixed-effects methods, we are able to quantify the firm-specific productivity, and relate this to the skills of workers in the firm. We use the bootstrap to estimate standard errors.

Not all firms participate for the full fifteen years in our sample. Survival

---

[1]They estimate regression in levels and in first differences. In the first case, they find statistical significance for workforce characteristics. In the second case, where they allow for firm-specific fixed effects, the effect of the workforce characteristics disappears. This may hint at assortative matching. However, this is not directly addressed by the authors.

[2]More specifically, they take the coefficient of a regression of the fraction of low-skilled employees on firm-specific productivity as a measure of assortative matching. Note however, that they only had access to a single cross-section, which requires them to make strong assumptions in order to calculate the firm-specific effect.

may depend on workforce characteristics, and hence the results may be affected by selectivity of surviving firms. We therefore estimate models on panels of different length to check on the robustness of our findings. We also check whether heterogeneity of search frictions across worker skills can explain PAM. For this purpose, we use data on job transitions to compute an index of search frictions for the different skills.

As we use Portuguese data, we briefly discuss some distinctive features of the Portuguese labor market (see also OECD, 1996, 1997, 1998, Vieira *et al.*, 2005, Cardoso and Portela, 2005, and references in these sources). Compared to most EU and OECD countries, Portugal has low unemployment, low average labor productivity, and very high employment protection. Centralized collective bargaining covers most of the workforce. Private costs to geographic mobility are high. To some extent, the institutional constraints may hamper job transitions into better matches, and thus the degree of assortativeness may be particularly visible in the matches that occur upon workers' entry into the labor market. In this regard, it is relevant that youth unemployment in Portugal represented in the 1990s almost the double (or more) of the overall rate.[3] Among young unemployed, around 50% have never worked before, and the concentration of long-term unemployment is high: it varied between 40% and 20% from the mid-1980s to the late 1990s.[4] Furthermore, youth unemployment is higher for individuals with tertiary education than for those with lower education. Along with the expansion of higher education during the 1980s and early 1990s, unemployment increased for university graduates. This did not increase the likelihood of these individuals moving into non-university jobs (Cardoso, 2005). This may suggest that individuals take quite long time until they reach their first match but that these first matches are to some extent assortative.

The paper proceeds as follows. The next section presents the empirical approach, with two subsections. In Subsection 2.1 we describe the specification and estimation method used to quantify the firm-specific productivity. In Subsection 2.2 we present the measures for assortative matching and we give details about the sensitivity analysis of those measures. Section 3 presents the data and describes the construction of the variables. Results are reported in Section 4. In Section 5, we investigate whether search frictions heterogeneity may be behind

---

[3]For EU 15, in the 1990s, youth unemployment never reached the double of the general unemployment rate.

[4]This concentration of long-term unemployment among the young was also seen in other southern European countries. For the time period considered, this group of countries had the highest concentration in the EU 15.

our findings on PAM. Section 6 concludes.

2   Empirical approach

Our ultimate objective in this study is to analyze empirically the way the two *ex ante* heterogeneous sides of the labor market sort themselves out by matching. For this, we first need to know the specific productivity of workers and firms. However, these are not directly observable in the data. What we can observe is the production of a firm that operate with its workforce. To obtain a measure of the specific productivity of workers and firms, we need thus to disentangle the contribution of each of them on the observed production. Intuitively, two firms having the same workforce composition can produce different outputs if their management skills or structure are different. In the same way, two firms with common management abilities or structure may have different outputs if their workforces differ.

Our empirical strategy to deal with this problem is based on a framework where heterogeneity in labor inputs and firm-specific productivity play both a central role. In particular, we assume that the observed output in firm i is the result of a relationship between different qualities of labor and firm-specific (unobserved) heterogeneity. The latter is derived from fixed-effects estimates of a translog approximation, which can be seen as a second order approximation to a generalized production function. Details about this translog approximation and the estimation of firm-specific productivity are presented in Subsection 2.1.

The productivity of the labor input of a firm is measured by the proportion of high-skilled labor in the total labor input of the firm. Assortative matching is then assessed by measures of association between the estimated firm-specific productivity and the proportion of high quality labor within the firm.[5] These measures are discussed in Subsection 2.2.

2.1   Estimation of the firm-specific productivity

We postulate that the output Y_i observed in firm i is determined by a functional relationship between various different observed qualities of labor and the firm-specific unobserved heterogeneity:

---

[5]Assortative matching is often measured by the correlation between job complexity and worker skill. We take the firm (instead of job) as the relevant unit, and so the workers skills are summarized to the firm level.

$$Y_i = F(L_{ij})v_i \tag{1}$$

where L_{ij} denotes the quality of labor aggregate offered by workers with skill level j in firm i, v_i stands for the firm-specific unobserved productive skill and $F(.)$ represents a medium or long term production objective which is predetermined with respect to the actual observed value Y_i.

We assume that in firm i, within the quality of labor aggregate of skill j, L_{ij}, workers with different demographic characteristics are perfectly substitutable inputs with potentially different marginal products. For example, in case we distinguished workers only by gender, L_{ij} would be defined as:

$$L_{ij} = M_{ij} + \phi_F F_{ij} \tag{2}$$

where M_{ij} and F_{ij} are, respectively, the number of males and females with skill j, and ϕ_F is the marginal productivity of women relative men, within skill j.

For our analysis, we define quality of labor as:

$$\begin{aligned} L_{ij} &= MA_1S_{ij} + \phi_{A_2}^j MA_2S_{ij} + \phi_{A_3}^j MA_3S_{ij} + \\ &+ \phi_F^j FA_1S_{ij} + \phi_F^j \phi_{A_2}^j FA_2S_{ij} + \phi_F^j \phi_{A_3}^j FA_3S_{ij} \end{aligned} \tag{3}$$

where MA_hS_{ij} (FA_hS_{ij}) stands for the proportion of hours offered by males (females) in age group h and schooling j, and ϕ_K^j are parameters to be estimated. ϕ_F^j is the marginal productivity of women relative to men within schooling level j, $\phi_{A_2}^j$ is the marginal productivity of workers in the second age group relative to those in the first age group within schooling level j, and so forth.[6] This specification is similar to Hellerstein and Neumark (1995) and Hellerstein *et al.* (1999). However, we allow here for greater flexibility in terms of marginal productivity differential between different demographic groups.

For $F(.)$ we specify the translog form, which may generally be viewed as a second order approximation to a generalized production function. We take this specification because we want to have $F(.)$ as flexible as possible in order to obtain the best possible estimate of the firm-specific productivity.[7] The flexibility of the translog form is well known and it is the reason why it is increasingly and widely

---

[6]Note that the definition of L implies that productivity differentials between groups are indicated by ϕ^j being estimated to be significantly different from one (rather than zero). For example, an estimate of 0.8 for ϕ_F^j would mean that women in schooling level j are on average 20% less productive than men with the same schooling level.

[7]Conventional Cobb-Douglas functions were also estimated. However the results for the translog approximation reject the Cobb-Douglas specification.

employed in empirical work. Basically, this form permits a greater variety of substitution patterns among input factors than functions based on constant and equal elasticities of substitution among all pairs of inputs. See Christensen *et al.* (1973).

With three levels of labor skills ($j = 1, 2, 3$, where 1 and 3 stand for the lowest and highest skill level, respectively), the translog specification yields the following relationship:

$$ln(Y_{it}) = \sum_{j=1}^{3} \gamma_j ln(L_{jit}) + 0.5 \sum_{j=1}^{3} \sum_{k=1}^{3} \gamma_{jk} ln(L_{jit}) ln(L_{kit}) + c_i + \epsilon_{it} \qquad (4)$$

with $\gamma_{jk} = \gamma_{kj}$ *for all* (j, k)

where Y_{it} denotes the real sales per worked hour of firm $i = 1, 2, ..., N$ in year $t = 1, 2, ..., T$, $c_i = ln(v_i)$ stands for time constant characteristics of firm i and ϵ_{it} are idiosyncratic errors, assumed to be independent and identically distributed over time and firms, with zero mean and equal variance. L_{jit}, previously defined, are time varying regressors assumed to be strictly exogenous, conditional on the unobserved firm-specific effect c_i.[8] The parameters γ_j and ϕ^j are to be estimated common to every firm within each industry. Two control variables were added to (4): the size of the workforce and an indicator for single establishment firms.

This specification incorporates the idea that two mediocre workers do not combine to make a good one, meaning that quantity of one skill type of labor cannot be completely substituted by any combination of the other skill types. Under this framework, in order to produce, a firm needs a workforce that combines the three labor skills.[9]

Under this specification, labor quality is allowed to differ across skill levels, and across age and gender within skill categories. Furthermore, skills are complementary and workers with different ages and gender are substitutable but allowed to have different marginal productivities. Notice that we allow the productivity differential between age groups to vary across skill.[10]

---

[8]Later we relax this assumption of strict exogeneity. See description of the robustness test and respective results in the end of Section 4.1.

[9]This is the reason for considering no more than three types of labor skills. In the estimation we can only use firms for which we have no zero observations on any of the skills. A more detailed skill classification would lead to a large number of different skills and would force us to discard a large number of firms from our data.

[10]This function is supermodular in c and L_3. This is sufficient for PAM in a frictionless environment. If strong search frictions hold, supermodularity is only necessary and the sufficient condition for PAM is then logsupermodularity. Notice however that the specified function does not have this property, since $(ln f_x)_{xy} = 0$.

Equation (4) is estimated by nonlinear least squares (NLS). We do not want to impose that c_i and L_{jit} are uncorrelated and therefore use the within transformation of (4):

$$
\begin{aligned}
ln(Y_{it}) - \overline{ln(Y_i)} \ = \ & \sum_{j=1}^{3} \gamma_j \Big[ln(L_{jit}) - \overline{ln(L_{ji})} \Big] + \\
& + 0.5 \sum_{j=1}^{3} \sum_{k=1}^{3} \gamma_{jk} \Big[ln(L_{jit}) ln(L_{kit}) - \overline{ln(L_{ji}) ln(L_{ki})} \Big] + \\
& + \epsilon_{it} - \overline{\epsilon_i}
\end{aligned}
\tag{5}
$$

The within or fixed-effects estimation procedure is convenient for our purposes. First, our interest is mainly in c_i as a measure of unobserved firm-specific productivity. We can retrieve an estimate of this (nuisance) parameter after having estimated the model. Second, the condition $E\{c_i \mid L_{jit}\} = 0$ will be violated in the case of assortative matching. Finally, our sample is possibly subject to non-random attrition and under some assumptions we can still consistently estimate the model parameters with the fixed-effect method. Lindeboom *et al.* (2002) provided a more detailed analysis of the possible sources of endogenous attrition and the advantages of the fixed-effects method.

Having estimated the parameters γ_j and ϕ_k^j, we proceed with the estimation of c_i as

$$
\hat{c}_i = \overline{ln(Y_i)} - \sum_{j=1}^{3} \widehat{\gamma_j} \overline{ln(\widehat{L_{ji}})} - 0.5 \sum_{j=1}^{3} \sum_{k=1}^{3} \widehat{\gamma_{jk}} \overline{ln(\widehat{L_{ji}}) ln(\widehat{L_{ki}})}
\tag{6}
$$

The c_i term captures stable key features such as the organizational and managerial skill, capital and technology endowment, to the extent that they are time-constant over the observed time interval. We use it as the measure of the firm-specific productivity skill.

The estimates $\hat{c}_i$ are unbiased for any sample size. However, since they are time-series averages, their variance only tends to zero as T tends to infinity. Therefore, $\hat{c}_i$ are consistent if T is sufficiently large.

2.2 Measures for assortative matching

We next focus on the relation between the firm-specific effect and the workforce skill. A positive association between these two indicates the presence of PAM. The workforce skill of a specific firm is measured by the average (over time) fraction

of highly-educated workers in that firm ($\overline{L_{3i}}$). We use three association measures: the traditional correlation coefficient, the rank correlation coefficient and the regression coefficient of a regression of the firm fixed effect on the proportion of highly educated workers in the firm. As with the estimation of the production function, these measures of assortative matching are estimated for each industry separately.

The traditional correlation coefficient identifies the sign and the degree of (im)perfection of the linear relation between the firm-specific productivity and the proportion of high-educated workers in the firm. A value of (+)-1 would indicate, for our data, $P(c_i = a + b\overline{L_{3i}}) = 1$, for some constant a and a constant $b(>) < 0$. Since this correlation coefficient is dimensionless, we also present the estimates for b. This regression coefficient provides us a quantitative interpretation, giving an estimate of the average change of firm-specific productivity in response to an one percentage point change in the proportion of highly-educated workers.

The rank correlation coefficient is an association measure that uses the ranks instead of the actual values of the variables. As the traditional correlation coefficient, it varies between -1 and 1, but it is less sensitive to extreme values and it captures broader forms of association than a linear relation. With the rank correlation coefficient we get thus a measure of the degree of ordering of the two heterogeneous sides of the labor market. A value of 1 would mean that the highest-c_i firm matches with the highest-$\overline{L_{3i}}$ workforce, and the second higher-c_i firm teams up with the second higher-$\overline{L_{3i}}$ workforce, and so on; but the underlying relation between c_i and $\overline{L_{3i}}$ is not required to be linear.

Not all firms in our sample participate in all waves and it is likely that the firm survival rate is related to the firm-specific unobserved productivity c_i. In the case of assortative matching the c_i will also be related to the skill level of the workforce. A consequence of this is that our association measures may change as the panel ages. Haltiwanger *et al.* (2007) investigate the adjustment of the workforce over the ageing of the firm and they find it to be consistent with both selection and learning. On the one hand, firms with better unobserved characteristics and better matches are found to have higher survival probabilities. On the other hand, new firms adjust their workforce composition towards the workforce of the longer existing firms. Under these mechanisms, a panel of long lived firms is likely to have firms with a higher average specific quality and with the best matches than one panel of shorter time interval. In this case, we would find stronger evidence for assortative matching in the former panel than in the latter.

To evaluate the impact of this dynamics, we compute our measures for assortative matching using four panels with different minimum number of observations

per firm: two, six, nine, twelve.

3   The "Quadros de Pessoal" data

Our analyses are based on *Quadros de Pessoal*, a longitudinal matched employer-employee data set gathered by the Portuguese Ministry of Labor and Solidarity (see also, e.g., Vieira *et al.*, 2005, and Cardoso and Portela, 2005, for descriptions and analyses of this dataset). The data are collected through firms' annual reports. Firms are legally obliged to do this. The public administration and domestic service are not covered and the coverage of the agricultural sector is low. On the other hand, the manufacturing and private services sector is almost fully covered.

The firm data contain information on sales, employment, detailed industry, location, legal setting and ownership. The worker information includes gender, age, occupation, schooling, tenure, mechanism of wage bargaining, normal and overtime hours of work and gross monthly earnings (split into several components). The reported data report on the status of the workers in a reference week (March up to 1993, October since 1994).

An identification code is assigned to every firm at the time it enters the data set for the first time. The Ministry performs various checks to guarantee that the firm code does not change. The identification code of the worker is a transformation of the worker's social security number. Based on these identification numbers, one can match workers and firms, and follow both over time. The firm and worker data cover the period 1986-2000.

For each worker, we compute the monthly total (normal plus extra) hours of work. Workers are then grouped according to the three demographic characteristics that formed the labor types in our specification: gender, age and schooling. We consider three age groups – under 30, between 30 and 50, and above 50 – and three schooling categories[11] – low education (less than 4 years), medium education (6 or 9 years) and high education (more than 10 years [12]). The grouping leads to 18 types of workers within the firm. The proportions of hours worked by the different types of workers within the firm are our measure of labor inputs.

There is some item non-response in the reported age and schooling (an average of 6% in each year[13]). We discarded these individual observations from our

---

[11] The reported education is the highest completed school degree of the worker.

[12] High education category includes both the secondary school and higher education.

[13] These percentages include actual non-response and schooling reported as 'Ignored' or 'Other'.

sample. In these cases, the workforce composition within a specific firm is based on the remaining workers of whom we observe all characteristics. We implicitly assume that the item non-response is random. This may introduce some systematic measurement error on L_{jit}. An alternative for this would be to eliminate every firm with at least one worker not fully characterized. This would however, result in the elimination of a large number of firms.

Beyond these workforce characteristics, we consider the following firm's variables: sales, total number of hours worked within the firm, industry, number of establishments, size of workforce and geographical location.[14]

The reported sales volume relates to the previous year and is measured in thousands of *escudos* (a thousand *escudos* is around 5 euros). We use real sales (prices 1997) per worked hour as the measure of output of the firm. In this way, we control for differences in working times between various labor types (notably males and females). To obtain sales per hour we compute the ratio of the volume of monthly sales[15] and the sum of monthly hours of all workers in the firm.[16]

Our dependent variable is real sales (per hour worked) and we therefore only keep firms in Manufacturing, Construction and Trade. In our translog function workers in different schooling categories are imperfectly substitutable inputs. A consequence of this is that output is zero in any firm that does not have workers in one of the three schooling categories that we consider. We therefore had to eliminate firms who had either no low-, medium or high-educated workers. We also excluded the observations in the top and bottom 1% of the sales distribution. Our within estimator requires at least 2 observations for a given firm. With all these selection criteria we end up with a panel of 39 543 firms and 204 537 firm-year observations. Around 45% of the firms are observed for 5 or more years, slightly more than 20% for at least 8 years and 6.5% for the full observation period. The participation pattern of firms in the panel is given in Table 1.

Descriptive statistics for the whole panel and individual industries are presented in Table 2.

In terms of the distribution of schooling, firms in our sample tend to employ

---

[14]The hours worked in the firm and the size of the workforce are not affected by the selection that we make when we deal with the item non-response. The variable hours worked is computed before that selection and the size of the workforce is directly reported by the firm.

[15]We assume monthly sales to equal the annual sales divided by 12.

[16]Until 1993, reported workers are those employed in the firm in March. From 1994 onwards, the reference period was October. We use the characteristics of the workforce of March of year t to explain sales of year $t-1$, for $t \leq 1993$ and the workforce of October of t to explain sales of t, if $t \geq 1994$. So, we do not use the data on sales of 1993. This reduces the maximum length of the panel to 13 waves.

Number of years	Absolute frequency	Relative frequency	Relative cumul. freq.
2	9679	24.48	24.48
3	6621	16.74	41.22
4	4950	12.52	53.74
5	3913	9.90	63.63
6	3150	7.97	71.60
7	2598	6.57	78.17
8	1833	4.64	82.81
9	1638	4.14	86.95
10	1296	3.28	90.23
11	1297	3.28	93.51
12	1152	2.91	96.42
13	1416	3.58	100.00
Total	39543	100.00	

Table 1: Participation pattern of firms in the panel

more low educated workers than any others: low educated workers constitute around 45% of the workforce and high educated workers represent about 19%. In Trade, high education takes a relatively high share (25%).

On average firms employ more men than women. As expected, this difference is especially large in Construction. The gender differences vary with education. The shares of men and women are closer among high educated workers, especially in Trade.

Also the distribution of age changes with schooling level: workers aged 30-50 dominate among the low educated, and medium and high education are primarily younger workers. Trade tends to employ more older workers than the average firm in the panel.

Furthermore, the average firm in our panel employs around 57 workers. The smaller firms are those in the Trade and the largest those in Manufacturing. Regarding the sales, firms have on average a volume of 8 thousands *escudos* per each hour of work. Firms in Manufacturing have the lowest average sales. On the other hand, Trade is the sector with the largest sales and the smallest workforce.

It is important to notice that the data show a great amount of heterogeneity in the workforce composition. The large standard deviations presented in Table 2 indicate that the workforce composition tends to be substantially different across firms. We found evidence for heterogeneity on several other dimensions. Even after removing year and industry means, we found large variation of the skill composition across firms. For example, the distribution of the proportion of low educated workers still presents an interquartile range of 31 percentage points

Variable	All Inds		Manufact.		Construct.		Trade	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Schooling level 1	45.2	23.3	51.2	23.0	58.2	23.3	37.0	20.4
Males, Under 30	7.4	11.1	8.3	11.2	15.2	14.8	5.0	9.0
Males, 30-50	16.5	14.8	16.9	14.4	29.1	15.5	13.4	13.6
Males, Above 50	7.0	9.5	6.8	8.7	12.1	12.2	6.1	9.2
Females, Under 30	3.8	8.8	6.3	11.2	0.3	1.7	2.3	6.1
Females, 30-50	8.3	11.6	10.8	12.9	1.0	3.3	7.5	10.8
Females, Above 50	2.2	5.1	2.1	4.4	0.5	2.3	2.8	6.1
Schooling level 2	35.9	18.9	35.5	19.9	27.7	18.9	38.1	17.4
Males, Under 30	13.0	13.6	13.3	13.9	14.9	14.6	12.3	13.1
Males, 30-50	8.6	10.2	7.2	8.8	8.7	10.2	9.8	11.2
Males, Above 50	1.4	3.9	1.1	2.9	1.1	3.3	1.8	4.7
Females, Under 30	7.5	12.5	9.5	14.3	1.2	3.6	7.1	11.4
Females, 30-50	4.9	7.9	4.2	6.5	1.5	3.8	6.4	9.3
Females, Above 50	0.5	2.4	0.3	1.5	0.2	1.3	0.7	3.1
Schooling level 3	18.8	15.2	13.3	12.1	14.1	12.0	24.9	16.1
Males, Under 30	4.6	7.6	3.5	6.0	4.2	6.6	5.8	8.9
Males, 30-50	4.9	7.6	3.8	5.9	4.4	6.5	6.0	8.9
Males, Above 50	0.9	3.1	0.7	2.2	0.7	2.5	1.1	3.8
Females, Under 30	4.8	7.9	3.1	5.2	2.9	5.4	6.8	9.8
Females, 30-50	3.4	6.4	2.1	4.2	1.9	4.3	4.8	8.0
Females, Above 50	0.2	1.8	0.1	0.9	0.1	0.9	0.4	2.4
Workforce size	57.4	170.7	84.5	203.4	62.8	188.6	31.4	123.4
Real sales per hour	8.2	10.0	5.0	6.6	5.9	8.5	11.6	11.6
No observations:	204 537		87 786		20 684		96 067	

Table 2: Descriptive statistics

Notes: The demographic groups within the workforce are measured in percentage.

(p.p.) and that range is of 15 p.p. for the proportion of high educated workers. We also observe substantial variation in the workforce composition over time (within the panel, which means, over firm's ageing), as we may see from Table 3.[17] We present here changes for 4-, 8- and 12-year time horizons.

In general terms, the share of low educated workers within the firms tend to decrease substantially over time, while the shares of the other skill groups both increase. Looking at the interquartile ranges and variation coefficients, we see

---

[17]Measures for age groups were also computed. Though we do not report them here, we also found substantial heterogeneity for age composition.

	Schooling		
	level 1	level 2	level 3
four-years changes: $L_{ij4} - L_{ij1}$			
mean	-5.1	3.9	1.2
$P_{75} - P_{25}$	16	19	10
variation coef.	-3.3	4.5	10
eight-years changes: $L_{ij8} - L_{ij1}$			
mean	-11.1	8.8	2.3
$P_{75} - P_{25}$	22	22	12
variation coef.	-1.7	2.1	5.6
twelve-years changes: $L_{ij12} - L_{ij1}$			
mean	-15.1	10.5	4.6
$P_{75} - P_{25}$	20	20	9
variation coef.	-1.1	1.7	2.8

Table 3: Descriptive statistics on time changes of workforce composition

Notes: The demographic groups within the workforce are measured in percentage. Variation coefficient is defined as the ratio between mean and standard deviation.

that time changes exhibit also large variation across firms. We also find that this heterogeneity across firms in terms of adjustments of the workforce composition tends to decrease over firm's ageing. For example, the variation coefficient of the time changes of the highest skill decreases from 10 to 2.8 when we move from a 4- to 12-years time horizon.

We estimate the firm-specific productivity and the measures of assortative matching on panels of different length. We do this to check the robustness of our findings. We use panels in which firms are observed for at least two, six, nine and twelve years in the period 1986-2000.

4 Results

4.1 Estimates for the translog specification

Here we present the estimates of the translog form described in Subsection 2.1. The models are estimated on the four different panels. Since our interest on these estimates lies on their use for quantifying the firm-specific productivity, here we only report the estimates obtained for the largest panel (Table 4). The parameters are estimated by industry using nonlinear least squares on equation (5). The table also provides information on statistical tests. For labor quality variables, we test

whether the coefficient is significantly different from zero. However, given our definition of labor quality aggregates, the test of interest for the demographic characteristics is whether the coefficient is significantly different from one. We indicate for each variable whether the relevant hypothesis is rejected.

Since the results for the parameters γ of the translog function are not directly interpretable, here we focus on the estimates of marginal productivities of the demographic groups (ϕ).

The coefficients for females suggest that women are more productive in Manufacturing. This is also the case for highly educated female workers in the Trade and Construction sector.

The estimated age profiles vary with labor quality. They suggest that productivity increases considerably with age among low educated workers. However, among medium- and high-skilled workers, individuals aged 30-50 are the most productive. The differential between age groups is much lower for high educated workers. The Construction sector is an exception. Here the young are the most productive among the low educated workers.

Finally, coefficients for workforce size indicate that large firms are on average less productive.

For the various reasons previously discussed, we estimate the productivity regression 4 using the fixed-effects estimator. However, this estimator has the drawback of requiring strict exogeneity of the regressors. In our case, this means that decisions on the workforce composition (L_{ij}) can not affected by past or contemporaneous shocks on productivity or demand (ϵ_i).

In general, this is a quite strong assumption. Though, it may be not very strong in the context of the Portuguese labor market, which presents the strictest regulation on employment protection within OECD. The strict regulation is very likely to act as a barrier to feedbacks from demand or productivity shocks to workforce composition.

As a robustness test, we estimate the productivity regression assuming workforce composition to be predetermined instead of strict exogenous. This allows for correlation between the regressors and past shocks. Under this assumption, we need to use the difference GMM estimator (Arellano and Bond, 1991). Notice that this is an estimator for linear models and to adjust to non-linear case would be quite cumbersome. Therefore, for this part of the analysis we use the linearized version of the translog specification.[18]

The results obtained with difference GMM turn out to be mediocre, as the

---

[18]The linearization implies that, within each type of labor L_{ij}, demographic groups are not distinguished.

various sets of instruments we used were all rejected by the Hansen test. Since the variables were found to be somewhat persistent, we proceed using system GMM (Arellano and Bover, 1995, Blundell and Bond, 1999). The extended sets of instruments are again rejected. In a third stage, we estimate a dynamic version of the productivity regression, allowing the shocks ϵ_i to be serially correlated. We estimate a large variety of specifications with different sets of instruments but results were not satisfying for any of the cases. Even if we find some sets of instruments to be valid and errors to be serially uncorrelated, the explanatory power of the linear regression productivity is always very poor, with workforce composition being basically insignificant.[19]

In few words, though these estimators allow for more flexibility in terms of the exogeneity of the regressors, they do not produce satisfying results in our setting.[20] We proceed our analysis based on the fixed-effects estimates of the translog specification previously presented.

4.2   Measures for assortative matching

We use the estimates of the parameters γ and ϕ of the translog approximation, to calculate the firm-specific unobserved productivity c_i. Next we relate the estimate $\hat{c}_i$ to a skill measure of the workforce. The skill level of workforce is measured as the time average of the share of high-educated workers in the firm.

In this subsection we present three different measures for the association between the skill of workforce and the firm-specific productivity: traditional correlation coefficient, rank correlation coefficient and regression coefficient (see Section 2.2). The association measures were computed with four different panels. This was done in order to check on the robustness of our findings. A high positive association is an indication of strong PAM. The results are presented in Table 5. Panel 1, 2, 3 and 4 are panels that include only firms that are observed for at least 2, 6, 9 and 12 years, respectively. The dimension of each panel is presented in the bottom part of the table. Together with the point estimates, we present information on their accuracy, obtained through bootstrapping.

A first look on the Table 5 gives a very clear qualitative conclusion. Our results provide evidence for PAM in the Portuguese labor market. The three

---

[19]Results for the several specifications and persistence of the series are available upon request.

[20]A good performance of these estimators is well-established in the estimation of production functions, which is slightly different from our case. Note that our objective is primarily to estimate the unobserved firm-specific productivity by correcting for observed workforce characteristics in a flexible way, while usual estimation of production functions use series of labor and capital inputs in order to estimate elasticities and returns to scale of those inputs.

measures are significantly positive for all sectors in almost every panel.

The size of the association measures suggests that PAM is also quantitatively important. There are however, some differences across sectors. All measures and panels provide consistent evidence that Construction is the sector in which PAM is stronger, followed by Manufacturing and then Trade. In Construction, the correlation between firm's and worker's skills is estimated to be close to 0.37 in the panel of the shortest length. Notice that the traditional and the rank correlation coefficients are very close being respectively of 0.371 and 0.365. This similarity is generally observed in the other sectors and remaining panels. This indicates that workers and firms in each sector tend to match following an ordering pattern that is almost fully described by a linear relation. In Manufacturing, these correlations are very close to 0.30 and in Trade close to 0.15.

The third panel of the table reports the estimates of the regression coefficient of the firm-specific productivity on the proportion of high-educated workers in the firm. The values are generally large and significant. These results indicate that the magnitude of changes in the workforce composition on firm-specific productivity is large.

As expected, we find higher association measures when we increase the length of the panel, i.e. when we condition on firm survival up to a certain number of years. The surviving firms have on average a higher firm-specific productivity and a higher skill level of the workforce. This holds for every sector. A notable increase is observed for the construction sector: the correlation coefficient between firm's and workers' skills increases from 0.37 in Panel 1 to 0.63 in Panel 4. The regression coefficient of this same industry goes up from 3.05 to 5.35. In Manufacturing, the change in the measures follows the same features. The correlation coefficient ranges between 0.30 and 0.37, taking intermediate values in Panels 2 and 3. Similarly, the regression coefficient increases from 2.23 to 2.52.

These measures of assortative matching were also computed taking the lowest education category as reference instead of the highest and they are presented in Table 6. The results are consistent with the PAM. In particular, the correlations between firm-specific productivity and the proportion of the lowest-educated workers are negative for all industries. Looking at the different panels, we find the same conclusion regarding the sensitivity of assortative matching to the length of the analyzed panel. In general, the measures become more and more negative as we increase the minimum number of observations per firm.

We also measured assortative matching taking all the analyzed industries together. We do not present the results here, but the sign of assortative matching for the whole sampled market is also positive and its extent is about an average

of the those obtained for each industry.

The results on the sensitivity of measures of assortative matching to the stability of sampled firms lend some credence to the hypothesis of learning and selection, a hypothesis already supported by previous empirical literature like Haltiwanger *et al.* (2007), as discussed in Section 2.2. It also means that the degree of PAM may depend on sample selection criteria, like firm survival.

5   Search frictions heterogeneity

The association measures tell us whether or not assortative matching is important. It is possible that PAM is driven by heterogeneity in the amount of search frictions across skill levels. All worker types may be attractive for all firm types, but the high-skill workers may simply be less constrained by search frictions than the low-skill workers. In this scenario, high-quality workers would reach the high-quality firms more easily, leading to the observation of PAM. Stated differently, if we do not find evidence for search frictions heterogeneity across skills then PAM is due to high-productivity firms (workers) not wanting to team up with low-productivity workers (firms).

Van den Berg and Van Vuuren (2003) have addressed this issue before. They call the two explanations for PAM "two-sided sorting" and "search friction heterogeneity". They distinguish between the two explanations by investigating whether sectors and regions where the PAM is high also have low search frictions. They find that sectors and regions with lower frictions display more PAM.

In this paper, we use data on job transitions to estimate an index of search frictions for the three skill levels within different submarkets. We then investigate whether the degrees of frictions tend to differ across skills and whether the magnitude of that inter-skill difference is correlated with the extent of assortative matching across submarkets.

Our index for search frictions is defined as the ratio between the probability of moving to another firm and the probability of leaving the labor force. Considering the history of each worker between period t and period $t + 1$, we define three possible outcomes: stay in the same firm, move to another firm of the private sector and leave the labor force of the private sector (unemployment, public sector, retirement,...).

The analysis is based on various short (three year long) panels of individual workers. Information on the first year is only used to select individuals who are employed for more than one year. We do this to ensure some homogeneity in the search environment faced by the sampled workers. The data on the second and

third year are then used to obtain the outcome of relevance.

For each worker we record information on schooling, age, gender and on location and industry of their employers. The schooling, age, gender and industry variables are defined as in the panel of firms. We keep only workers employed in Manufacturing, Trade or Construction. Location is aggregated to five large regions: North, Centre, Lisbon and Tagus Valley, Alentejo and Algarve, and a last one with the islands Azores and Madeira. Submarkets are defined by combinations of industry and region.

Tables 7, 8 and 9 report the results for the index of search frictions in each of the industries. For sake of simplicity, we discuss here the results for four pairs of years and three regions. The last column of the tables gives the inters-kill differential, computed as the difference between the indicator for high-skilled workers and low-skilled workers. This is our measure for search friction heterogeneity.

Table 7 shows that in Manufacturing the inter-skill differential varies between -0.051 and 0.368. For Construction, results of Table 8 indicate that the inter-skill differential is in the range -0.093 and 0.316, being some of these not significant. The situation for Trade is quite similar: values are between -0.026 and 0.282, and insignificant for some years.

These tables offer similar broad conclusions: the values of the inter-skill differential vary within a quite large range, and some of them are not significantly different from zero. Furthermore, in some cases the indicator takes on a negative sign, which contradicts the hypothesis that high-skilled workers face lower search frictions. We can also observe that there is a clear global time trend: the inter-skill differential increased from 1987 to 1999 and in the last years the inter-skill differential is generally positive.[21]

These results indicate that there is some search friction heterogeneity across skills, notably in the more recent years. It has to be noted though that our measure of search frictions deviates from what is commonly used in the literature: the ratio of job offer arrival rate and the separation rate (see e.g. Ridder and Van den Berg, 2003, and Mortensen, 2003). Unfortunately, our data do not provide the joint distribution of job durations and wages that is necessary for the estimation of this ratio, so we are forced to rely only on information on job transitions. Basically, our index differs from the traditional one in the sense that the observed job-to-job transitions are those received offers that were accepted, so they depend on the relative position of the workers in the wage distribution. Considering that high-skill workers tend to have higher positions in the wage distribution, we can

---

[21]For the cases not reported here (remaining years and regions), the situation is similar. Only for the region "Islands" significance is rarer, perhaps because of the smaller size of the sample.

expect that our index underestimates more heavily the search frictions for high-skill than it does for low-skill. This would result in an underestimation of the inter-skill difference of search frictions.

One can reasonably assume that this underestimation is roughly the same across submarkets, and thus use the obtained results to investigate in a valid way whether the pattern of heterogeneity is systematically related with the degree of assortative matching across submarkets.[22]

Table 10 presents descriptive statistics for both search frictions differences and assortative matching. The means for the measures of assortative matching indicate that PAM measured by regional submarkets is clearly lower. The correlation between search frictions differences and assortative matching is estimated to be positive but not high. Considering the correlation coefficient as the measure of assortative matching, the correlation between search frictions heterogeneity and assortative matching is around 0.20. Taking the regression coefficient, this correlation becomes lower, about 0.14.

These results do not provide a clear evidence that behind the found PAM lies or does not lie search frictions heterogeneity across workers skills. Data that is more informative on search frictions would be necessary to provide a more clear evidence.

6   Conclusion

Labor market theories provide different and contradicting predictions regarding the equilibrium pattern of assortative matching between worker and firms. We use a unique Portuguese data set to test these predictions. These data are particularly suitable for our purposes, for various reasons: the data are longitudinal in both work and firm dimensions, the output of all firms in the private sector is known for several years, a detailed characterization of the complete workforce of the firm is available, and the number of hours worked by each employee is reported. With this information we are able to extract the firm-specific unobserved productivity and use these to test whether positive assortative matching (PAM) is important in the Portuguese labor market.

The empirical results on the sign and extent of assortative matching are unambiguous. There is strong PAM in the Portuguese labor market. In other words, firms and workers of similar productivities tend to match together. We also find that the degree of assortative matching varies across industries.

---

[22]We estimate the measures for assortative matching for each submarket and year.

As discussed in Section 1, the Portuguese labor market has institutional features that may boost PAM. It would be an interesting topic for further research to examine whether other countries, with different institutions, display different degrees of PAM.

A sensitivity analysis based on panels of different lengths shows that assortative matching is stronger when measured among long-lived firms only. It is not easy to interpret this result. After all, long-lived firms compete with short-lived firms. However, long-lived firms are on average more productive, so it seems that as time proceeds their part of the market moves towards the complete-segmentation equilibrium assignment outcome in a perfect market (Becker, 1973).

The results on the question whether PAM is explained by cross-skill heterogeneity of search frictions are not completely unambiguous. There is some evidence for search frictions heterogeneity across skills, but we do not find a strong correlation between the degree of search friction heterogeneity and the extent of assortative matching across submarkets.

References

Abowd, J., J.F. Kramarz, P. Lengermann, and S. Perez-Duarte (2004), "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States" [*mimeo*].

Arellano, M. and S.R. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, 58, 277–297.

Arellano, M. and O. Bover (1995), "Another look at the instrumental-variable estimation of the error-components models", *Journal of Econometrics*, 68, 29–52.

Becker, G.S. (1973), "A theory of marriage", *Journal of Political Economy*, 81, 813–846.

Blundell, R. and S.R. Bond (1999), "Gmm estimation with persistent panel data: an application to production functions", Working Paper 99/4, The Institute for Fiscal Studies.

Burdett, K. and M.G. Coles (1999), "Long-term partnership formation: marriage and employment", *The Economic Journal*, 109, F307–334.

Cardoso, A.R. (2005), "Jobs for young university graduates: is it worth having a degree?", Working Paper 1311, IZA.

Cardoso, A.R. and M. Portela (2005), "The provision of wage insurance by the firm: Evidence from a longitudinal matched employer-employee dataset", Working paper, IZA, Bonn.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau (1973), "Transcendental logarithmic production frontiers", *The Review of Economics and Statistics*, 55, 28–45.

Haltiwanger, J.C., J.I. Lane, and J.R. Spletzer (1999), "Productivity differences accross employers: the roles of employer size, age, and human capital", *American Economic Review*, 89, 94–98.

Haltiwanger, J.C., J.I. Lane, and J.R. Spletzer (2007), "Wages, productivity, and the dynamic interaction of businesses and workers", *Labour Economics*, 14, 575–602.

Hellerstein, J.K. and D. Neumark (1995), "Are earning profiles steeper than productivity profiles? Evidence from Israeli firm-level data", *The Journal of Human Resources*, 30, 89–112.

Hellerstein, J.K., D. Neumark, and K.R. Troske (1999), "Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations", *Journal of Labor Economics*, 17, 409–446.

Lindeboom, M., F. Portrait, and G.J. van den Berg (2002), "An econometric analysis of mental-health effects of major events in life of older individuals", *Health Economics*, 11, 505–520.

Mortensen, D.T. (2003), *Wage Dispersion*, MIT Press, Cambridge.

OECD (1996), "Employment outlook 1996", Technical report, OECD, PARIS.

OECD (1997), "Thematic review of the transition from initial education to working life", Technical report, OECD, PARIS [Online at http://www.oecd.org].

OECD (1998), "Employment outlook 1998", Technical report, OECD, PARIS.

Ridder, G. and G.J. van den Berg (2003), "Measuring labor market frictions: a cross-country comparison", *Journal of the European Economic Association*, 1, 224–244.

Shi, S. (2001), "Frictional assignment. i. efficiency", *Journal of Economic Theory*, 98, 232–260.

Shimer, R. (2005), "The assignment of workers to jobs in an economy with coordination frictions", *Journal of Political Economy*, 113, 996–1025.

Shimer, R. and L. Smith (2000), "Assortative matching and search", *Econometrica*, 68, 343–369.

Van den Berg, G.J. and A. van Vuuren (2003), "The effect of search frictions on wages", Working Paper 3979, CEPR.

Vieira, J.A.C., A.R. Cardoso and M. Portela (2005), "Gender segregation and the wage gap in Portugal: An analysis at the establishment level", *Journal of Economic Inequality*, 3, 145–168.

	Industry		
	Manufacturing	Construction	Trade
Labor Quality 1 (L_1)	**0.058***	**-0.155***	**0.030***
	(0.010)	(0.019)	(0.013)
Female	1.696†	1.539	0.732†
	(0.120)	(0.566)	(0.119)
Aged 30-50	4.751†	0.001†	6.163†
	(0.716)	(0.002)	(2.553)
Aged over 50	7.600†	0.001†	8.874†
	(1.370)	(0.001)	(3.845)
Labor Quality 2 (L_2)	**0.194***	**0.073***	**0.112***
	(0.014)	(0.027)	(0.016)
Female	1.200†	1.145	1.128
	(0.069)	(0.284)	(0.099)
Aged 30-50	4.890†	25.289	3.355†
	(0.374)	(14.289)	(0.398)
Aged over 50	3.985†	30.099	3.015†
	(0.578)	(19.323)	(0.535)
Labor Quality 3 (L_3)	**0.175***	**0.420***	**0.116***
	(0.011)	(0.030)	(0.013)
Female	1.999†	1.369†	1.405†
	(0.167)	(0.143)	(0.139)
Aged 30-50	1.882†	1.374†	1.659†
	(0.155)	(0.139)	(0.156)
Age over 50	1.676	1.484	0.813
	(0.362)	(0.345)	(0.231)
$(L_1)^2$	**0.025***	**-0.033***	**0.024***
	(0.007)	(0.010)	(0.006)
$(L_2)^2$	0.078*	0.036*	0.066*
	(0.006)	(0.008)	(0.009)
$(L_3)^2$	0.046*	0.098*	0.030*
	(0.004)	(0.011)	(0.006)
$L_1 * L_2$	**-0.057***	**0.001**	**-0.013***
	(0.007)	(0.002)	(0.006)
$L_2 * L_3$	-0.026*	-0.004	-0.009
	(0.004)	(0.006)	(0.005)
$L_1 * L_3$	-0.035*	0.015*	-0.014*
	(0.005)	(0.006)	(0.005)
Workforce size	**-0.096***	**-0.204***	**-0.152***
	(0.008)	(0.015)	(0.008)
Number of observations	87 786	20 684	96 067
F-test	205.16	134.01	75.58

Table 4: Results for the translog function by industry, for Panel 1

Notes: This panel includes all the firms observed at least in two years. Fixed firm effects and a control variable for the firm being a single establishment are included in the function. * indicates rejection at 5% level of H_0: coefficient= 0. For the demographic characteristics within each of the labor qualities, the test of interest is instead H_0: coefficient= 1. Rejection at 5% level of this hypothesis is indicated by †. Standard errors are in brackets. The F-test is the test of global significance of the model. The p-value of each of the presented F-tests is 0.00.

Panel	Industry		
	Manufacturing	Construction	Trade
Correlation coefficient			
Panel 1 (≥ 2)	0.295**	0.371**	0.152**
	(0.032)	(0.069)	(0.038)
Panel 2 (≥ 6)	0.358**	0.486**	0.233**
	(0.032)	(0.113)	(0.039)
Panel 3 (≥ 9)	0.381**	0.485**	0.226**
	(0.064)	(0.142)	(0.059)
Panel 4 (≥ 12)	0.365**	0.631**	0.230**
	(0.059)	(0.122)	(0.056)
Rank correlation coefficient			
Panel 1 (≥ 2)	0.329**	0.365**	0.153**
	(0.037)	(0.080)	(0.039)
Panel 2 (≥ 6)	0.399**	0.469**	0.230**
	(0.031)	(0.122)	(0.040)
Panel 3 (≥ 9)	0.411**	0.437*	0.233**
	(0.065)	(0.169)	(0.061)
Panel 4 (≥ 12)	0.381**	0.626**	0.230**
	(0.062)	(0.148)	(0.054)
Regression coefficient			
Panel 1 (≥ 2)	2.277**	3.047*	0.995**
	(0.265)	(1.439)	(0.251)
Panel 2 (≥ 6)	2.660**	3.678*	1.362**
	(0.252)	(1.412)	(0.247)
Panel 3 (≥ 9)	2.885**	3.473	1.250**
	(0.591)	(1.776)	(0.358)
Panel 4 (≥ 12)	2.524**	5.353**	1.237**
	(0.492)	(1.482)	(0.318)
Number of observations			
Panel 1 (≥ 2)	16040	4382	20259
Panel 2 (≥ 6)	6978	1503	6754
Panel 3 (≥ 9)	3580	672	3062
Panel 4 (≥ 12)	1534	252	1071

Table 5: Measures for assortative matching by industry and panel

Notes: Assortative matching is measured by the relation between the firm fixed effect and its respective time average proportion of workers in the highest schooling level. ** and * indicate statistical significance of each as measure for assortative matching at 1% and 5%, respectively.

| | Industry | | |
Panel	Manufacturing	Construction	Trade
	Correlation coefficient		
Panel 1 (≥ 2)	-0.172**	-0.309**	-0.158**
	(0.035)	(0.082)	(0.050)
Panel 2 (≥ 6)	-0.175**	-0.342**	-0.199**
	(0.039)	(0.115)	(0.050)
Panel 3 (≥ 9)	-0.174**	-0.269	-0.203**
	(0.061)	(0.158)	(0.058)
Panel 4 (≥ 12)	-0.149**	-0.400**	-0.247**
	(0.053)	(0.110)	(0.061)
	Rank correlation coefficient		
Panel 1 (≥ 2)	-0.175**	-0.294**	-0.157**
	(0.036)	(0.070)	(0.049)
Panel 2 (≥ 6)	-0.179**	-0.294**	-0.184**
	(0.039)	(0.105)	(0.049)
Panel 3 (≥ 9)	-0.171**	-0.229	-0.185**
	(0.058)	(0.156)	(0.058)
Panel 4 (≥ 12)	-0.139**	-0.358**	-0.221**
	(0.050)	(0.112)	(0.056)
	Regression coefficient		
Panel 1 (≥ 2)	-0.692**	-1.270**	-0.787**
	(0.137)	(0.817)	(0.249)
Panel 2 (≥ 6)	-0.679**	-1.236	-0.871**
	(0.147)	(0.636)	(0.225)
Panel 3 (≥ 9)	-0.694**	-0.860	-0.864**
	(0.255)	(0.796)	(0.266)
Panel 4 (≥ 12)	-0.577*	-1.636**	-1.034**
	(0.219)	(0.587)	(0.264)
	Number of observations		
Panel 1 (≥ 2)	16040	4382	20259
Panel 2 (≥ 6)	6978	1503	6754
Panel 3 (≥ 9)	3580	672	3062
Panel 4 (≥ 12)	1534	252	1071

Table 6: Measures for assortative matching by industry and panel

Notes: Assortative matching is measured by the relation between the firm fixed effect and its respective time average proportion of workers in the lowest schooling level. ** and * indicate statistical significance of each measure for assortative matching at 1% and 5%, respectively.

Years	Number of observations	Schooling			Interskill differential
		Low	Medium	High	
			North		
87/88	258236	0.234	0.197	0.183	-0.051**
					(0.007)
93/94	277390	0.144	0.191	0.174	0.030**
					(0.007)
96/97	252829	0.098	0.144	0.143	0.044**
					(0.008)
99/00	274401	0.171	0.363	0.376	0.205**
					(0.012)
			Center		
87/88	59777	0.261	0.272	0.260	-0.001
					(0.018)
93/94	90281	0.108	0.146	0.138	0.030**
					(0.008)
96/97	88589	0.089	0.146	0.155	0.066**
					(0.013)
99/00	99460	0.200	0.344	0.359	0.159**
					(0.024)
		Lisbon and Tagus Valley			
87/88	170152	0.172	0.160	0.170	-0.002
					(0.007)
93/94	145636	0.175	0.228	0.234	0.059**
					(0.007)
96/97	125644	0.091	0.162	0.277	0.186**
					(0.020)
99/00	125685	0.158	0.306	0.526	0.368**
					(0.020)

Table 7: Indicators for search frictions for Manufacturing, by region and year

Notes: The columns 3, 4 and 5 report our search frictions indicator: the ratio between observed probabilities of moving to another firm and of leaving the labor force of the private sector. The inter-skill differential is computed as the difference between columns 3 and 5. Standard errors (between brackets) were obtained through bootstrapping. Statistical significance at the 5% level is indicated by **.

Years	Number of observations	Schooling			Interskill differential
		Low	Medium	High	
		North			
87/88	29898	0.264	0.162	0.280	0.015
					(0.039)
93/94	40734	0.150	0.133	0.181	0.031
					(0.018)
96/97	42312	0.134	0.128	0.248	0.114*
					(0.025)
99/00	55127	0.247	0.356	0.500	0.253*
					(0.043)
		Center			
87/88	7490	0.201	0.121	0.108	-0.093*
					(0.032)
93/94	14776	0.154	0.129	0.131	-0.024
					(0.023)
96/97	16098	0.152	0.150	0.172	0.020
					(0.037)
99/00	22619	0.277	0.331	0.393	0.116*
					(0.044)
		Lisbon and Tagus Valley			
87/88	26116	0.180	0.129	0.169	-0.012
					(0.015)
93/94	39065	0.170	0.161	0.275	0.105*
					(0.014)
96/97	40429	0.132	0.138	0.248	0.116*
					(0.018)
99/00	50646	0.256	0.308	0.572	0.316*
					(0.029)

Table 8: Indicators for search frictions for Construction, by region and year

Notes: The columns 3, 4 and 5 report our search frictions indicator: the ratio between observed probabilities of moving to another firm and of leaving the labor force of the private sector. The Inter-skill differential is computed as the difference between columns 3 and 5. Standard errors (between brackets) were obtained through bootstrapping. Statistical significance at the 5% level is indicated by *.

Years	Number of observations	Schooling			Interskill differential
		Low	Medium	High	
			North		
87/88	49225	0.161	0.174	0.176	0.015
					(0.009)
93/94	66534	0.126	0.179	0.188	0.062*
					(0.008)
96/97	83248	0.071	0.127	0.180	0.110*
					(0.008)
99/00	102782	0.167	0.329	0.410	0.242*
					(0.015)
			Center		
87/88	16471	0.145	0.160	0.119	-0.026
					(0.016)
93/94	26596	0.107	0.164	0.138	0.031*
					(0.014)
96/97	35468	0.073	0.128	0.114	0.041*
					(0.015)
99/00	43841	0.159	0.321	0.328	0.169*
					(0.015)
		Lisbon and Tagus Valley			
87/88	96655	0.193	0.180	0.206	0.014
					(0.008)
93/94	120639	0.173	0.231	0.256	0.083*
					(0.005)
96/97	145580	0.090	0.128	0.190	0.101*
					(0.052)
99/00	169592	0.197	0.366	0.479	0.282*
					(0.012)

Table 9: Indicators for search frictions for Trade, by region and year

Notes: The columns 3, 4 and 5 report our search frictions indicator: the ratio between observed probabilities of moving to another firm and of leaving the labor force of the private sector. The Inter-skill differential is computed as the difference between columns 3 and 5. Standard errors (between brackets) were obtained through bootstrapping. Statistical significance at the 5% level is indicated by **.

	Interskill diff.	Correlat. coeff.	Regress. coeff.
	(a)	(b)	(c)
Mean	0.067	0.097	0.672
Stand. error	0.084	0.118	1.018
corr(a, b)		0.198	
corr(a, c)		0.141	

Table 10: Descriptive statistics of search frictions indices and measures for assortative matching

Notes: For each measure, we have 150 observations. Correlation and regression coefficients refer to the relation between the firm fixed effects and the proportion of high-skilled workers. These were estimated based on the panel including all the firms observed at least two years.