Enabling Data Analytics for Actions Tackling Skills Shortages & Mismatch

Project Title:	**TRAILS - Enabling data analytics for actions tackling skills shortages & mismatch**
Contract No:	101132673
Instrument:	Research and Innovation Action
Thematic Priority:	HORIZON-CL2-2023-TRANSFORMATIONS-01
Start of project:	1 January 2024
Duration:	36 months

Deliverable No: D4.1

NOVEL I: Using machine learning to measure skills matched to needs

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101132673

Funded by the European Union

Document Control Page

Deliverable Name	NOVEL I: Using machine learning to measure skills matched to needs
Deliverable Number	D4.1
Work Package	WP4
Associated Task	4.1
Covered Period	M01-M24
Due Date	31/12/2025
Completion Date	28/12/2025
Submission Date	30/12/2025
Deliverable Lead Partner	UNINA
Deliverable Author(s)	Luca Coraggio, Lorenzo Incoronato, Monica Langella, Armando Miano, Marco Pagano, Marco Stenborg Petterson, Vincenzo Pezone and Annalisa Scognamiglio
Version	1

Dissemination Level		
PU	Public	**X**
CO	Confidential to a group specified by the consortium (including the Commission Services)	

Document History

Version	Date	Change History	Author(s)	Organisation
1	December 28, 2025	Initial version	Luca Coraggio, Lorenzo Incoronato, Monica Langella, Armando Miano, Marco Pagano, Marco Stenborg Petterson, Vincenzo Pezone and Annalisa Scognamiglio	UNINA, Tilburg

Internal Review History

Name	Institution	Date
Luca Coraggio	UNINA	December 27, 2025
Marco Pagano	UNINA	December 27, 2025

Legal Notice
This document has been produced in the context of the TRAILS Project. The TRAILS project is part of the European Community's Horizon Europe Program for research and development and is as such funded by the European Commission. All information in this document is provided 'as is' and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. For the avoidance of all doubts, the European Commission has no liability concerning this document, which is merely representing the authors' view.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

Acronym	Explanation
CEO	Chief executive officer
INPS	Istituto Nazionale della Previdenza Sociale
ISCO	International Standard Classification of Occupations
ISTAT	National Statistical Italian Centre
JAQ	Job Allocation Quality
LISA	Longitudinal integration database for health insurance and labour market studies (Statistics Sweden)
O*NET	Occupational Information Network
OROA	Operating Return on Assets
PSID	Panel Study of Income Dynamics
SME	Small and Medium-sized Enterprise
SUN2000	Swedish Standard Classification of Education
UI	Unemployment Insurance

ACKNOWLEDGEMENTS

The authors acknowledge the exceptional contribution of Lorenzo Sinno, who provided outstanding research assistance for this project. Not only did he process the data for Portugal and produce most of the corresponding results, but he also enhanced that part of the report with additional, insightful results. This report benefited substantially from his commitment and creativity.

EXECUTIVE SUMMARY

Efficiently matching workers to jobs that best utilize their skills is a central determinant of productivity growth, wage dynamics, and firms' economic performance. Yet measuring job–skill mismatch in a way that is economically meaningful, scalable, and comparable across countries has long posed a challenge. Traditional mismatch indicators—based on education–occupation gaps or externally defined skill requirements—are often static, coarse, and difficult to harmonize across institutional contexts. This report addresses these limitations by developing and applying an innovative, data-driven measure of job match quality based exclusively on administrative employer–employee data.

The report introduces and applies a machine–learning–based indicator of Job Allocation Quality (JAQ), which measures how efficiently workers are allocated to occupations given their observable characteristics. Rather than comparing workers' skills to externally imposed job requirements, JAQ infers an implicit allocation rule from observed job assignments in highly productive firms, which are assumed to allocate labor relatively efficiently. Using this revealed benchmark, individual worker–job matches are evaluated based on how closely they align with each worker's predicted best occupational assignment. Job mismatch is thus defined as a deviation from empirically observed efficient allocation patterns rather than from normative or expert-defined standards.

The empirical analysis draws on rich matched employer–employee administrative data from four European countries—Sweden, Portugal, Italy, and the Netherlands—covering long time horizons and millions of workers and firms. Despite differences in institutional settings, data availability, and labor market structures, the report applies a harmonized methodology across countries, enabling direct comparison of results. At the worker level, JAQ is measured both as a binary indicator (whether the worker is assigned to her most suitable occupation) and as a continuous probability-based measure. At the firm level, JAQ captures the share of well-matched workers and the average quality of matches within the firm.

Several robust findings emerge across countries. First, job match quality improves sharply early in workers' careers and then plateaus. In all countries analyzed, match quality increases steeply with labor market experience and job tenure during the first years of employment, with much smaller gains thereafter. This pattern highlights the importance of learning—by firms about workers' abilities and by workers about their comparative advantage—and suggests that most allocative efficiency gains occur early in working lives.

Second, better job matches are consistently associated with higher wages. Across countries where earnings data are available, workers whose job assignments are more closely aligned with their predicted suitability earn higher wages, even after controlling for observable characteristics, occupations, firms, and—in more demanding specifications—worker fixed effects. Although the magnitude of the wage premium varies, the positive relationship between match quality and earnings is stable and statistically significant, validating JAQ as an economically meaningful measure of job matching rather than a purely statistical construct.

Third, firms with higher job allocation quality are more productive. At the firm level, JAQ is strongly and positively correlated with labor productivity, measured by sales or value added per employee, across all four countries. This relationship remains robust after controlling for firm size, industry, capital intensity, workforce composition, and fixed effects. By contrast, the relationship between match quality and profitability is weaker and often insignificant, suggesting that productivity gains from better matching are at least partly shared with workers through higher wages rather than fully accruing to firm owners.

Fourth, job match quality is systematically related to firm size and organizational maturity. In all countries, match quality increases rapidly with firm size up to a threshold—typically around 20 to 30 employees—after which it levels off. Older firms also tend to exhibit higher match quality. These findings point to fixed organizational costs in developing effective personnel management and job assignment practices, implying that small and young firms face structural disadvantages in efficiently allocating labor.

Alongside these robust patterns, the report documents meaningful cross-country differences, particularly in the time-series behavior of match quality. While aggregate JAQ is relatively stable over time in all countries, its cyclical dynamics differ in ways that reflect exposure to major macroeconomic shocks—most notably the euro-area sovereign debt crisis. Sweden, which lies outside the euro area, exhibits clearly pro-cyclical movements in match quality, with declines during the global financial crisis and partial recovery thereafter. In contrast, Portugal and Italy—both heavily affected by the euro-area sovereign crisis—display more persistent and subdued dynamics, suggesting that prolonged fiscal consolidation, credit constraints, and labor market rigidities slowed reallocation and rendered mismatch more structural during the crisis years. The Netherlands occupies an intermediate position, with smoother dynamics consistent with stronger labor market adjustment mechanisms and less severe sovereign stress. These differences underscore that mismatch is not purely structural but responds to macroeconomic shocks in ways shaped by institutional and financial constraints.

The analysis also shows that match quality varies across sectors and occupations, with greater mismatch in sectors characterized by complex task structures, such as manufacturing, and better matching in more standardized sectors. By contrast, demographic characteristics such as gender, immigrant status, and contract type are only weakly and inconsistently related to match quality, suggesting that allocative inefficiencies are driven primarily by career histories, task complexity, and firm-level assignment practices rather than by observable worker traits.

Overall, the report delivers three main messages. First, job mismatch is economically meaningful and quantitatively important, with clear implications for wages and productivity. Second, mismatch is dynamic rather than immutable: it declines with experience, varies over the business cycle, and responds to institutional and macroeconomic conditions. Third, machine-learning methods applied to administrative data offer a powerful and scalable tool for measuring job match quality in a way that is comparable across countries and over time.

From a policy perspective, the findings suggest that interventions aimed at improving labor market efficiency should focus on early career stages, facilitating learning and mobility, and strengthening managerial and organizational capabilities within firms, especially small and young firms. More broadly, the strong consistency of the core results across diverse European labor markets highlights the potential of JAQ as a building block for real-time monitoring of skills mismatch and labor allocation efficiency within the European Union.

1. INTRODUCTION

1.1 Purpose of the deliverable

The efficient allocation of workers to jobs that best utilize their skills is a central determinant of productivity, wage growth, and firm performance. Because labour services are intrinsically heterogeneous and labour markets are decentralized, frictions in search and imperfect information inevitably generate mismatches between workers' skills and job requirements. A large body of literature documents that such mismatches are associated with lower wages, weaker career progression, higher turnover, and lower firm productivity. Yet measuring mismatch in a way that is economically meaningful, scalable, and comparable across countries and over time remains a significant challenge.

A wide range of empirical approaches has been proposed to measure occupational mismatch. As illustrated by Deliverable D1.1 of the TRAILS project, traditional measures compare workers' educational attainment to job-specific requirements, identifying over- and under-education. While straightforward, these measures are coarse and fail to capture heterogeneity within educational groups, on-the-job learning, and multidimensional skills. Task- and skill-based approaches improve on this by mapping workers' characteristics into multidimensional skill vectors and comparing them to externally defined job requirements, often drawn from expert classifications or survey data. While these measures are conceptually rich, they rely on static benchmarks, making them challenging to harmonize across institutional contexts and over time. A third strand of occupational mismatch measures infers mismatch indirectly from realized labour market outcomes, such as wage residuals or mobility patterns, capturing the consequences rather than the structure of job–skill allocation.

This paper builds on and extends a different approach: a machine-learning-based measure of job allocation quality (JAQ) that exploits matched employer–employee administrative data to infer the quality of worker–job matches from observed allocation patterns in highly productive firms. Rather than comparing workers to externally imposed notions of job requirements, JAQ uses data-driven benchmarks to infer how workers' observable characteristics are mapped into occupational assignments when allocation is relatively efficient. Mismatch is then measured as a deviation from this revealed benchmark.

The purpose of the paper is to apply the JAQ methodology to matched worker–firm data from several European Union (EU) countries over an extended period. The analysis has three main objectives. First, we develop a harmonized measure of job–skill mismatch based solely on administrative data that are widely available across European countries. Because JAQ relies on standard worker characteristics, occupational codes, and firm identifiers, it avoids dependence on surveys, expert evaluations of job requirements, or country-specific skill taxonomies, making it well-suited for applications to countries that collect administrative worker-firm matched data and make them available for research, such as EU countries. Second, we assess the external validity of previous evidence, which is based on Swedish data, by examining whether the relationships between job allocation quality, worker outcomes, and firm performance generalize to other EU labour markets with different institutional features, educational systems, and firm structures. Third, we exploit the time-series depth of the data to study how mismatch evolves and how it responds to major economic and institutional events. By tracking JAQ within firms and countries over long horizons, we investigate whether mismatch is primarily structural or whether it varies systematically over the business cycle, during financial crises, and following labour market reforms.

1.2 Relation with other deliverables and tasks

This report builds on deliverable D1.1 by Coraggio et al. (2024), which provided an extensive survey of the theoretical and empirical literature on occupational mismatch and pointed to the new insights that can be gained from innovative measures of match quality built by applying machine-learning algorithms to administrative worker-firm matched data. This paper can be considered as a proof of concept of this idea, as it applies such methods to administrative data for four EU countries: (i) it shows that such measures can be constructed for all EU countries; (ii) it validates them by showing how they correlate with worker-level and firm-level outcomes; (iii) it also highlights the limitations that are currently encountered in constructing them because of administrative constraints in accessing them due to confidentiality concerns and national regulations, and because of cross-country differences in the structure of national databases and to differences in the availability of data regarding certain worker and/or firm characteristics. However, also learning about these current limitations and difficulties can be considered as a helpful insight, as it enables us to point to steps to be taken if researchers are to build fully comparable real-time measures of job-mismatch for all EU countries, exploiting the unique information that Europe has accumulated over time via its administrative worker-firm data.

1.3 Structure of the Document

The remainder of the paper proceeds as follows.

Section 2 describes the construction and interpretation of the JAQ measure, clarifies its relationship to existing mismatch indicators, and lays out its expected relationships with worker and firm characteristics and outcomes, as well as its time-series response to macroeconomic shocks.

Section 3 presents country-level data and results for Sweden, Portugal, Italy, and the Netherlands, following a harmonized structure that facilitates comparison across countries and over time.

Finally, Section 4 provides concluding remarks, highlighting both the substantive empirical findings that emerge consistently across all four countries and those that differ across them, and draws the main new insights for research and policy from reliance on the JAQ measure.

2. Methodology

This section presents the job allocation quality (*JAQ*) measure and explains how it is used to study mismatch at the worker and firm level, as well as its evolution over time. The methodology closely follows the framework introduced by Coraggio et al. (2025) while emphasizing features relevant to multi-country and longitudinal analyses.

2.1 Construction and meaning of JAQ

The *JAQ* measure is based on the idea that firms assign workers to jobs according to an implicit allocation rule that maps workers' observable characteristics into occupational positions. Let X_{it} denote a vector of characteristics of worker *i* at time *t*, including demographics, education, and labour market history, and let J_{it} denote the occupation held by the worker. The assignment rule can be represented as $J_{it} = g(X_{it}, Z_{ft})$, where the vector Z_{ft} captures firm characteristics such as industry and size.

Instead of assuming a parametric form for $g(\cdot)$, we use a Random Forest classifier to estimate the conditional probability $P(J = j \mid X, Z)$, for each possible occupation *j,* from a learning sample of worker-job assignments in high-productivity firms. A key identifying assumption is that high-productivity firms are more likely to implement an efficient assignment rule. Accordingly, the algorithm $P(J = j \mid X, Z)$ is trained on worker–job matches observed in firms belonging to the top decile of the productivity distribution within narrowly defined industry–size cells.

Using the estimated model, we predict for each worker the probability of being assigned to each occupation, conditional on the worker's characteristics. The occupation associated with the highest predicted probability is interpreted as the job for which the worker is most suitable, given her observable characteristics and the revealed allocation rule. This procedure yields a binary indicator, *eJAQ*, which equals 1 if the worker's observed occupation coincides with the occupation for which she has the highest predicted suitability, and 0 otherwise, as well as a continuous measure, *epJAQ*, equal to the predicted probability associated with the worker's observed occupation.

The firm-level *JAQ* is constructed by aggregating worker-level match quality *eJAQ* across all employees within a firm-year. Higher values of *JAQ* indicate that a larger share of the workforce is allocated to jobs that are well aligned with workers' predicted comparative advantage. As done at the worker level, we also aggregate the continuous worker-level measure *epJAQ* at the firm level by averaging it across the firm's employees. We denote this firm-level average by *pJAQ*, which is thus the continuous analogue of the dichotomic *JAQ* measure of match quality.

2.2 Comparison with other mismatch measures

Relative to existing mismatch measures surveyed in the literature, *JAQ* occupies a distinct position. Unlike education-based indicators of over- and under-education, it does not collapse skills into a single dimension and naturally incorporates experience and career histories. Unlike task- and skill-based measures that rely on expert taxonomies or surveys, it infers job requirements endogenously from administrative data. It therefore adapts to country-specific time-changing institutional and technological contexts.

JAQ also differs from other realized-match approaches. Measures based on the average characteristics of incumbents or tenured workers within occupation–firm cells, such as those proposed by Fredriksson, Hensvik, and Skans (2018), implicitly assume that current allocations are optimal on average and are typically not applicable to all workers. By contrast, *JAQ* benchmarks assignments against those observed in high-productivity firms, relies on a machine learning algorithm rather than simple averages, and can therefore be computed for all employees, including new hires and younger workers.

Finally, *JAQ* differs from outcome-based mismatch measures derived from wages or mobility patterns in that it targets the allocation mechanism itself rather than its ex-post consequences.

2.3 JAQ and worker-level outcomes

At the worker level, *JAQ* captures how well an individual's skills are aligned with her job assignment. Previous evidence shows that better-matched workers earn higher wages, experience faster improvements in match quality early in their careers, and are less likely to separate from their employer. In the country-level analyses of Section 3, we study how wages, experience, and job mobility relate to *JAQ* in each institutional setting.

2.4 JAQ and firm-level outcomes

Aggregated to the firm level, *JAQ* provides a measure of how efficiently firms allocate human capital across tasks. Previous work by Coraggio et al. (2025) based on Swedish administrative data from 2000 to 2010 shows that firms with higher *JAQ* are more productive and that improvements in managerial quality translate into better job allocation. In Section 3, we extend this work to administrative data from Italy, the Netherlands and Portugal, as well as to a longer data set from Sweden, encompassing data from 2000 to 2010. For each of these countries we investigate the relationship between firm-level JAQ and firm productivity, as measured by value added per employee, and to profitability measures, as measured by return on assets.

2.5 Time-series variation in JAQ

A key advantage of the *JAQ* metric is that it can be computed repeatedly over time for the same firms and workers. This enables a dynamic analysis of mismatch. For each country, we will examine how the average *JAQ* evolves over time, how match quality changes within firms, and how *JAQ* correlates with macroeconomic conditions, such as recessions, expansions, and crisis episodes. In particular, this enables us to test whether aggregate match quality displays a procyclical pattern, consistent with the findings by Baley et al. (2022) and Bowlus (1995). This time-series perspective prepares the ground for interpreting mismatch as a dynamic outcome shaped by economic shocks and institutional change rather than as a purely static inefficiency.

3. Country-level Evidence

In this section, we provide evidence based on the machine-learning measures of worker-level and firm-level match quality presented in Section 2 for four countries: Sweden (Section 3.1), Portugal (Section 3.2), Italy (Section 3.3), and the Netherlands (Section 3.4). Each country section consists of seven parts:

1. Description of the administrative worker and firm data sources used for the relevant country.

2. Descriptive statistics regarding worker characteristics and the balancing between the learning and the main sample.

3. Correlation between our worker-level match quality measure (*eJAQ*) and worker characteristics. Here we describe how *eJAQ* varies over the working life of the average worker in the population: we expect it to increase with workers' age and experience, as managers learn about their characteristics (Fredriksson et al., 2018), and employees themselves adapt their skills via on-the-job training (Guvenen et al., 2020).

4. Relationship between individual wages and match quality. Here we explore how individual wages evolve as match quality evolves over workers' careers: insofar as match quality improves with tenure, we expect the resulting productivity gains to be partly appropriated by workers, and thus wages to be positively related to *eJAQ*. We do so both via graphical analysis and by estimating a regression of the logarithm of the annual earnings of worker i in year t on (i) a job indicator, (ii) the measure $eJAQ_{it}$ for worker i in year t, (iii) the workers' characteristics included in the machine-learning algorithm, (iv) the characteristics of the firm that employs the worker in each year (e.g., 2-digit industry dummies, firm age, indicators for family firm, listed company, presence of a human resources manager), and (v) year dummies.

5. Relationships between firm-level measures of match quality (*JAQ* and *pJAQ)* and firm characteristics, namely, age, size, and employees' age, experience, tenure, and education.

6. Relationships firm-level measures of match quality (*JAQ* and *pJAQ*) and measures of firm performance, i.e., sales per employee, value added per employee, and operating return on assets. Here too, we use graphical analysis and panel regressions to determine whether these two measures of match quality capture meaningful variation in the quality of workforce allocation, rather than merely statistical noise or firm heterogeneity in productivity.

7. Time-series behavior of match quality in the relevant country. Here, we aggregate the firm-level *JAQ* and *pJAQ* measures to the country level to explore whether aggregate measures of match quality display any trends or correlation with recessions or expansions.

Due to data limitations for some countries, the analysis is not exactly symmetric across all four countries. For instance, for Italy we have not yet been able to access wage data, so that Section 3.3 does not (yet) contain evidence regarding the relationship between wages and individual match quality measures in the INPS data. However, this shortcoming is offset by additional evidence on an issue that is not (yet) investigated in the other three countries, namely, the relationship between match quality and job switches.

3.1 Sweden

This section describes the data and presents the results regarding our worker-level and firm-level measures of match quality obtained from the administrative worker-firm matched data for Sweden.

3.1.1 Data sources

To develop and estimate the *JAQ* measure proposed for Sweden, we use registry data. This data set is ideal for our purposes because it allows us to observe the entire population of workers and firms in Sweden over a relatively long period, including several variables on workers' job histories, such as occupations and wages over their careers.

The bulk of our data comes from the Statistics Sweden LISA database that covers the whole Swedish population of individuals who are at least 16 years old and reside in Sweden at the end of each year. This longitudinal matched employer-employee database integrates information from registers held by various government authorities. We have data for the 1990–2020 interval, but our analysis focuses on the 2002–12 interval because occupation information is too sparse before 2000, and a major change occurred in the occupation classification system in 2012. However, we draw on 1990–2001 data in constructing worker job histories.

The estimation of a worker's suitability for a given job is based on the same type of information that would typically be included in individual resumes available to managers assigning workers to jobs, namely, background information, education, and past work experience. Background information, drawn from LISA, includes age, gender, an indicator for immigrant status, residence municipality and a mobility indicator equal to one for workers employed in a county different from the county of birth. As for education, we observe both the education level (basic, high school, vocational, or university) and the education subject (no specialization, law, business and economics, health and medicine, natural sciences, teaching, engineering, social sciences, services, or other specializations). Finally, past work experience is captured by labor market experience (measured as years since graduation), tenure at the current firm, number of firms and number of two-digit industries where an individual previously worked, total number of unemployment days since 1992 (when unemployment data starts in LISA), years of experience in each occupation, years of experience in each 2-digit industry, and years of experience in each decile of the distribution of firms' number of employees or total assets.

The firm-level variables drawn from LISA are firm age, 2-digit industry, size (measured by the number of employees), sales, and total assets. We identify jobs based on international ISCO-88 (COM) 2-digit classification of occupations, based on data provided primarily by official wage statistics drawn from yearly surveys of around 11,000 companies. Companies with more than 500 workers are surveyed every year and the remainder is a random sample of firms. Occupation data is gathered for around a million workers each year. The second source is a yearly survey sent out by mail to around 30,000–47,000 companies that are not selected for inclusion in the official wage statistics survey (a total of around 150,000 private sector companies per year). The surveys are sent out on a rolling basis: all 150,000 companies are surveyed at least once in five-years' time.

In extracting our sample of firms from the LISA database, we apply two screens by firm size: we only retain firms whose median number of employees in the sample period is between 10 and 6,000. The lower bound is due to the sparsity of occupational information for firms with less than 10 employees: including these firms would introduce large noise in the estimation of the job-employee matching rule. The upper bound of 6,000 employees excludes from the sample very large firms that may otherwise dominate the estimates of the job-employee matching rule, despite featuring a quite different structure from other firms, e.g., a more layered corporate hierarchy and a richer set of possible occupations. After applying these filters, our sample comprises 64,203 firms, employing an yearly average of 1,660,611 employees.

3.1.2 Descriptive statistics

Table 1.1 presents descriptive statistics regarding job mismatch and worker characteristics in the Swedish data. The table shows that, on average, the predicted probability associated with workers' observed occupation (*epJAQ*) is 0.23, while the fraction of workers whose observed occupation coincides with their most suitable job (*pJAQ*) is 0.39. On average, employees are 40 years-old, have 19 years of work experience and 5 years of tenure in their current job, and display considerable mobility, being employed in 3 different industries and in 4 different firms over their careers. They also feature considerable geographical mobility, as about one third of them lives away from the place where they were born. Female and immigrant workers are 38% and 15% of the total workforce, respectively.

Next, we turn to evidence regarding the balancing of worker characteristics between the main sample and the learning samples used to estimate the Random Forest algorithm. The two samples are sufficiently similar to share common support, as shown in Figure 1.1, which displays the distributions of predicted wages for workers in the two samples. For both samples, the predictions are obtained from wage regressions estimated on the main sample, using the worker characteristics included in the ML algorithm as explanatory variables. The figure shows that the support of the two distributions largely overlaps, even though the learning sample places more weight on high predicted wages than the main sample. This evidence supports our assumption that the learning sample can be used to estimate an allocation rule relevant to workers in firms included in the main sample.

Table 1.1 - Job mismatch and worker characteristics (Sweden)

• Variable	Mean	P50	P10	P25	P75	P90	SD
epJAQ	0.23	0.18	0.02	0.07	0.34	0.53	0.20
eJAQ	0.39	0.00	0.00	0.00	1.00	1.00	0.49
Labor Income	3,284.43	3,144.56	1,347.35	2,389.56	3,919.12	5,049.26	1,812.42
Log Labor Income	8.63	8.75	7.90	8.47	8.97	9.22	0.65
College Degree	0.13	0.00	0.00	0.00	0.00	1.00	0.33
Age	40.30	40.00	23.00	30.00	50.00	58.00	12.36
Female	0.38	0.00	0.00	0.00	1.00	1.00	0.48
Immigrant	0.15	0.00	0.00	0.00	0.00	1.00	0.35
Experience	19.18	18.00	2.00	8.00	29.00	38.00	13.08
Tenure	5.11	3.00	0.00	1.00	8.00	13.00	5.24
Number of Industries	2.86	3.00	1.00	2.00	4.00	5.00	1.59
Number of Firms	3.90	3.00	1.00	2.00	5.00	7.00	2.23
Unempl. Days since 1992	222.14	28.00	0.00	0.00	284.00	710.00	390.46
Graduated in Recession	0.19	0.00	0.00	0.00	0.00	1.00	0.39
Lives where Born	0.64	1.00	0.00	0.00	1.00	1.00	0.48

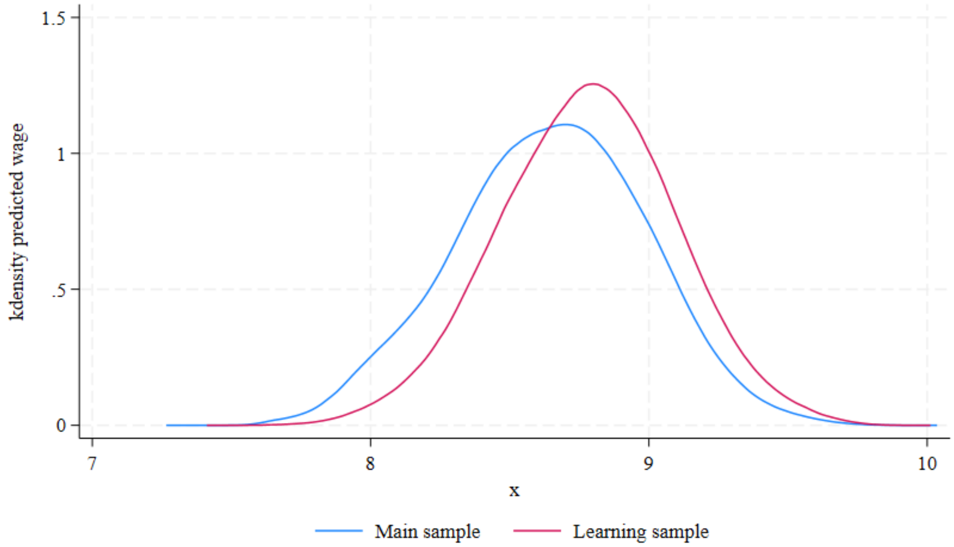

Figure 1.1 - Balancing of characteristics between the main sample and the learning sample

3.1.3 Match quality and worker-level characteristics

Figures 1.2 to 1.4 show that, in the Swedish data, our worker-level measure of job allocation quality, *eJAQ*, is strongly increasing in workers' age, experience and firm-level tenure. In all three cases, the relationship is non-linear: *eJAQ* rises steeply in the first years of a worker's life, experience and job tenure, and then it tends to grow much more slowly or even achieve a plateau.

In particular, the relationship with age in Figure 1.1 shows that workers' allocation quality increases by 29% from age 20 to 30, but only by 4.2% per decade from age 30 to 60, on average. In other words, after a worker turns 30, the average improvement per decade in his/her match quality slows to one seventh of the first decade's improvement. The relationship with job market experience shown in Figure 1.2 and that with job tenure at the firm shown in Figure 1.3 feature even greater concavity, being close to piece-wise piecewise linear functions: *eJAQ* increases by 43% in the first 5 of work experience and by 39% in the first 5 years of tenure in the firm, and then remains almost unchanged in the subsequent 25 years. The difference relative to its age pattern shown in Figure 1.1 can be explained by the fact that workers take up their first job at different ages, depending on when they complete their education. Individuals who do not enrol in university degree programs start gaining work experience and experience a substantial increase in their job match quality in their early 20s. In contrast, those who obtain a university degree delay their first job and improve their match quality in their late 20s. The relationship with age effectively averages the steeper tenure-related increases of workers with different ages of entry into the labor market, so that it exhibits a less steep increase across ages between 20 and 30.

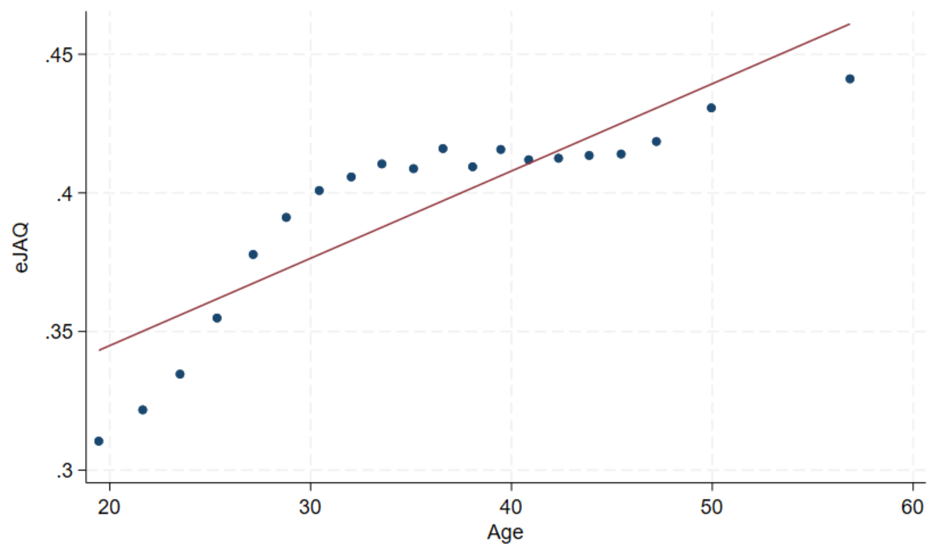

Figure 1.2. individual workers' job match quality and age

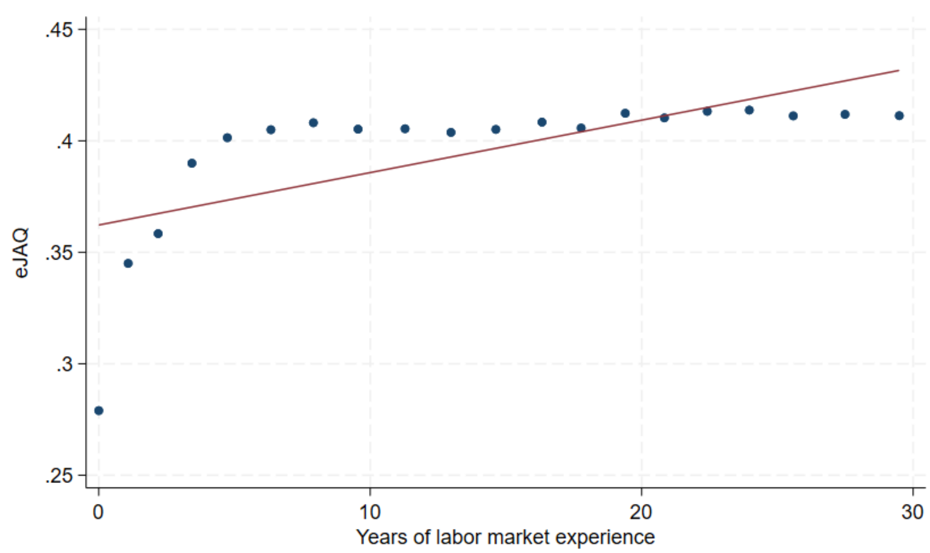

Figure 1.3. individual workers' match quality and labor market experience

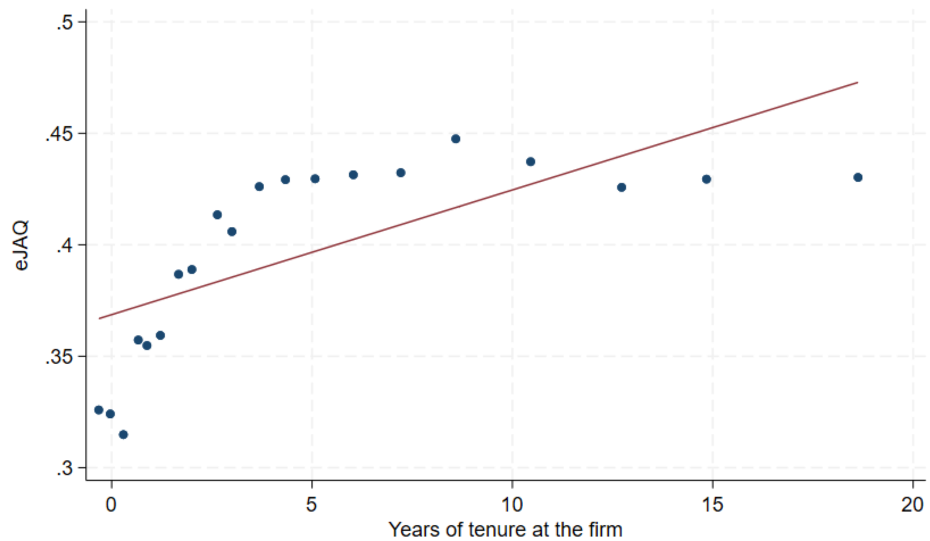

Figure 1.4. individual workers' match quality and years of tenure at the firm

3.1.4 Match quality and worker-level wages

Table 1.2 reports the estimates from panel regressions of individual workers' wages on our two worker-level measures of match quality, namely, *eJAQ* and *epJAQ*. Column 1 of Panel A reports the estimate of the regression coefficient of *eJAQ* in a regression that includes only job and year effects and the machine learning variables. The resulting estimate is 0.0389: a worker allocated to her most suitable job (eJAQit = 1) is estimated to earn 3.89% more than a mismatched worker with the same characteristics or with the same job. The estimate is very similar upon controlling for 2-digit industry dummies and firm characteristics (column 2), decreases by more than half of its value upon considering only within-worker variation in *eJAQ* (column 3) and controlling for unobserved heterogeneity across firms (column 4): the estimated effect of match quality on the wage becomes 1.6% in a specification including worker, jobs and year effects, but is highly statistically significant.

The table also presents estimates of the correlation between *epJAQ* and labor earnings in Panel B, providing a robustness check of the results obtained using the *eJAQ* indicator with a continuous measure of workers' suitability for jobs. The estimates shown in Panel B indicate that labor earnings are also positively and significantly correlated with this second measure of job match quality over workers' careers. The 0.1178 coefficient estimate in column 1 indicates that a 10-percentage-point increase in a worker's epJAQ (amounting to half of its standard deviation) is associated with a 1.2% increase in labor earnings. This effect is qualitatively similar and precisely estimated in the specification with industry fixed effects and firm-level controls (column 2), but again halves in value in the specifications with worker fixed effects (column 3) and both worker and firm fixed effects (column 4).

Table 1.2- Match quality and workers' labor income (Sweden)

	Outcome variable: log(annual labor income)			
	(1)	(2)	(3)	(4)
Panel A				
eJAQ	0.0389***	0.0388***	0.0153***	0.016***
	(0.0004)	(0.0004)	(0.0004)	(0.001)
Panel B				
epJAQ	0.1178***	0.1349***	0.0624***	0.079***
	(0.0009)	(0.0010)	(0.0015)	(0.004)
N	9,865,810	9,865,810	9,865,810	9,865,810
Occupation FEs	yes	yes	yes	Yes
Year Fes	yes	yes	yes	Yes
Worker X	yes	no	no	No
Firm X	no	yes	no	No
Worker Fes	no	no	yes	Yes
Firm Fes	no	no	no	Yes

3.1.5 Match quality and firm-level characteristics

We now explore how firm-level measures of match quality, namely, firms' fraction of well-matched employees (*JAQ*) and their average match quality (p*JAQ*) correlate with their characteristics, namely, their age and size, and their employees' age, experience, tenure, and education. Figures 1.5 and 1.6 show that both firm-level measures of match quality are strongly and positively correlated with firm age: older firms are more established, and therefore are more likely to have developed better personnel management systems, and to attract the best talent, which may be easier to match to jobs.

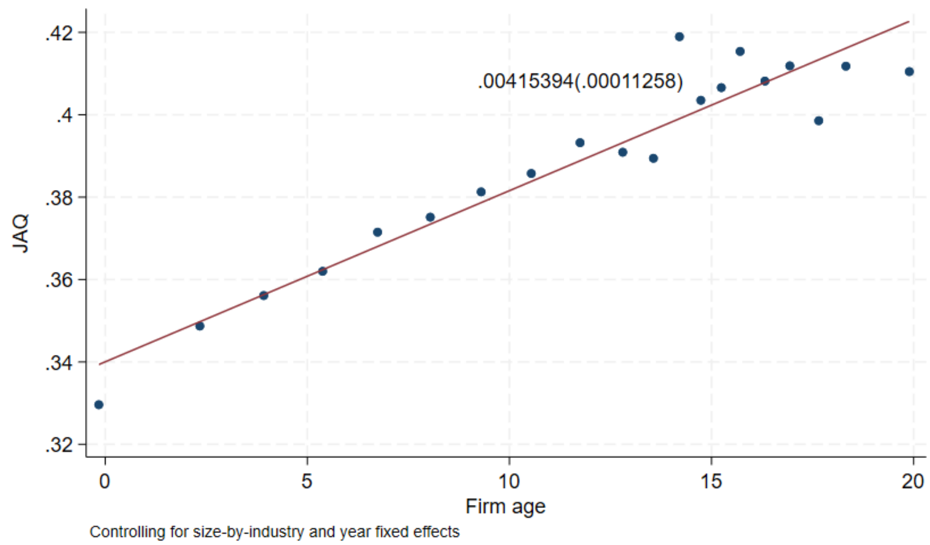

Figure 1.5. Firm-level fraction of well-matched employees (*JAQ*) and firm age

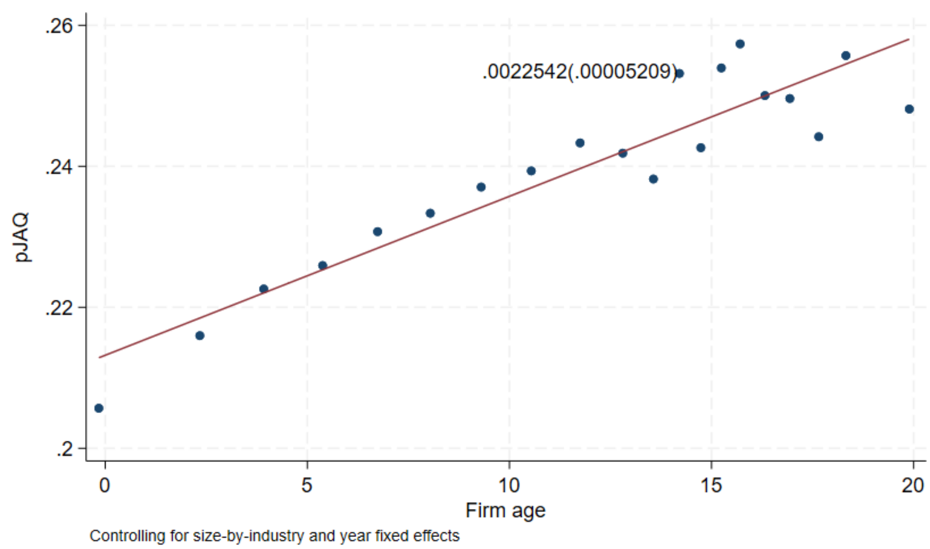

Figure 1.6. Firm-level average match quality (p*JAQ*) and firm age

Figures 1.7 and 1.8 indicate that match quality exhibits a strong positive correlation with the average age of firms' employees up to age 50, after which match quality drops. This is consistent with the positive relationship between match quality and age at the individual worker level seen in Figure 1.2.

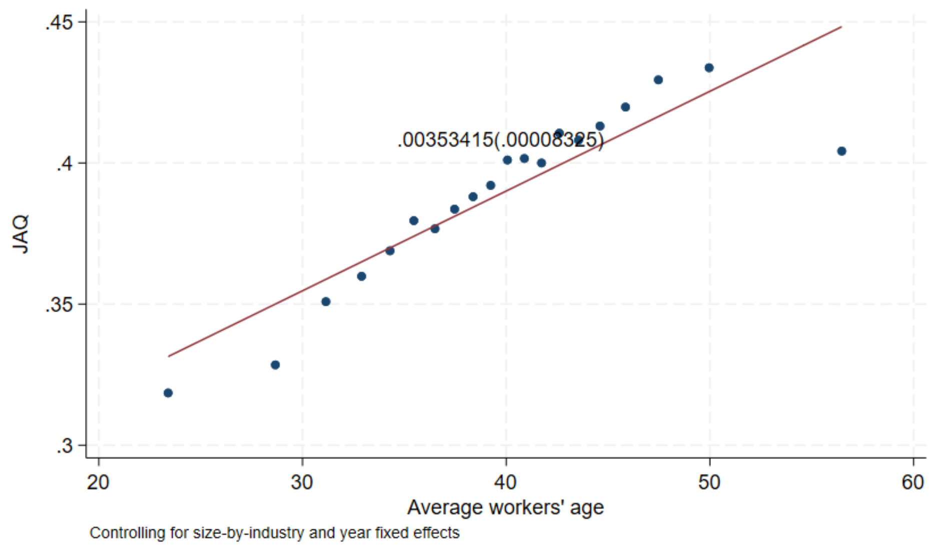

Figure 1.7. Firm-level fraction of well-matched employees (*JAQ*) and average employees' age

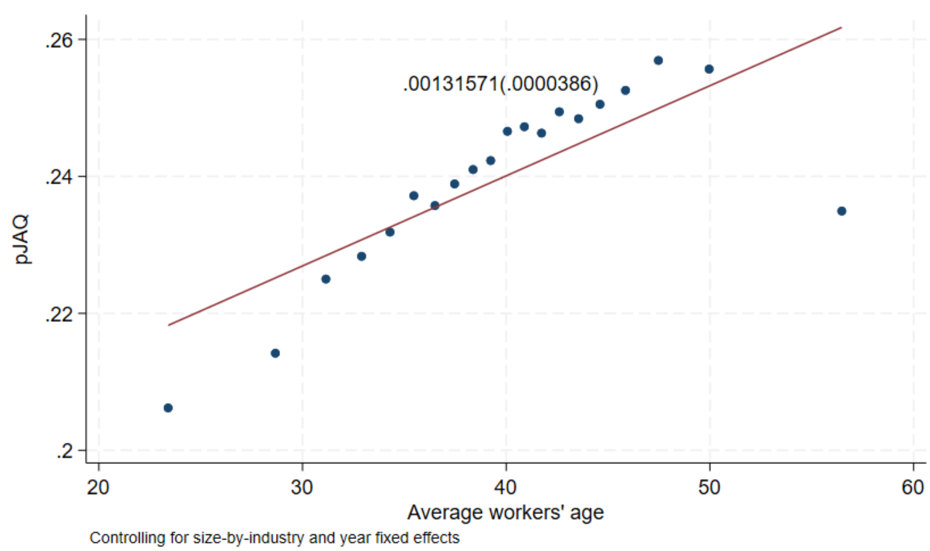

Figure 1.8. Firm-level average match quality (p*JAQ*) and average employees' age

Figures 1.9 and 1.10 show that match quality is steeply increasing with the number of employees up to a critical threshold of about 30 employees, after which it plateaus, suggesting that efficient personnel management systems have a minimum scale.

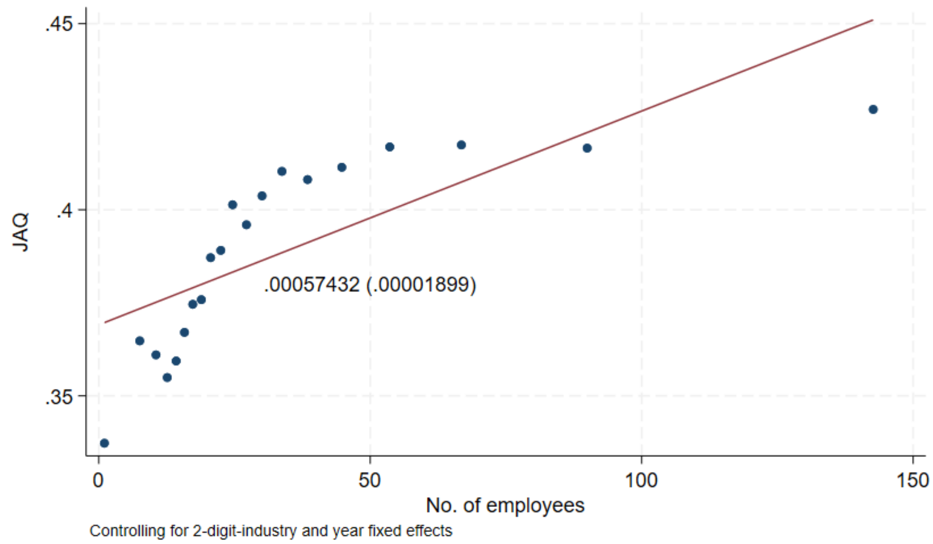

Figure 1.9. Firm-level fraction of well-matched employees (*JAQ*) and number of employees

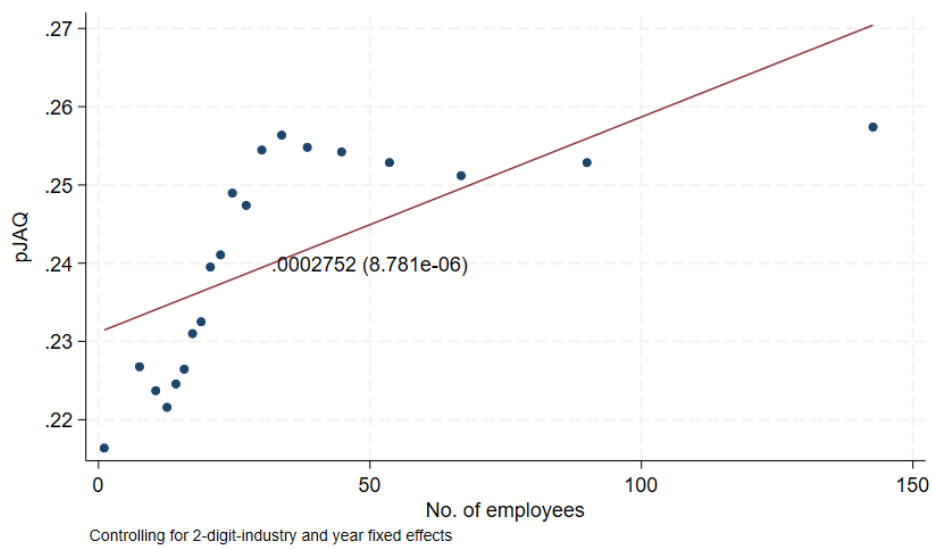

Figure 1.10. Firm-level average match quality (p*JAQ*) and number of employees

Figures 1.11 and 1.12 indicate that match quality increases with the share of employees with a college degree, suggesting that more meritocratic firms attract a more educated workforce.

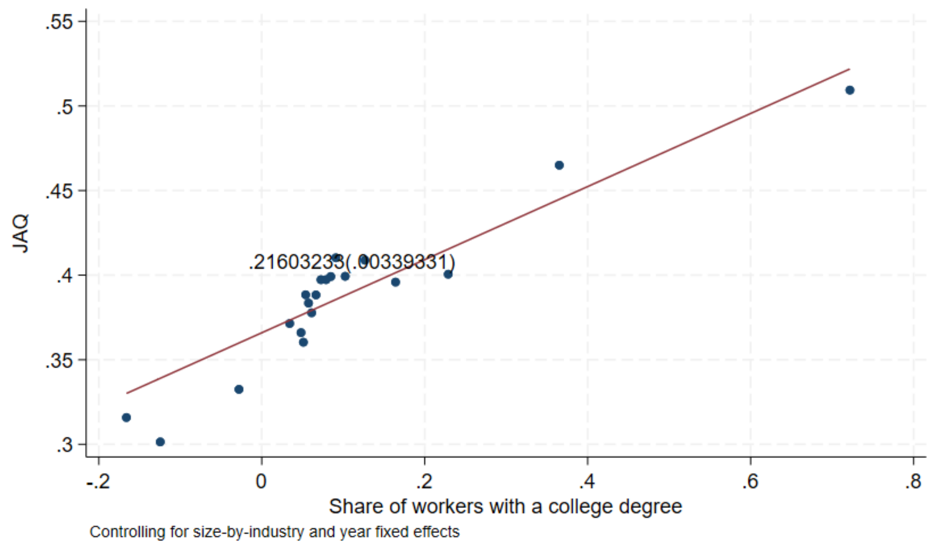

Figure 1.11. Firm-level fraction of well-matched employees (*JAQ*) and share of employees with college degree

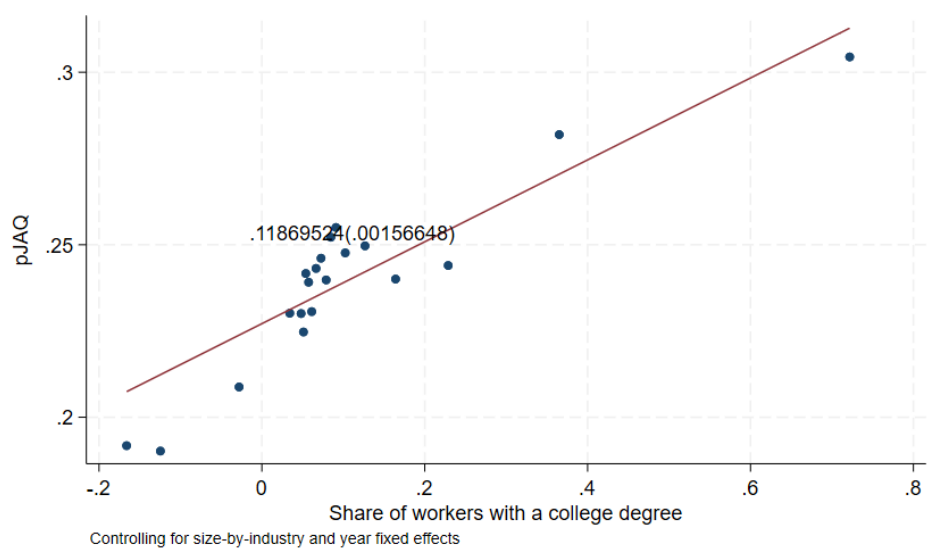

Figure 1.12. Firm-level average match quality (p*JAQ*) and share of employees with college degree

Figures 1.13 and 1.14 show that match quality is strongly positively correlated with the average experience of firms' employees over their first 30 years, after which match quality drops. This is consistent with the relationship observed at the individual worker level in Figure 1.3.

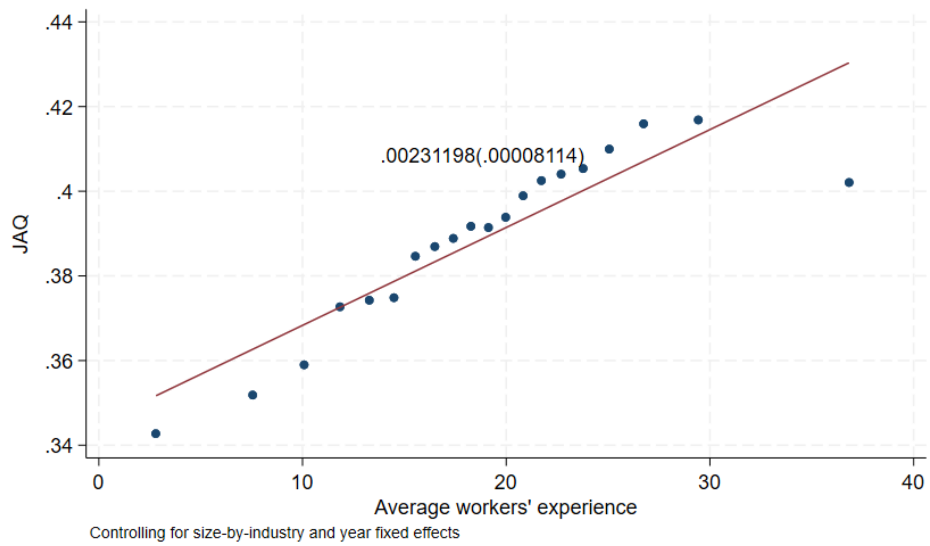

Figure 1.13. Firm-level fraction of well-matched employees (*JAQ*) and average employees' experience

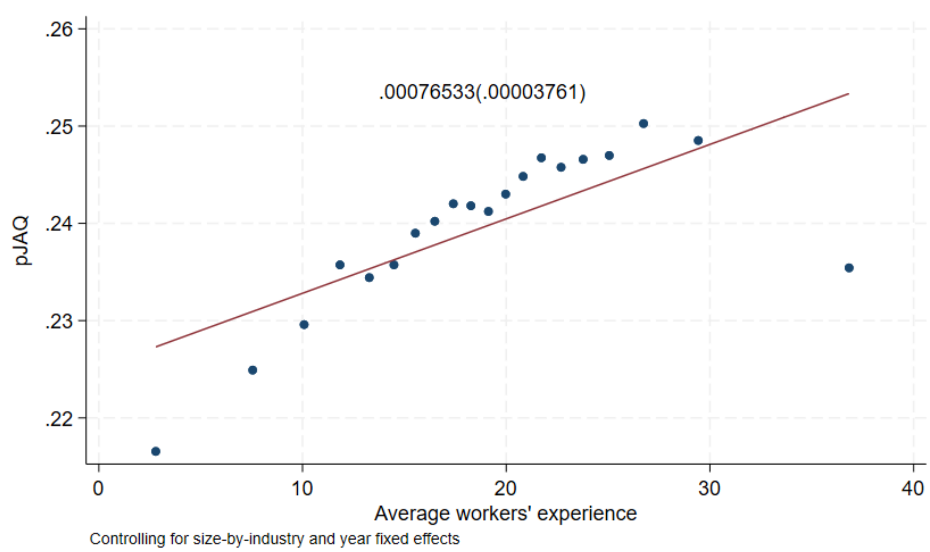

Figure 1.14. Firm-level average match quality (p*JAQ*) and average employees' experience

Figures 1.15 and 1.16 show that match quality is strongly positively correlated with the average tenure of firms' employees, especially in the first years of job tenure, and plateaus after 10 years. This is also consistent with the relationship observed at the individual worker level in Figure 1.4.

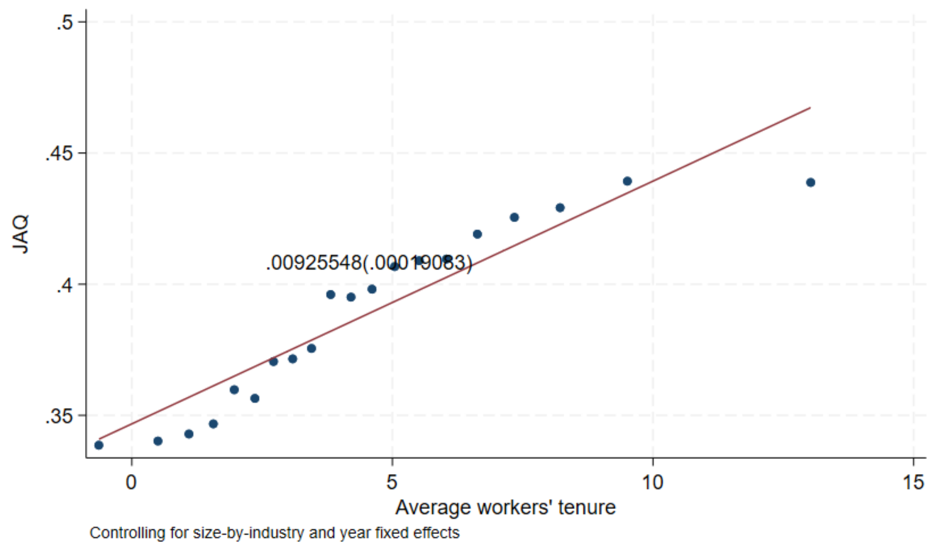

Figure 1.15. Firm-level fraction of well-matched employees (*JAQ*) and average tenure of employees

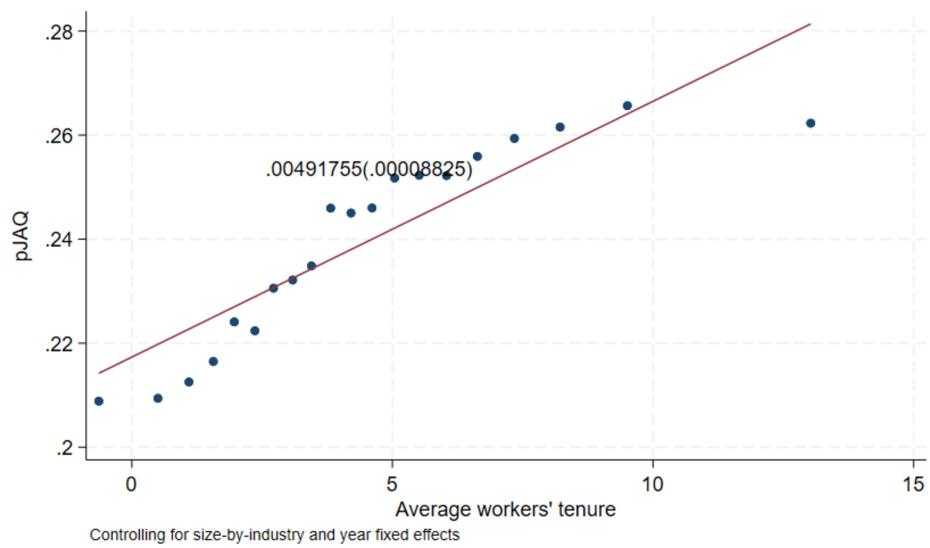

Figure 1.16. Firm-level average match quality (p*JAQ*) and average tenure of employees

3.1.6 Match quality and firm-level outcomes

Figures 1.17 and 1.18 illustrate the relationships between JAQ and firm-level productivity (as measured by the logarithm of value-added per employee) in the main sample. They show that firm-level productivity, as measured by value added per employee, correlates positively with JAQ across firms: it shows partial regression plots of value added per employee against this measure of job-worker match quality, conditioning on year effects and 2-digit industry effects. Figures 1.19 and 1.20 show that the relationship with profitability is instead weak and non-monotonic, being initially decreasing and then mildly increasing.

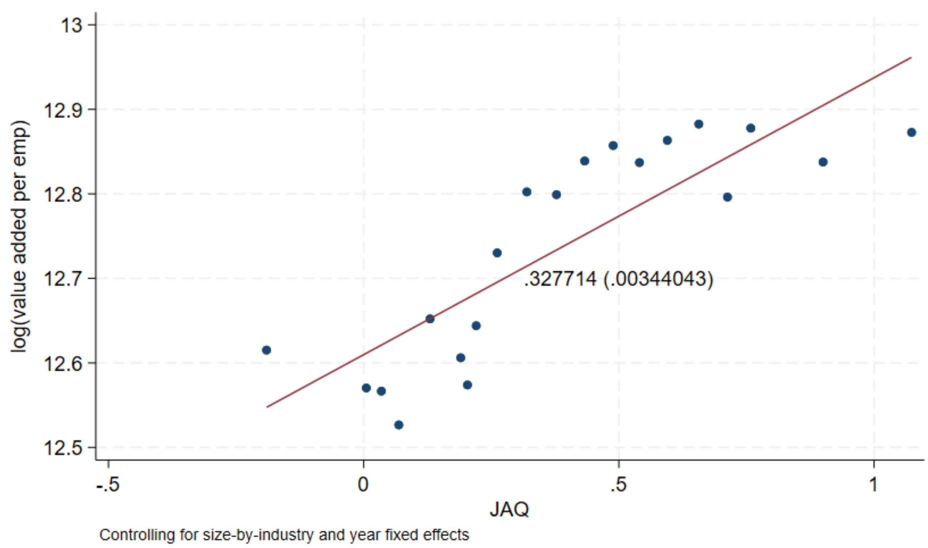

Figure 1.17. Firm-level fraction of well-matched employees (*JAQ*) and firm productivity (logarithm of value added per employee)

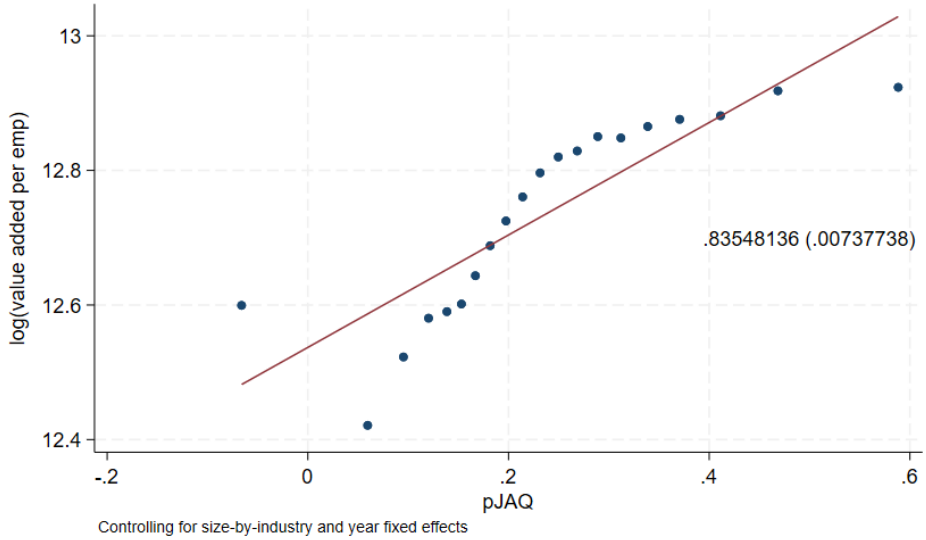

Figure 1.18. Firm-level average match quality (p*JAQ*) and firm productivity (logarithm of value added per employee)

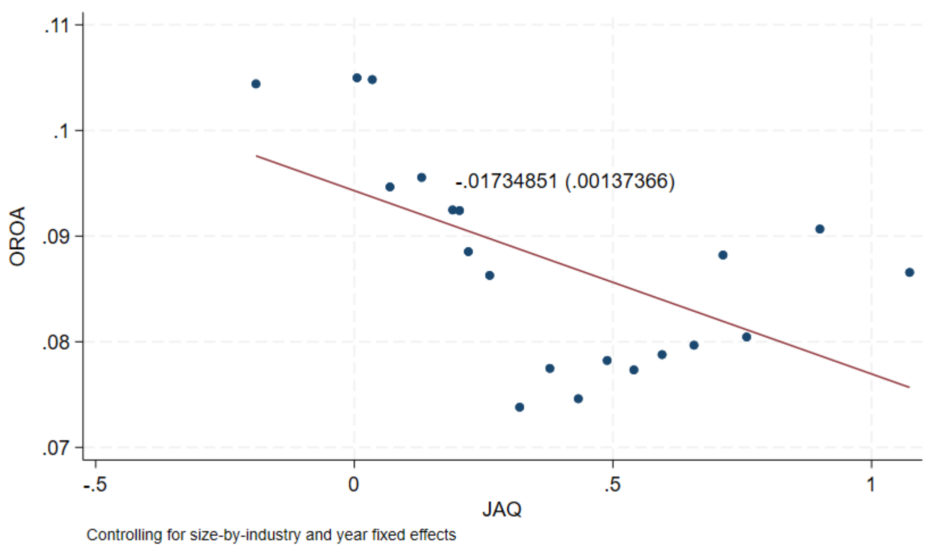

Figure 1.19. Firm-level fraction of well-matched employees (*JAQ*) and firm profitability (ROA)

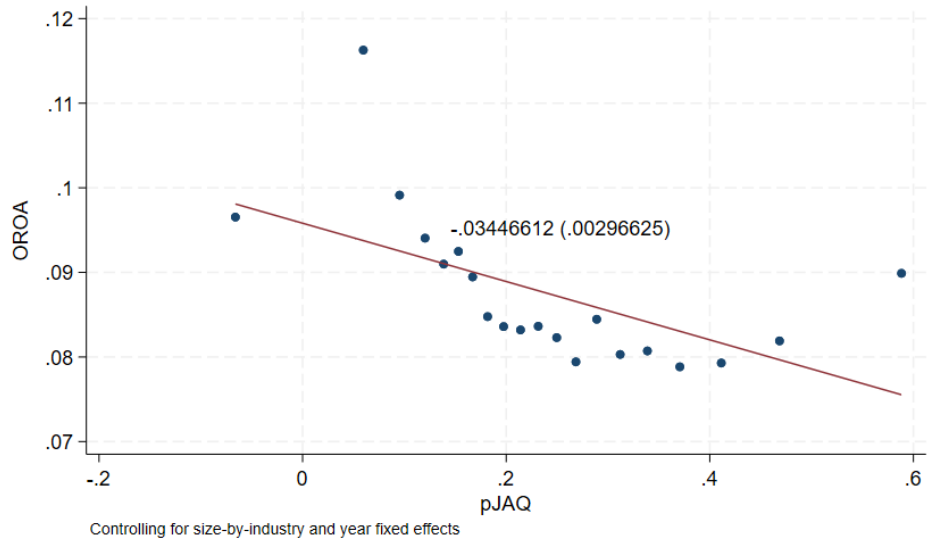

Figure 1.20. Firm-level average match quality (p*JAQ*) and firm profitability (ROA)

Table 1.3 explores further the firm-level correlation between productivity (as well as profitability) and JAQ, controlling for other determinants of productivity. All specifications presented in the table include year and municipality dummies: year dummies control for aggregate movements in productivity, while municipality dummies control for productivity differentials across locations. The latter may arise not only from location-related technological advantage but also from access to deeper and more diversified local labor markets. Hence, the relationship between productivity and JAQ captured by our estimates is not driven by differences in the availability of workers or labor market conditions across firms' locations. In Panel A of Table 1.3, column 1 reports the OLS estimates from a regression of log sales per employee on JAQ, which includes only year dummies. We find a highly significant coefficient of 0.210, implying that a 10-percentage-point increase in JAQ is associated with a 2.10% increase in sales per employee.

In column 2 of Table 1.3, the dependent variable is the log of value added per employee, and the coefficient of *JAQ* is again positive and highly significant: a 10-percentage-point increase in JAQ is associated with an average increase in value added per employee of 2.82%.

Table 1.2 -Firm productivity, profitability, and fraction of well-matched employees (JAQ) (Sweden)

PANEL A	(1) log(sales/emp)	(2) log(va/emp)	(3) OROA	(4) log(sales/emp)	(5) log(va/emp)	(6) OROA
JAQ	0.210***	0.282***	-0.011***	0.047***	0.107***	-0.000
	(0.010)	(0.005)	(0.002)	(0.005)	(0.004)	(0.002)
log(K/L)				0.794***	0.309***	-0.025***
				(0.004)	(0.003)	(0.001)
log(L)				-0.004	-0.010***	-0.006***
				(0.002)	(0.002)	(0.001)
Sh. w/ college				0.229***	0.191***	-0.018***
				(0.010)	(0.009)	(0.004)
Industry	No	No	No	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
PANEL B						
JAQ	0.263***	0.155***	-0.012***	0.049***	0.070***	-0.005**
	(0.010)	(0.006)	(0.002)	(0.006)	(0.005)	(0.002)
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Occ.	Yes	Yes	Yes	Yes	Yes	Yes
PANEL C						
JAQ	0.195***	0.114***	-0.010***	0.036***	0.049***	-0.004
	(0.010)	(0.006)	(0.002)	(0.006)	(0.005)	(0.002)
Group-by-yr FEs	Yes	Yes	Yes	Yes	Yes	Yes
Occupations	Yes	Yes	Yes	Yes	Yes	Yes
Workers X	Yes	Yes	Yes	Yes	Yes	Yes
Firm X	No	No	No	Yes	Yes	Yes
PANEL D						
JAQ	-0.002	0.015***	-0.002	-0.003	0.011**	-0.004*
	(0.006)	(0.005)	(0.002)	(0.006)	(0.005)	(0.002)
Group-by-yr	No	No	No	Yes	Yes	Yes
Firm and yr FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	258887	258993	258993	258885	258991	258991
No. Firms	60543	60566	60566	60543	60566	60566
y Mean	12.454	12.737	0.088	12.454	12.737	0.088
y St. Dev.	1.108	0.622	0.225	1.108	0.622	0.225

In panels B, C and D of the table, we control for various possible sources of omitted variable bias, namely, firm characteristics, differences in firms' occupation structures, and differences in workers' quality across firms.

First, the correlation between productivity and our measures of job allocation quality is robust to the inclusion of 2-digit industry indicators, log number of employees, log capital, and the fraction of employees with at least a college degree among the regressors, as shown by the estimates in columns 4 and 5 of the table. The estimated coefficients of JAQ in columns 4 and 5 drop in magnitude, but remain positive and significantly different from zero.

A second possible concern in the previous specifications is that the firms being compared may have different occupation structures. Two otherwise comparable firms may structure their internal hierarchy differently: if, for instance, a firm has an inefficiently large number of managerial positions relative to technical ones compared to other firms in its industry, and those managerial positions are harder to fill with suitable employees, it is likely to end up both with lower productivity and lower JAQ, creating a spurious correlation between the two variables. Hence, in Panel B we show a specification where we control for firms' occupation structure (the fraction of workers in firm f assigned to job j in year t) and firm controls (age, family firm status, state-owned status, listed firm status, an indicator for the presence of a human resources manager, its log number of employees and its log of total assets).

A third possible source of omitted variable bias is that firms with higher *JAQ* may feature higher-quality workers, irrespective of the job they are allocated to. Thus, in Panel C, we augment the specification by including workers' characteristics used in the machine learning algorithm, averaged across all workers employed in firm *f* in year *t*. Specifically, Panel C adds controls for workers' age, gender, an indicator for immigrant status, residence municipality, a mobility indicator equal to one for workers employed in a county different from the county of birth, education level (basic, high school, vocational, or university), education subject (no specialization, law, business and economics, health and medicine, natural sciences, teaching, engineering, social sciences, services, or other specializations), labor market experience (measured as years since graduation), tenure at the current firm, number of firms and number of two-digit industries where an individual previously worked, total number of unemployment days since 1992 (when the unemployment data starts in

LISA), years of experience in each occupation, years of experience in each 2-digit industry, and years of experience in each decile of the distribution of firms' number of employees or total assets. In Panel D, we also include firm and year fixed effects, as well as interactions between an indicator for firm size (small, medium, or large), industry, and year effects, so that effectively the coefficients are estimated within each of the industry-size bins used to estimate the machine-learning algorithm.

Hence, the results obtained in Panels B, C and D are qualitatively similar to those in Panel A: the estimated coefficients of *JAQ* drop in magnitude but remain positive and statistically significant in columns 1, 2, 4, and 5. The result that productivity correlates positively and significantly with JAQ is robust even to the inclusion of firm and year fixed effects, as shown in Panel D of the table.

In almost all specifications shown in Table 1.3, profitability, as measured by operating return on assets (OROA), is either not significantly related or negatively correlated with our measure of efficient job allocation, as shown in columns 3 and 6. A possible interpretation of this finding is that, in Swedish firms, the productivity gains from better job-worker matches translated mostly into higher wages rather than into increases in firm profitability.

3.1.7 Time series patterns in match quality

Figures 1.22 and 1.23 show the time series of aggregate match quality in Sweden from 2002 to 2012, obtained by averaging the firm-level *JAQ* and *pJAQ* measures across firms for each year in our sample. Both figures show that the aggregate match quality is significantly higher in the learning sample than in the main sample, as expected: it is about twice as large in the former as in the latter. Neither time series features an appreciable trend. However, aggregate match quality displays some cyclicality (consistent with the findings by Baley et al., 2022, and Bowlus, 1995), especially in the learning sample: it rose slightly from the early to mid-2000s, when Sweden experienced a robust expansion, with declining unemployment and strong productivity growth, among the highest in the OECD; conversely, in the learning sample match quality declined in 2008-09 when the country was significantly impacted by the global financial crisis, as the fall in external demand was exacerbated by the high exposure of Sweden's concentrated banking system to real estate in the Baltic states, and by its high levels of household debt and overvalued housing. This external shock led to a significant, widespread decline in economic activity, marking a sharp recession. In the 2010-12 recovery, match quality stabilized in the learning sample and recovered slightly in the main sample, settling around 2-2.25% growth rates by 2012.

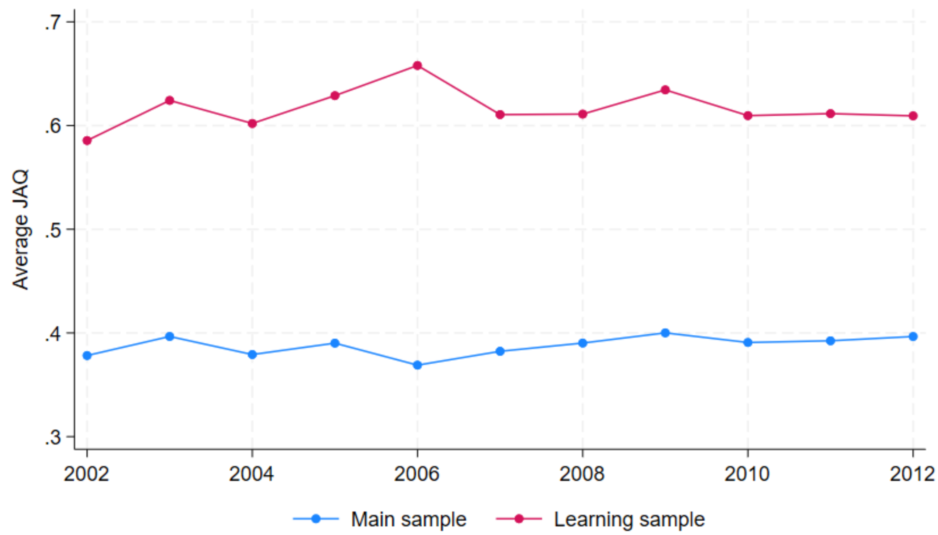

Figure 1.21. Aggregate fraction of well-matched employees (*JAQ*), in the main and in the learning samples, 2002-2012

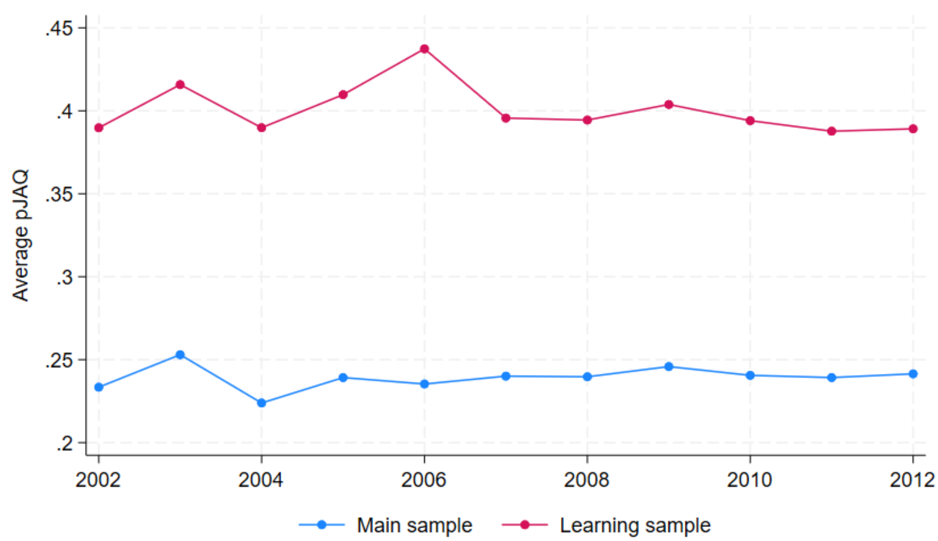

Figure 1.22. Aggregate average match quality (p*JAQ*), in the main and in the learning samples, 2002-2012

3.2 Portugal

This section describes the data and presents the results regarding our worker-level and firm-level measures of match quality obtained from the administrative worker-firm matched data for Portugal.

3.2.1 Data sources

To develop and estimate the JAQ measure for Portugal, we rely on administrative data provided by Statistics Portugal (Instituto Nacional de Estatística, INE). The Portuguese data are particularly well suited for our purposes, as they allow us to observe workers and firms longitudinally, together with detailed information on occupations, wages, education, and firm characteristics.

Our analysis combines two main data sources. First, we use Quadros de Pessoal (QP), a matched employer–employee database that covers the universe of private-sector employees in Portugal. Quadros de Pessoal is based on a mandatory yearly survey collected by the Ministry of Labour and includes detailed information on workers' demographic characteristics, education, occupation, wages, contractual arrangements, and firm identifiers. We observe worker-level data starting from 1986, which allows us to reconstruct long individual job histories and labor market trajectories.

Second, we use firm-level balance-sheet information from the Sistema de Contas Integradas das Empresas (SCIE), which covers the universe of Portuguese firms from 2004 onward. For earlier years, firm-level accounting information is drawn from the Inquérito às Empresas Harmonizado (IEH), available from 1990 to 2004. The SCIE database provides comprehensive information on firms' financial statements, including value added, sales, assets, capital intensity, profitability, leverage, liquidity, investment, and R&D expenditures.

Although worker data are available from 1986, our primary analysis focuses on the 2004–2022 period. Prior to 2004, differences in firm identifiers and classification systems would make it difficult to reliably match workers in Quadros de Pessoal to firms observed in the accounting data. Nevertheless, we exploit the pre-2004 worker data to construct rich measures of past labor market experience and employment histories, which enter the estimation of workers' suitability for different jobs.

The set of worker-level variables used to estimate job–worker match quality closely mirrors those employed in the Swedish analysis. Background characteristics include age, gender, immigrant status, and region of residence. Educational information is particularly detailed in the Portuguese data: we observe education level and field of study at a fine level of disaggregation (up to four digits). Work experience is captured through several dimensions, including tenure at the current firm, cumulative labor market experience (measured as the number of years the worker is observed in the dataset), and detailed occupational histories. Occupational classifications are administrative records directly reported by firms and harmonized over time.

Firms are characterized by age, size (measured by employment), sector, and a wide set of financial variables drawn from balance sheets and income statements. Sector classifications are harmonized across years, and our analysis excludes public administration, education, health, and defense, focusing instead on the private-sector economy where occupational assignment and internal job allocation are most relevant.

To ensure reliable estimation of the job–worker matching rule, we restrict attention to firms whose median employment over the sample period is at least 10 workers. Firms below this threshold tend to display sparse occupational structures, which would introduce substantial noise into the estimation of match quality. After applying these filters, our sample includes 69,802 distinct firms observed over the 2004–2022 period, employing on average 1,224,459 workers per year.

3.2.2 Descriptive statistics

In this section, we present descriptive statistics on the characteristics of the workers in our sample, and how their match quality differs depending on these characteristics in 2022, using a total of 1,264,222 observations. The tables in this section report the share of workers in each category (computed relative to the full sample size) and the fraction of well-matched workers (i.e., the mean of *eJAQ*) and the average match quality (i.e., the mean of *epJAQ*) in each category.

Table 2.1 shows that in Portugal male employees are 60% of the total population and are more frequently well-matched, based on both our worker-level match quality measures. The fraction of immigrants in Portugal is slightly less than 10%, and they are about as well-matched as natives, in contrast to findings based on other mismatch measures, which indicate that they are more frequently mismatched than natives. Interestingly, match quality appears to correlate with employees' contract type: permanent workers, who account for slightly more than 2/3 of the workforce, have slightly better match quality than holders of fixed-term contracts. This probably reflects the greater work experience of permanent workers, as firms offer permanent contracts only to workers with a proven track record. It may then appear surprising that full-time workers are significantly worse matched than part-time workers. This may, however, reflect the fact that part-time workers are assigned to simple, entry-level jobs, in which there is little scope for mismatch, while this does not apply to full-time workers.

Table 2.2 presents statistics on the sectoral distribution of workers and match quality in the population. Manufacturing, the largest sector in terms of employment (with over 1/3 of the total workforce), features the highest level of mismatch, followed by mining. Workers in information, finance, real estate, consulting, transportation, restaurants, and hotels have intermediate match quality levels. In contrast, employees in the transportation, retail, and especially construction sectors have the best match quality in the economy. This sectoral distribution of match quality probably reflects the much greater diversity of possible jobs that workers can be assigned to in manufacturing than in transportation, retail, and construction, hence the greater scope for labor misallocation in the former than in the latter sectors.

Table 2.3 instead describes the distribution of workers across educational levels and fields, and how match quality varies across these levels and fields. Note that the educational field variable is conditional on having at least a tertiary education, which explains why its sample share is very low. Match quality clearly rises with workers' educational level, and it is highest for graduates in mathematics, physics, and engineering, and lowest for graduates in law, vocational/services, and in the humanities. This suggests that more advanced degrees and more technical training, by leading employees to be more narrowly specialized, enable employers to match them more easily to jobs.

The table also shows horizontal-mismatch categories (defined only for workers with at least tertiary education, but expressing their shares relative to the full sample) and vertical-mismatch categories. Vertical and horizontal mismatch benchmarks are based on the modal category within 2-digit occupation-sector-size cells. Interestingly, workers who are well-matched vertically, and especially horizontally, are also better matched based on *eJAQ* and *epJAQ* measures.

Table 2.1 – Descriptive statistics about workers' characteristics and their match quality in 2022

Variable	In sample (%)	Well matched (*eJAQ*)	Average *epJAQ* (%)
Gender			
Men	60.11	45.65	27.76
Women	39.89	41.22	23.86
Immigrant status			
Native	90.48	43.87	26.11
Immigrant	9.52	43.98	27.02
Contract type			
Fixed-term contract	31.12	42.18	25.47
Permanent contract	68.88	44.65	26.53
Working time arrangement			
Full-time	95.22	43.63	25.84
Part-time	4.50	49.21	33.73

Table 2.2 – Descriptive statistics about workers' sectoral allocation, occupational qualifications, and match quality in 2022

Variable	In sample (%)	Well matched (*eJAQ*)	Average *epJAQ* (%)
Sector			
Construction	10.60	62.74	40.47
Information, Finance, Real estate; Consultancy	13.90	44.23	27.74
Manufacturing	35.79	34.20	19.23
Mining	0.32	34.32	28.39
Transportation, Restauration, Accommodation	17.83	45.47	26.86
Wholesale and Retail	21.57	49.29	29.18
Qualification			
Highly qualified	8.02	48.27	28.50
Intern	3.99	34.64	22.69
Low manager	5.94	39.09	23.98
Medium manager	5.30	49.23	29.89
Qualified	41.62	46.37	27.18
Semi-qualified	19.07	37.35	22.16
Top manager	7.42	50.21	31.61
Unqualified	8.64	41.15	24.55

Table 2.3 – Descriptive statistics about workers' education and match quality in 2022

Variable	In sample (%)	Well matched (*eJAQ*)	Average *epJAQ* (%)
Education level			
Low education	44.45	43.85	26.24
Medium education	35.01	42.47	24.76
High education (tertiary)	16.52	45.40	27.81
Very high education (MSc/PhD)	3.70	51.62	33.01
Education field			
Humanities	1.63	39.57	20.67
Social sciences	1.07	39.73	21.86
Economics & Business	3.10	40.46	24.07
Law	0.28	34.33	19.20
Math / Physics / CS	2.45	61.33	42.25
Engineering	4.98	57.64	36.15
Medicine / Health	0.53	41.47	24.61
Vocational / Services	0.44	35.02	18.38
Vertical mismatch (modal benchmark)			
Under-educated	7.14	39.94	23.63
Well-matched	60.08	46.83	28.66
Over-educated	32.46	39.37	22.30
Horizontal mismatch (tertiary only; modal benchmark)			
Horiz. matched	7.36	60.70	38.06
Horiz. mismatched	5.97	37.80	23.20

.

3.2.3 Match quality and worker-level characteristics

In this section, we describe how our measures of match quality correlate with workers' characteristics. Figures 2.2 to 2.4 show that, as in the Swedish data, our worker-level measure of job allocation quality, *eJAQ*, also increases strongly with workers' age, experience, and firm-level tenure. And, as in Sweden, in all three cases the relationship is non-linear: eJAQ rises steeply in the first years of a worker's life, experience, and job tenure, and then tends to grow much more slowly, or even drop somewhat, for workers over 50.

In particular, the relationship with age in Figure 2.1 shows that workers' allocation quality increases by approximately ¼ (from 39% to 48%) between ages 20 and 30, then effectively flattens between ages 30 and 50, and declines slightly from age 50 to 60. The relationship with job market experience shown in Figure 2.2 and that with job tenure at the firm shown in Figure 2.3 also feature a concave shape, being much steeper in the first 5 years of experience than in subsequent years, as in Sweden: in Portugal, *eJAQ* increases by about 20% over the first 5 years of experience and by 6% over the following 20 years, and by about 25% over the first 5 years of tenure at the firm while remaining almost unchanged over the subsequent 15 years.

As in the case of Sweden, the different age pattern shown in Figure 2.1, where *eJAQ* rises over the first decade rather than over the first 5 years, can be explained by workers taking up their first job at different ages depending on when they complete their education.

These consistent patterns of increasing match quality over the first years of experience and tenure among Portuguese workers suggest that their productivity and wages should also increase over the same period. This is indeed the case, as shown in the following subsection.

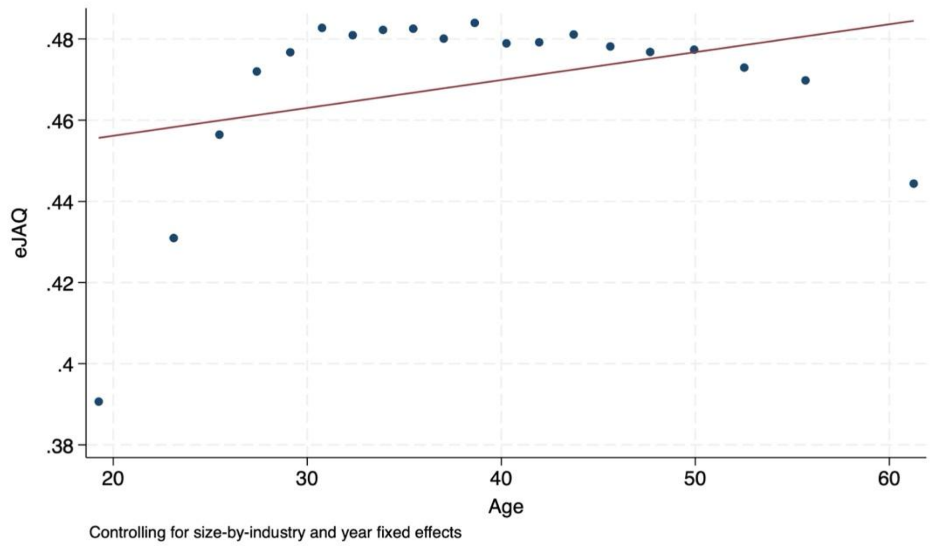

Figure 2.1. individual workers' job match quality and age

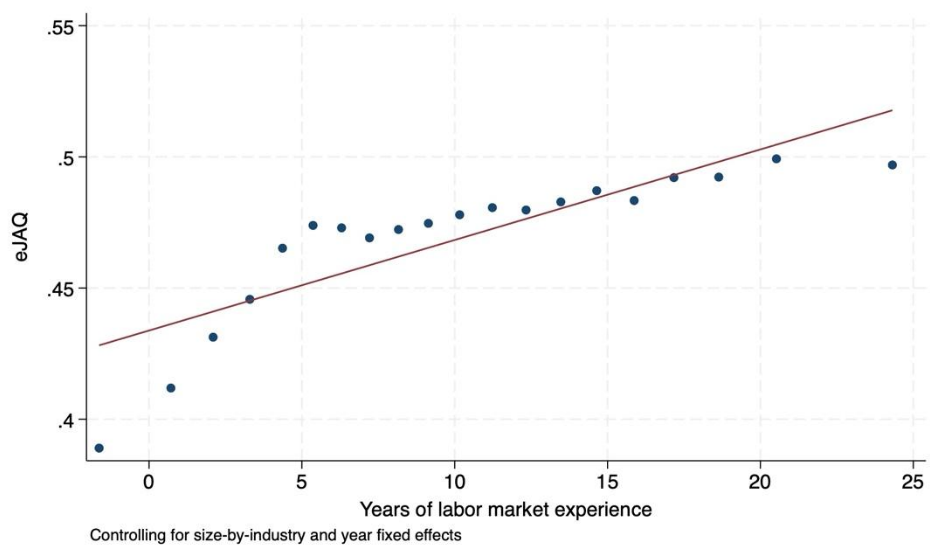

Figure 2.2. individual workers' match quality and labor market experience

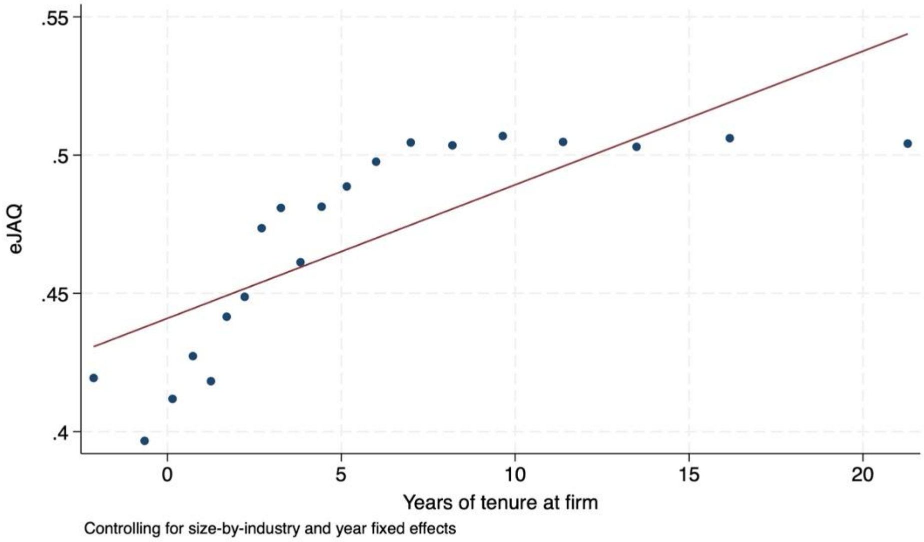

Figure 2.3. individual workers' match quality and years of tenure at the firm

3.2.4 Match quality and worker-level wages

Figures 2.5 and 2.6 show that there is a positive correlation between our two measures of match quality and the logarithm of monthly labor income, for most values of labor income. The relationship appears to be non-linear: the steepest increases in match quality appear to occur in the intermediate labor income region, and the correlation becomes negative in the top labor income region, possibly because the best-paid jobs are the hardest to fill with the right workers.

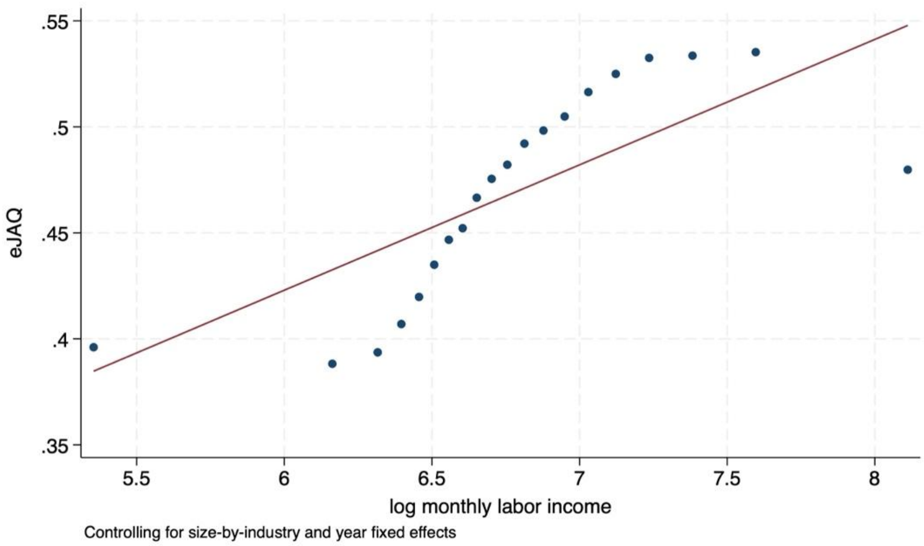

Figure 2.4. ***eJAQ*** **and log monthly labor income**

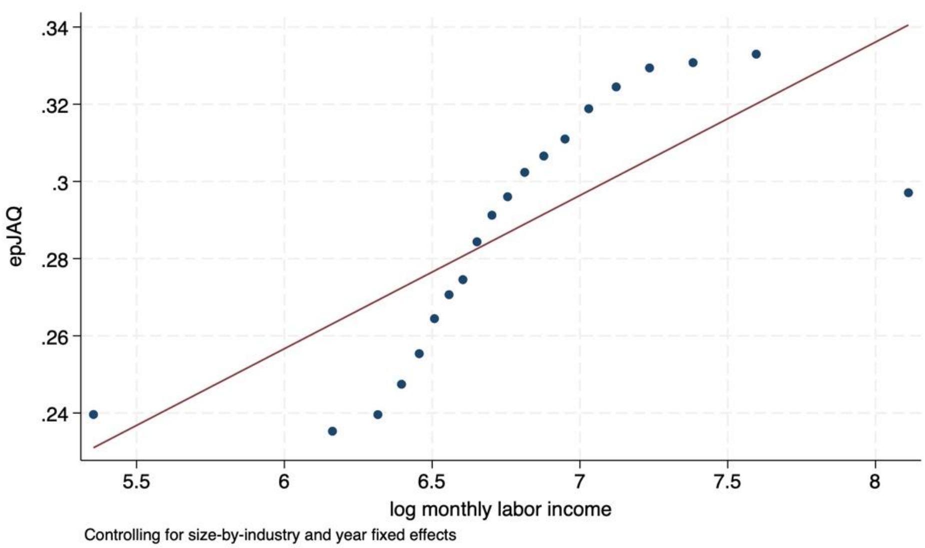

Figure 2.5. ***epJAQ*** **and log monthly labor income**

3.2.5 Match quality and firm-level characteristics

We now explore how firm-level measures of match quality correlate with firm characteristics. Figures 2.7 and 2.8 show that both firm-level measures of match quality are positively correlated with firm size, as measured by the logarithm of their number of employees.

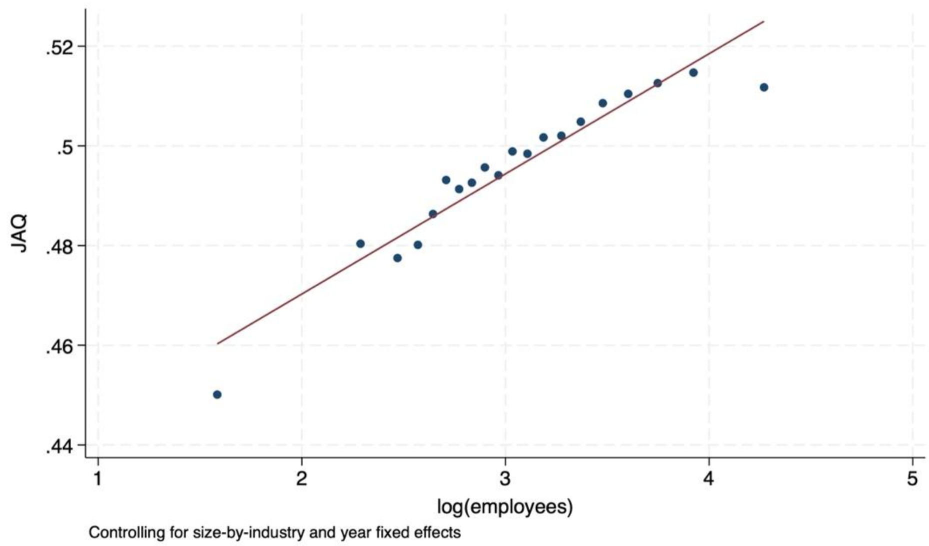

Figure 2.6. Firm-level fraction of well-matched employees (*JAQ*) and number of employees

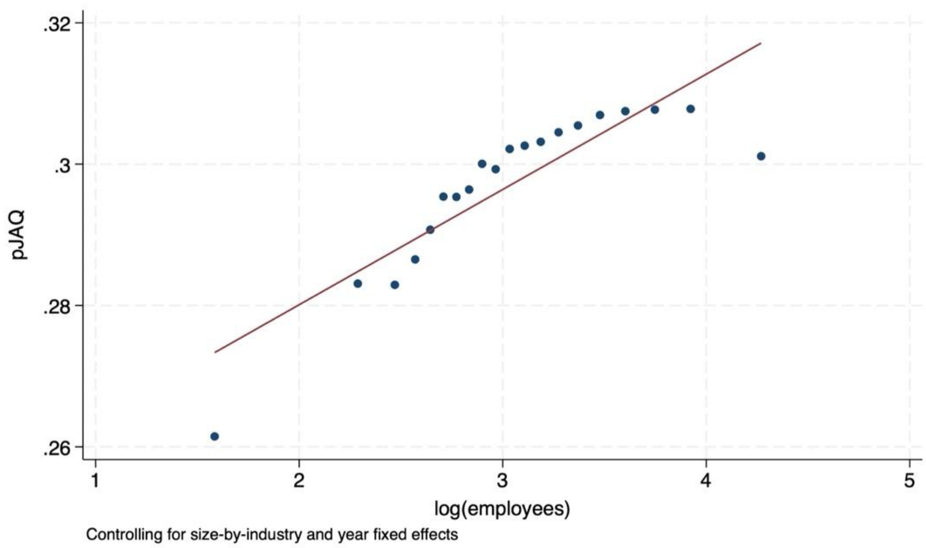

Figure 2.7. Firm-level average match quality (p*JAQ*) and number of employees

The following figures illustrate the relationship between match quality and firm-level employee characteristics. Again, the results are in line with those based on Swedish data. Figures 2.9 and 2.10 show that match quality is strongly and positively correlated with the average age of employees.

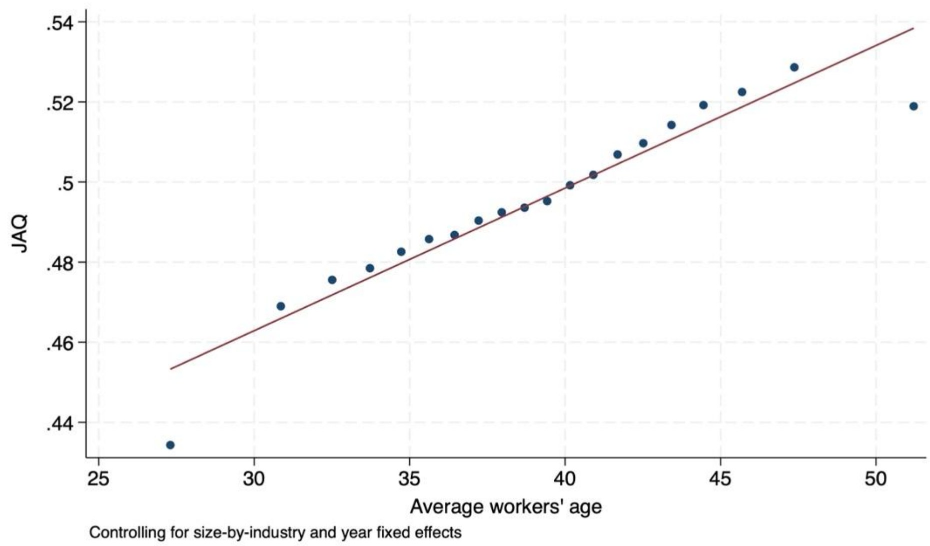

Figure 2.8. Firm-level fraction of well-matched employees (*JAQ*) and average employees' age

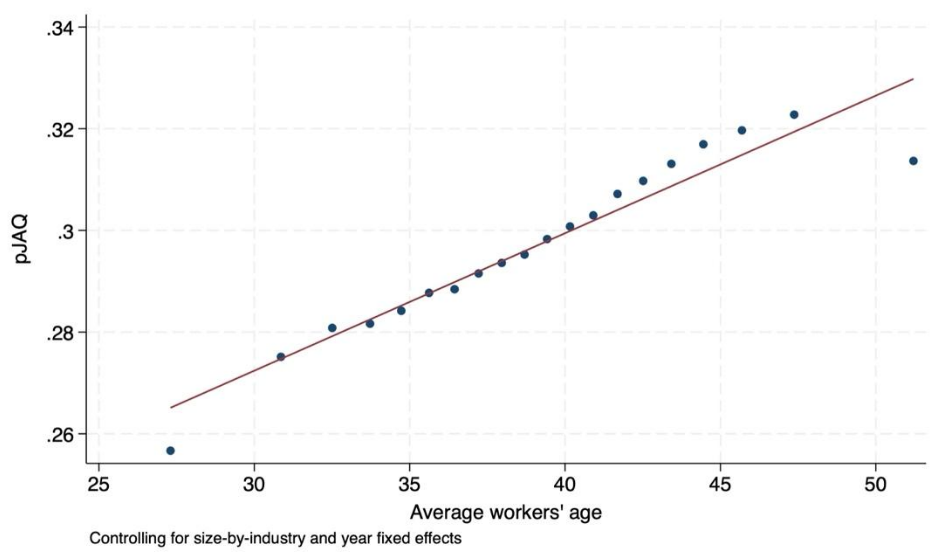

Figure 2.9. Firm-level average match quality (p*JAQ*) and average employees' age

Figures 2.11 and 2.12 show that match quality is strongly and positively correlated with employees' human capital, as measured by the fraction of employees with tertiary education.

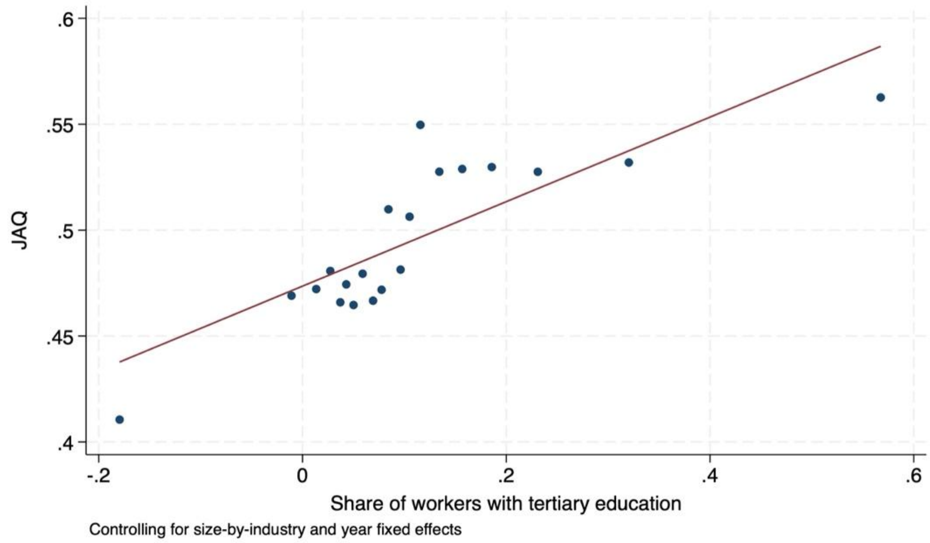

Figure 2.10. Firm-level fraction of well-matched employees (*JAQ*) and share of employees with tertiary education

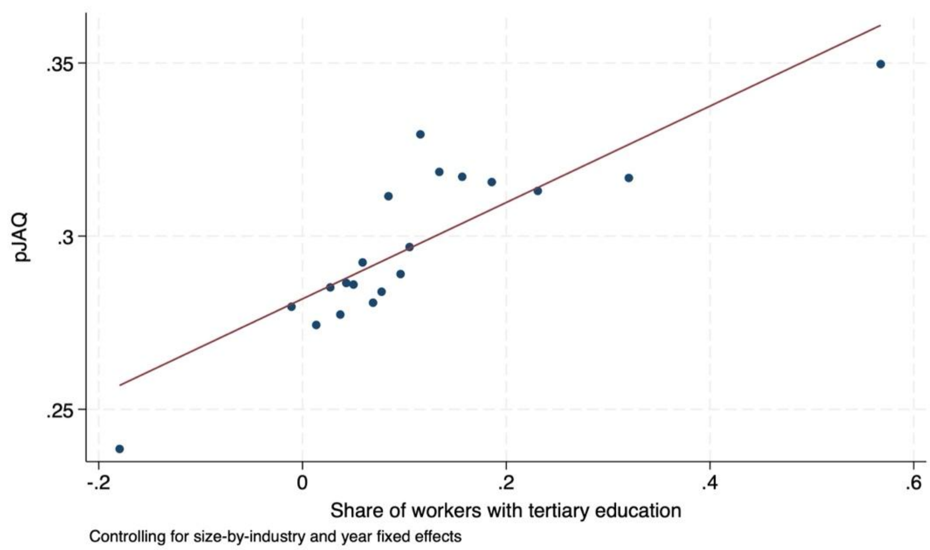

Figure 2.11. Firm-level average match quality (p*JAQ*) and share of employees with college degree

Figures 2.13 and 2.14 show that match quality is strongly and positively correlated with employees' work experience within the relevant firm, as measured by their average tenure. The relationship is strongest in the first 5 years of tenure, consistent with the worker-level results in Figure 2.4.

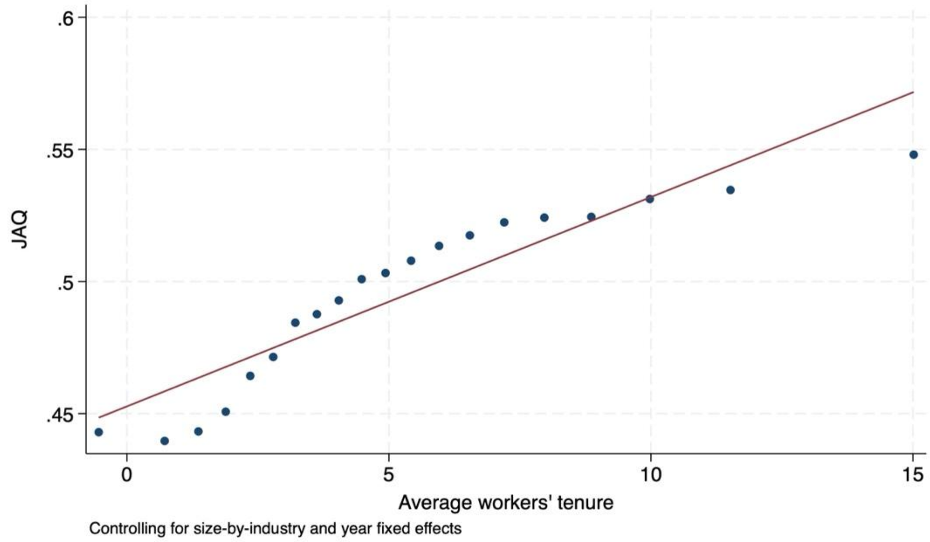

Figure 2.12. Firm-level fraction of well-matched employees (*JAQ*) and average tenure of employees

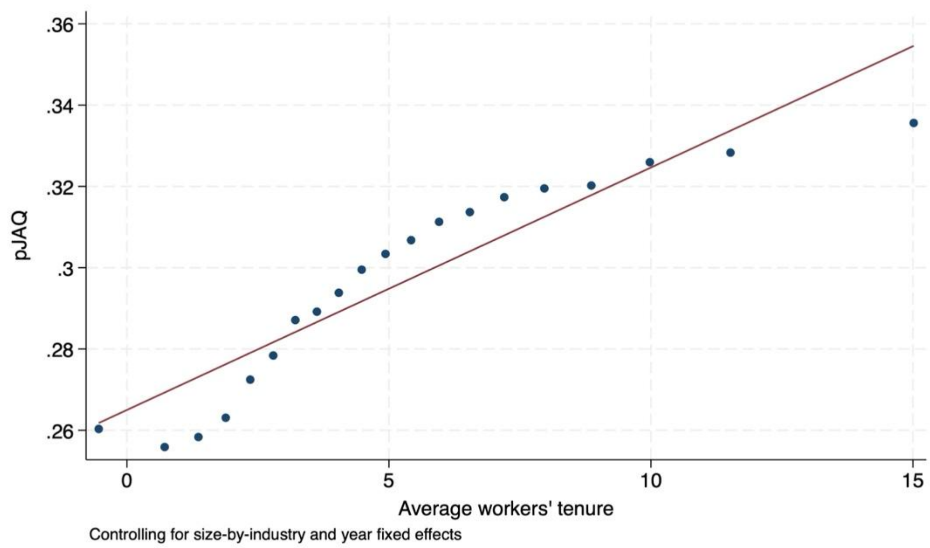

Figure 2.13. Firm-level average match quality (p*JAQ*) and average tenure of employees

The following figures explore relationships between match quality and characteristics not explored in Swedish data. Figures 2.15 and 2.16 show that match quality increases with capital intensity at low levels, but decreases at high levels. Hence, investment in personnel management complements capital investment, but very capital-intensive technologies require skills that are harder to find.

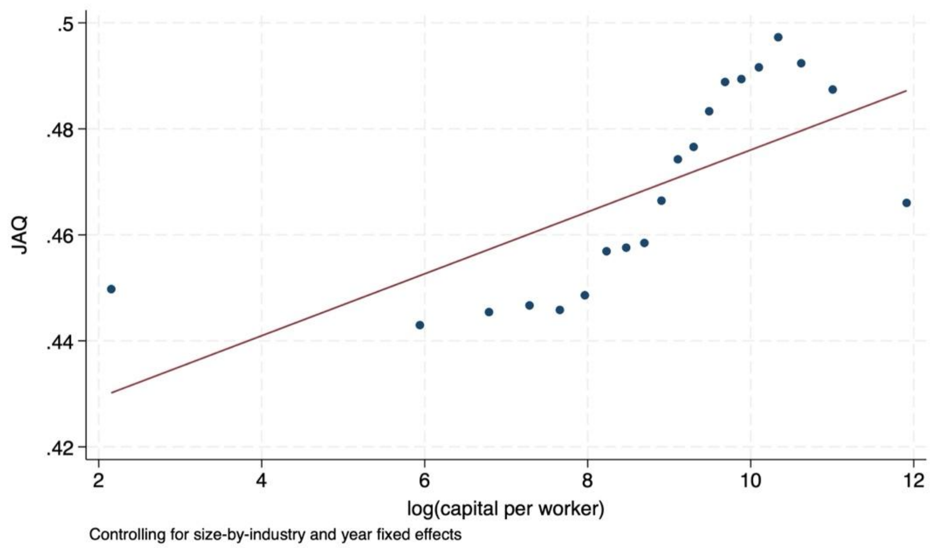

Figure 2.14. Firm-level fraction of well-matched employees (*JAQ*) and logarithm of capital-to-labor ratio (K/L)

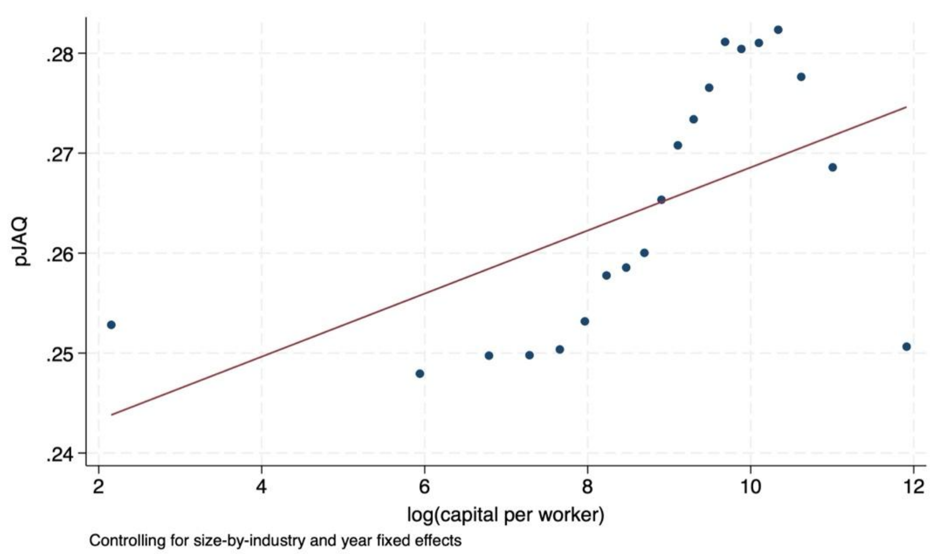

Figure 2.15. Firm-level average match quality (p*JAQ*) and logarithm of capital-to-labor ratio (K/L)

Figures 2.17 and 2.18 show that firm-level match quality increases with the average firm-level wage, consistent with the regression results for Sweden in Table 1.2 and with the idea that better matches map into higher productivity, which in turn translates into higher wages. Interestingly, as in Figures 2.5 and 2.6, the steepest increases in match quality appear to occur in the intermediate wage region.

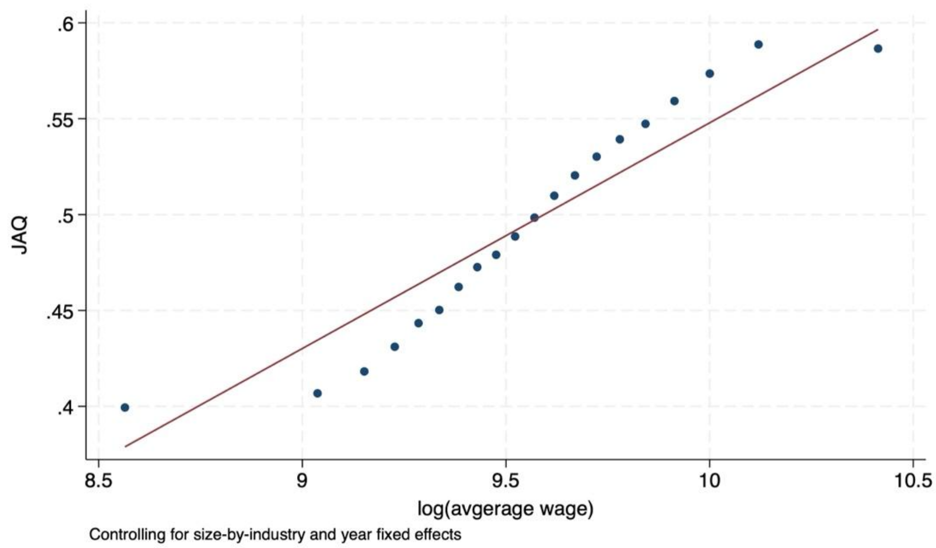

Figure 2.16. Firm-level fraction of well-matched employees (*JAQ* and logarithm of average wage paid (wage bill over number of employees)

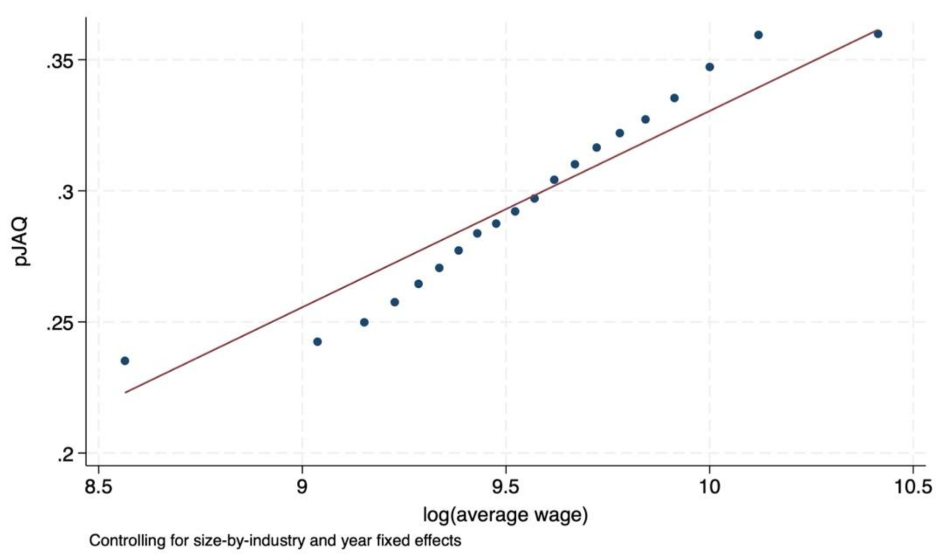

Figure 2.17. Firm-level average match quality (p*JAQ*) and logarithm of average wage paid (wage bill over number of employees)

Figures 2.19 and 2.20 investigate an aspect for which we have no evidence from the other three countries, namely, the relationship between firm-level match quality and the logarithm of firms' training per worker. Interestingly, the relationship is mostly positive and strong (except for the highest values of training per worker), suggesting a beneficial role of training in improving match quality.

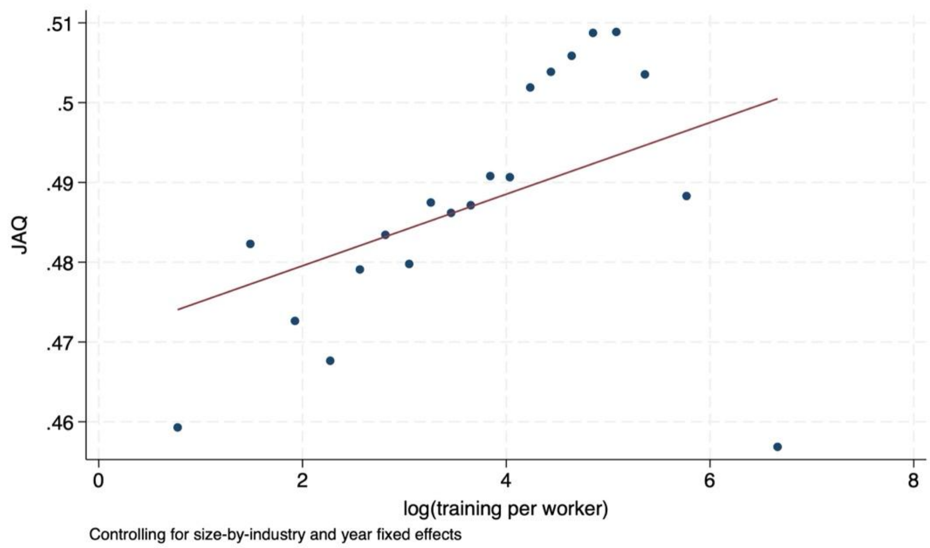

Figure 2.18. Firm-level fraction of well-matched employees (*JAQ*) and logarithm of training expenses per worker

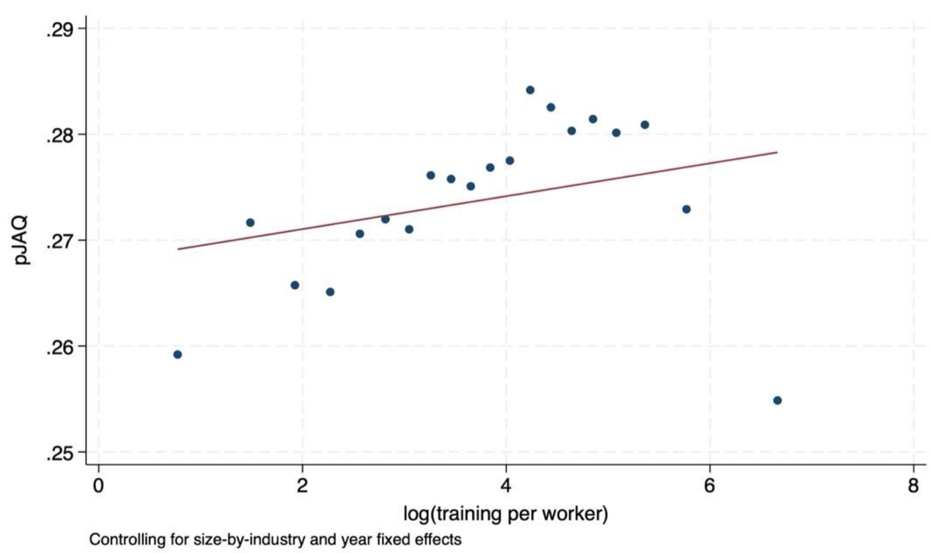

Figure 2.19. Firm-level average match quality (p*JAQ*) and logarithm of training expenses per worker

The next figures examine another aspect not covered for the other three countries: the relationship between match quality and firms' balance sheets. Figures 2.21 and 2.22 show that the relationship with the leverage ratio is positive at low levels, peaks around 10% leverage, then decreases slightly up to 50% leverage, and drops sharply for higher values. This indicates that while higher match quality increases firms' debt capacity, excessive leverage may limit their ability to match workers to jobs.

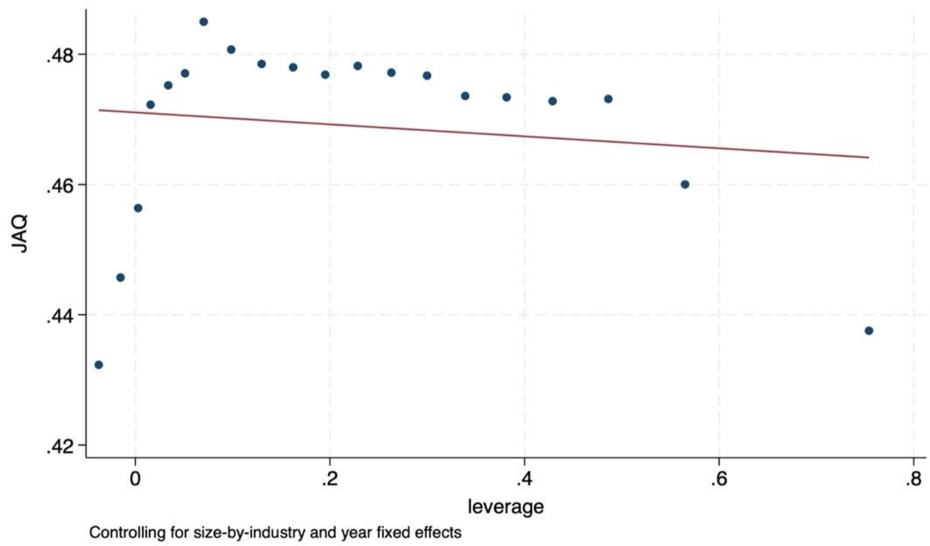

Figure 2.20. Firm-level fraction of well-matched employees (*JAQ*) and firm leverage

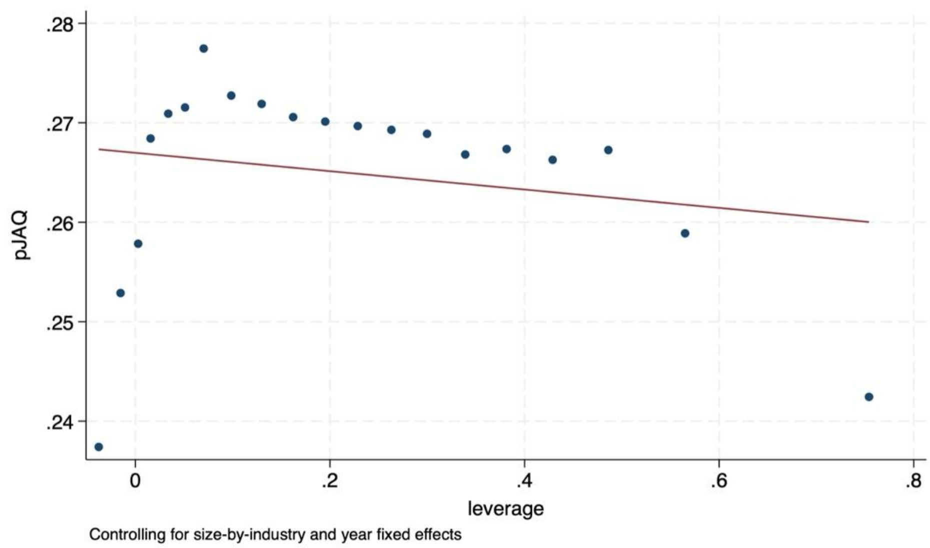

Figure 2.21. Firm-level average match quality (p*JAQ*) and firm leverage

Figures 2.23 and 2.24 instead show that the relationship with firms' liquidity, as measured by their cash-to-asset ratio, is consistently negative. This suggests that higher match quality reduces firms' need to rely on cash for operations, possibly reflecting their superior management techniques.

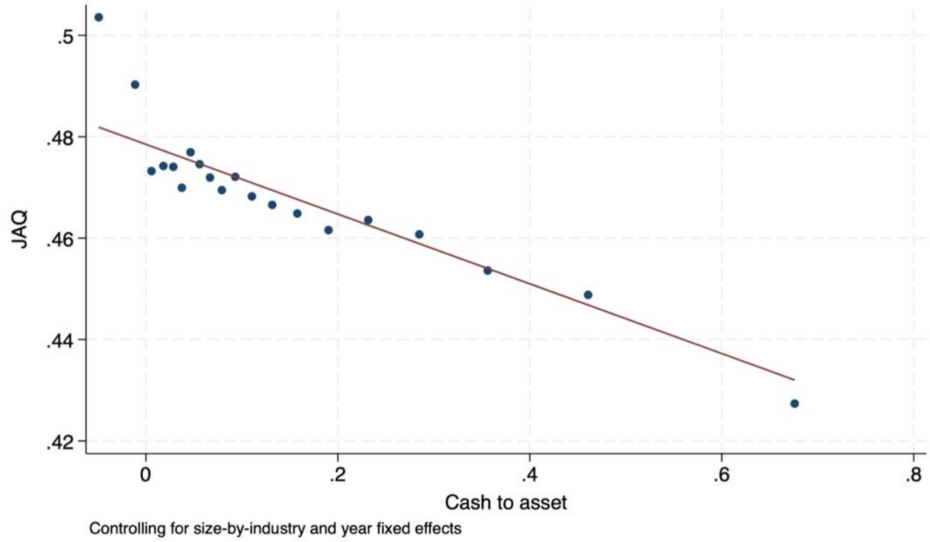

Figure 2.22. Firm-level fraction of well-matched employees (*JAQ*) and firm liquidity

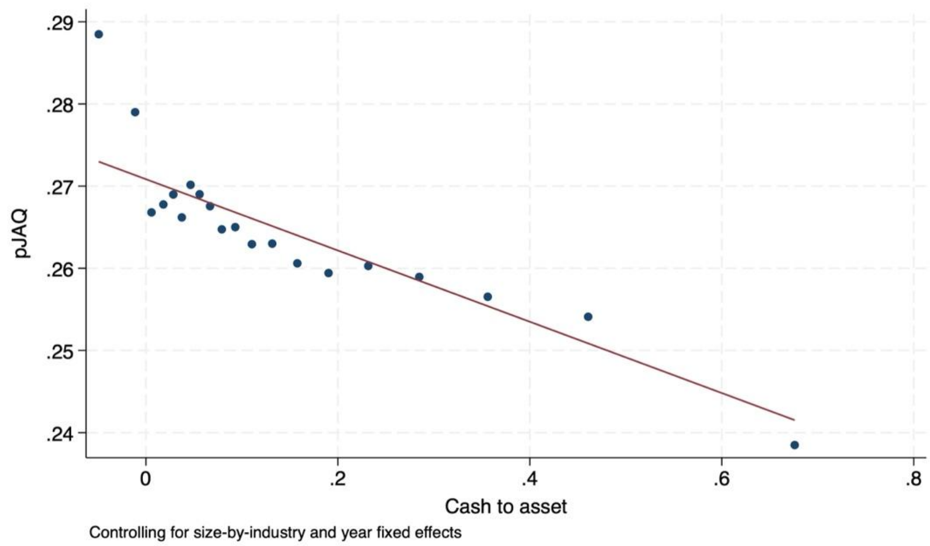

Figure 2.23. Firm-level average match quality (p*JAQ*) and firm liquidity

3.2.6 Match quality and firm-level outcomes

Figures 2.25 and 2.26 illustrate the relationships between match quality and firm-level productivity, as measured by the logarithm of value-added per employee, using partial regression plots of this variable against *JAQ* and *pJAQ*, conditioning on year effects and size-by-industry effects. As in the Swedish data, productivity correlates positively with match quality across firms.

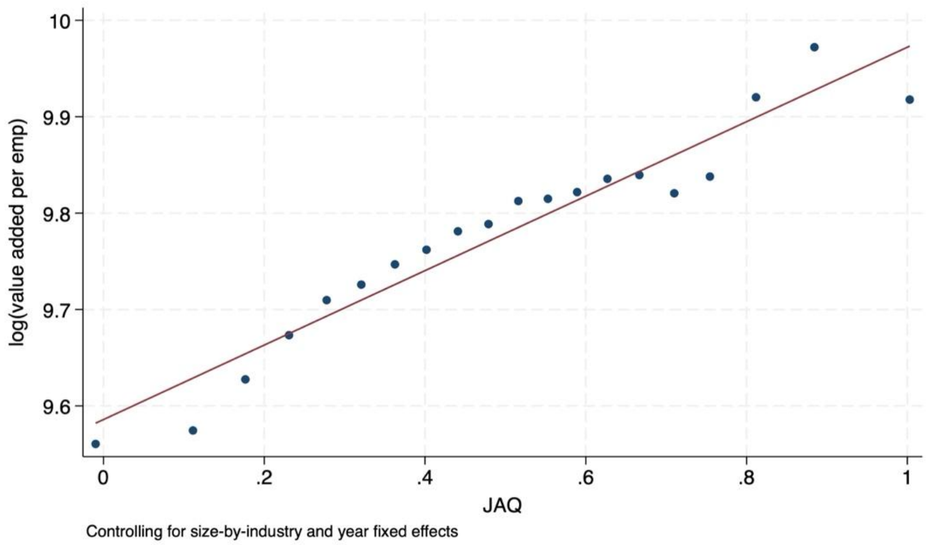

Figure 2.24. Firm-level fraction of well-matched employees (*JAQ*) and firm productivity

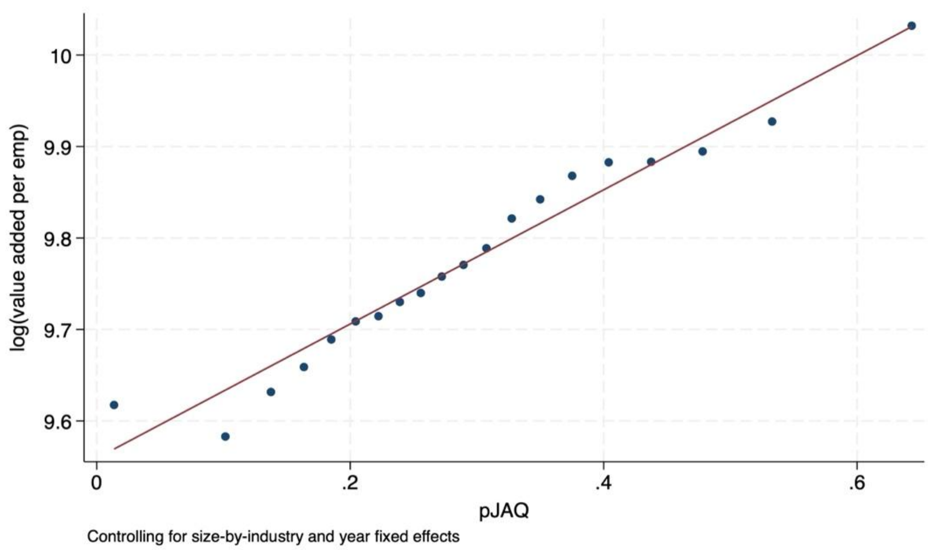

Figure 2.25. Firm-level average match quality (p*JAQ*) and firm productivity

Figures 2.27 and 2.28 show that firm-level match quality also correlates positively with firm profitability, as measured by the return on assets (ROA). This finding differs sharply from that established for Sweden, where match quality and profitability are negatively correlated in Figures 1.19 and 1.20, and either uncorrelated or negatively correlated in the estimates of Table 1.2.

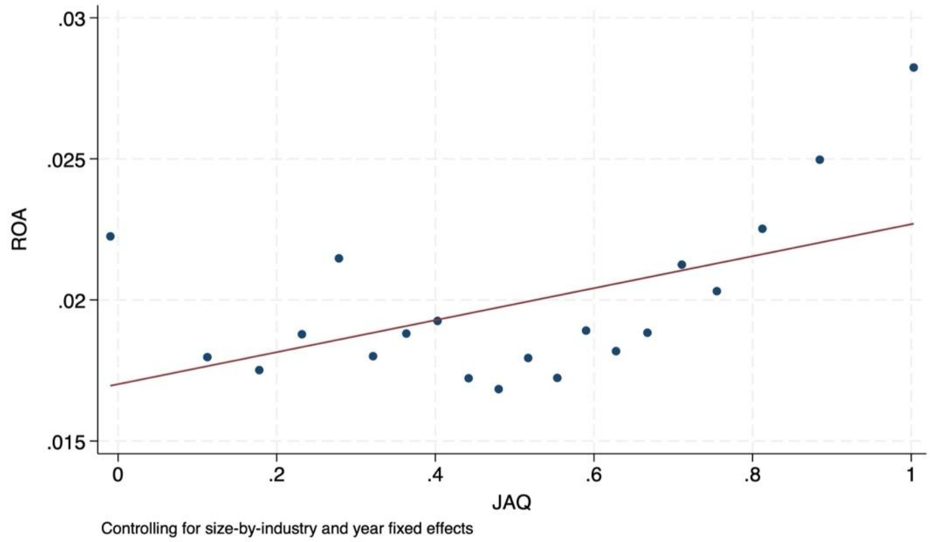

Figure 2.26. Firm-level fraction of well-matched employees (*JAQ*) and firm profitability (ROA)

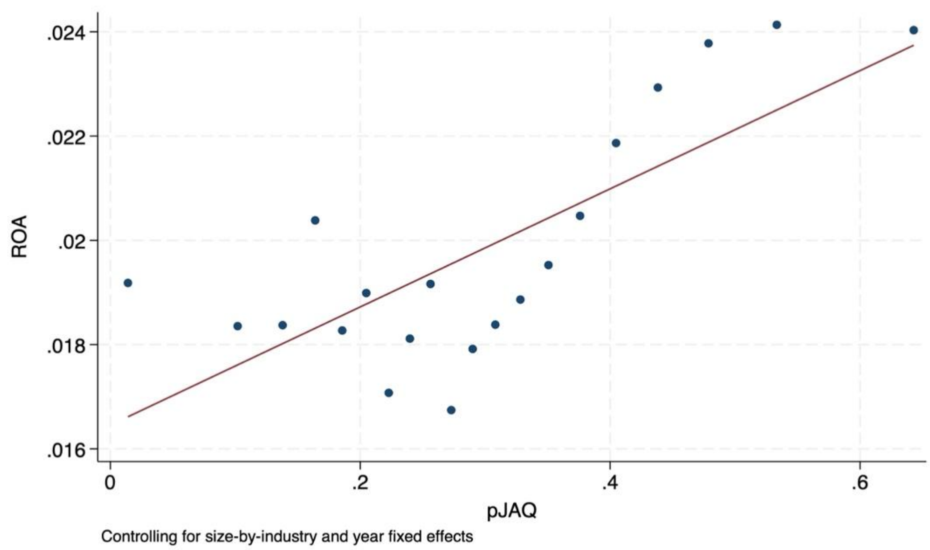

Figure 2.27. Firm-level average match quality (p*JAQ*) and firm profitability (ROA)

3.2.7 Time series patterns in match quality

Figures 2.29 and 2.30 show the time series of aggregate match quality in Portugal from 2004 to 2022, obtained by averaging the *JAQ* and *pJAQ* measures across firms for each year of our sample. Both figures show that the aggregate match quality is significantly larger in the learning sample than in the main sample, as expected: in the former, it is about one-third larger than in the latter.

The other striking finding is the strongly pro-cyclical behavior of both measures of aggregate match quality. This is more evident than in Swedish data, as Portugal's macroeconomic performance between 2004 and 2022 saw significant volatility: strong growth before 2008, a deep recession in 2009-13, followed by a remarkable recovery in 2017-19, the sharp COVID recession in 2020, and the strong 2021-22 rebound. In particular, the 2009-13 recession featured a “double dip”, with the GDP growth rate being -3.12% in 2009 due to the global financial crisis and -2.88 per year in 2011-12 due to the bank-sovereign debt crisis, unemployment peaking to 17.7% in 2013 and a sovereign bailout. This violent and protracted recession coincides with a significant and persistent decline in both of our aggregate measures of occupational match quality, which stabilize only as the 2017-19 recovery sets in, with GDP growth rates of about 3% per year and the country reattaining pre-crisis GDP levels.

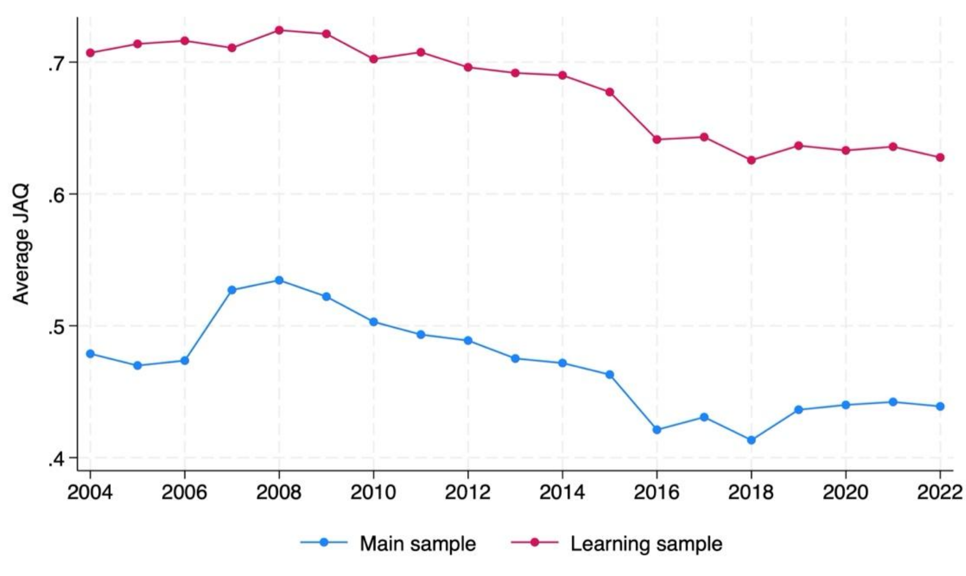

Figure 2.28. Aggregate fraction of well-matched employees (*JAQ*), in the main and in the learning samples, 2004-2022

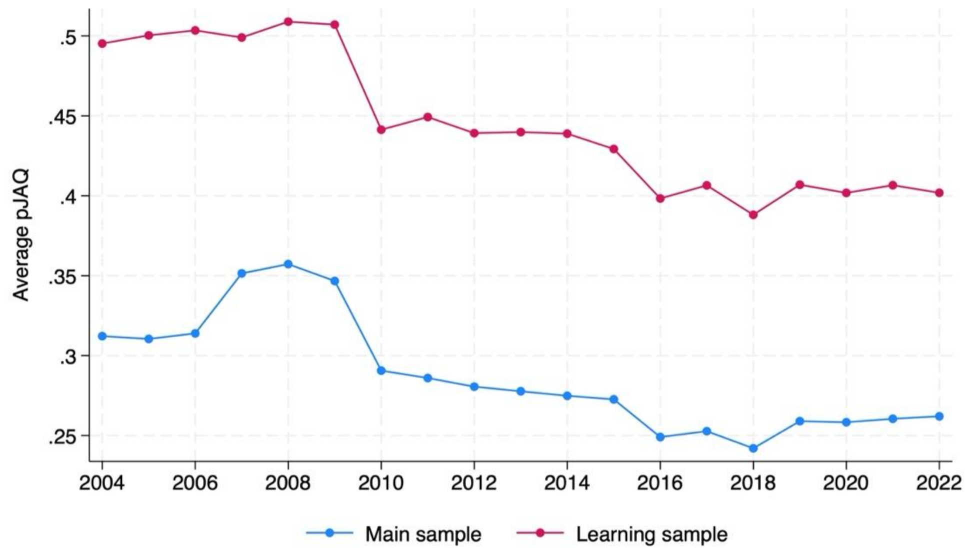

Figure 2.29. Aggregate average match quality (p*JAQ*), in the main and in the learning samples, 2004-2022

3.3 Italy

This section describes the data and presents the results regarding our worker-level and firm-level measures of match quality obtained from the administrative worker-firm matched data for Italy.

3.3.1 Data sources

For Italy, our analysis is based on administrative employee–employer-matched data provided by INPS, the Italian National Social Security Institute, which is responsible for collecting and administering compulsory social security contributions for employment relationships in Italy. As a result, its archives represent the most comprehensive source of information on private-sector employment in the country. The INPS employee–employer data have been available for a long time: data for private-sector employees have been available since the 1970s, with continuous coverage through the most recent available years and yearly updates.

The unit of observation is the individual employment relationship. Each worker is assigned a unique anonymized identifier that allows tracking across employers and over time, while a unique employer code identifies each firm. This structure enables the reconstruction of complete employment histories, including job starts, separations, and transitions between firms. On average, the dataset covers approximately 15 to 18 million private-sector employees per year, corresponding to the near-universe of dependent workers subject to INPS contributions. The number of active employers observed annually typically ranges between 1.5 and 2 million firms, reflecting the highly fragmented structure of the Italian productive system.

At the employee level, the data include detailed demographic and contractual information. Available variables typically comprise gender, year of birth, occupational category, contract type, employment status, and the number of weeks worked within each year. Earnings are reported as gross labor income subject to social security contributions and are recorded with high accuracy due to their direct link to contribution payments. These features make the data particularly suitable for analyses of wage distributions, employment stability, and career trajectories.

Each employment relationship is linked to an employer, for which the data report industry classification according to the ATECO system, geographic location, and number of employees. While the data do not include balance-sheet variables, ownership structures, or measures of governance, the employer identifiers allow aggregation of worker-level records to construct firm-level panels.

The longitudinal and administrative nature of the INPS data represents a key strength for our analysis. Because the data are collected for legal and fiscal purposes, they are not affected by sampling error, non-response, or recall bias, and variable definitions are stable over time. This makes the data particularly well-suited for estimating our measures of job match quality over a long period.

At the same time, the data have some limitations. Information on workers' education, skills, and job tasks is limited, and there are no direct measures of firm productivity, financial performance, or governance arrangements. Access to the data is subject to strict confidentiality constraints and is typically granted through secure research environments or specific institutional agreements.

3.3.2 Descriptive statistics

Table 3.1 presents summary statistics on labor income, education, employment contract type, occupation, age, and gender of private employees in the INPS sample as of April 2015. Figure 3.1 shows common support between the main sample and the learning sample. Unlike Sweden, to investigate common support, we use worker characteristics to predict firm-level value added rather than wages, since wage data are not yet included in our dataset (they will be in the future).

Table 3.1 - Employee characteristics: summary statistics, as of April 2015

Variable	Mean	P50	P10	P25	P75	P90	SD
Labor Income	2,518.52	2,082.58	1,196.59	1,642.82	2,757.94	3,903.07	2,566.00
College Degree	0.11	0.00	0.00	0.00	0.00	1.00	0.09
Permanent Contract	0.87	1.00	0.00	1.00	1.00	1.00	0.11
White Collar	0.34	0.00	0.00	0.00	1.00	1.00	0.22
Age	42.36	42.50	27.66	34.16	50.5	56.58	10.54
Female	0.41	0.00	0.00	0.00	1.00	1.00	0.24
Number of Observations				6,776,798			

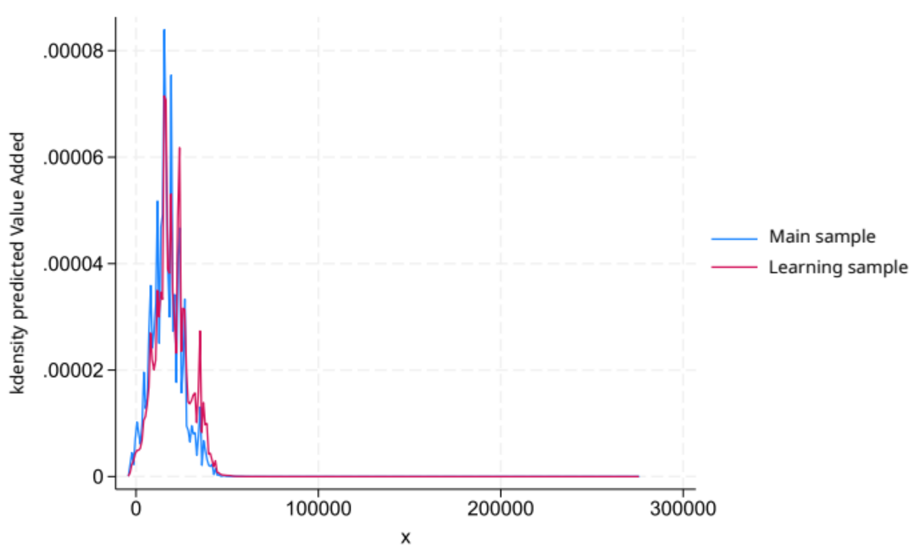

Figure 3.1 - Balancing of characteristics between the main sample and the learning sample

3.3.3 Match quality and worker-level characteristics

The analysis of the relationships between workers' match quality and their characteristics is to be added.

3.3.4 Match quality and worker-level wages

The analysis of the relationship between workers' match quality measures and individual workers' wages will be added as soon as wage data are included in our dataset. However, Table 3.2 presents regressions analyzing the relationship between separations and worker match quality in the Italian data, an issue for which we have no evidence so far in the other three countries. The regressions in this table are estimated using a random sample of 5% of all the workers. The dependent variable is a dummy equal to 1 when a worker separates from his/her previous employer, and 0 otherwise, and the explanatory variables are lagged values of individual match quality measures, i.e., *eJAQ* in Panel A and *epJAQ* in Panel B.

Table 3.2 - Match quality and worker-firm separations

	Outcome variable: separation dummy variable	
	(1)	(2)
Panel A		
Lagged eJAQ	-0.0354***	-0.0294***
	(0.0022)	(0.0022)
Panel B		
Lagged epJAQ	-0.0650***	-0.0483***
	(0.0025)	(0.0025)
N	386,296	378,849
Occupation FEs	no	no
Year FEs	yes	yes
Worker X	yes	yes
Firm X	no	yes
Worker FEs	no	no
Firm Fes	no	no

The estimates shown in column 1 refer to a specification that includes year fixed effects and worker-level control variables, and those in column 2 to a specification that also contains firm-level control variables. In both specifications, better match quality is associated with a significant reduction in workers' separations.

3.3.5 Match quality and firm-level characteristics

In this section, we describe and discuss the relationships between average firm-level match quality, as measured by *JAQ* and *pJAQ*, and firm-level characteristics. As or the other countries, we start with graphical evidence in the form of bin scatter plots, namely, quantile plots of firm characteristics on the vertical axis and measures of match quality on the horizontal axis, always controlling for year fixed effects and size-class-by-industry fixed effects. For these graphs, we used only information on broad (ATECO 07) sectors, unlike for Sweden, where we relied on more granular (2-digit) data for industries in some graphs. We expect more granular information about sectors to become available to us for Italy in the future as well.

Figures 3.2 and 3.3 show the relationship between JAQ or pJAQ and firm age, respectively. In both graphs, the correlation is positive and significant, particularly in the first 10 years after the firm's creation, as observed in Sweden and Portugal. The causality may go in either direction: on the one hand, better match quality might increase firm resilience and thus lead firms to survive longer; on the other hand, older firms may employ workers with higher average tenure or attract better workers, who are easier to allocate.

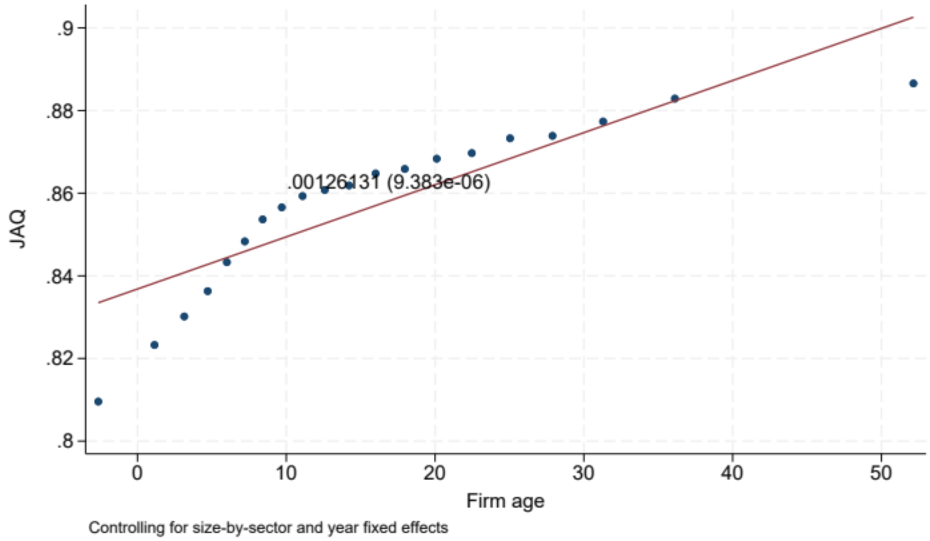

Figure 3.2 - Firm-level fraction of well-matched employees (JAQ) and firm age

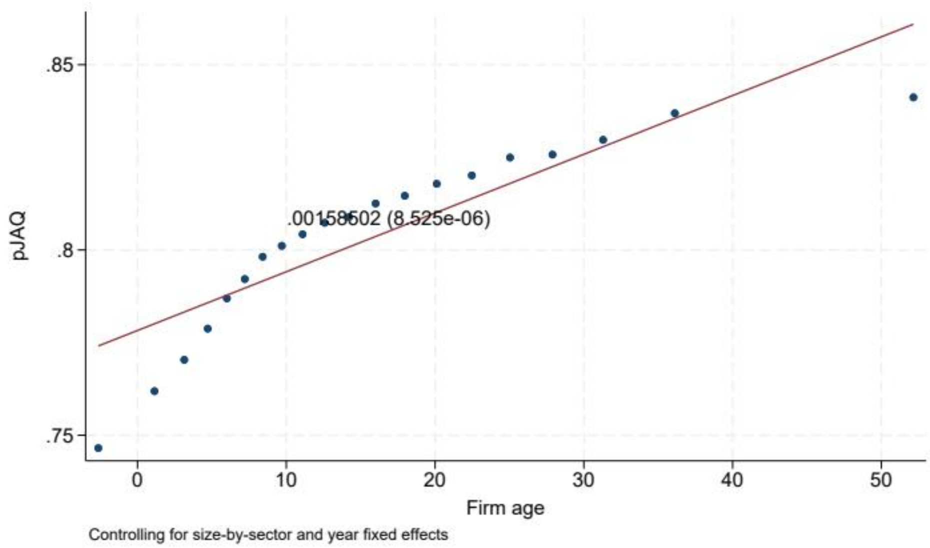

Figure 3.3 - Firm-level average match quality (p*JAQ*) and firm age

Figures 3.4 and 3.5 show that on the whole the correlation between firm-level measures of match quality and firm size is negative, as will be seen to be the case in the Netherlands as well. This may be because the allocation problem becomes more difficult as firm size grows, as larger firms have a wider array of different occupations to fill.

Qualitatively, the inverse U-shaped relationship is also similar to that observed in the Swedish and Portuguese data. However, in the Italian data, the inverse U-shape is more pronounced, with a peak at a lower firm size. Note also that for large firms, JAQ starts increasing again, possibly because these firms have more resources to cover the fixed costs of identifying better matches, such as more sophisticated management and personnel allocation systems.

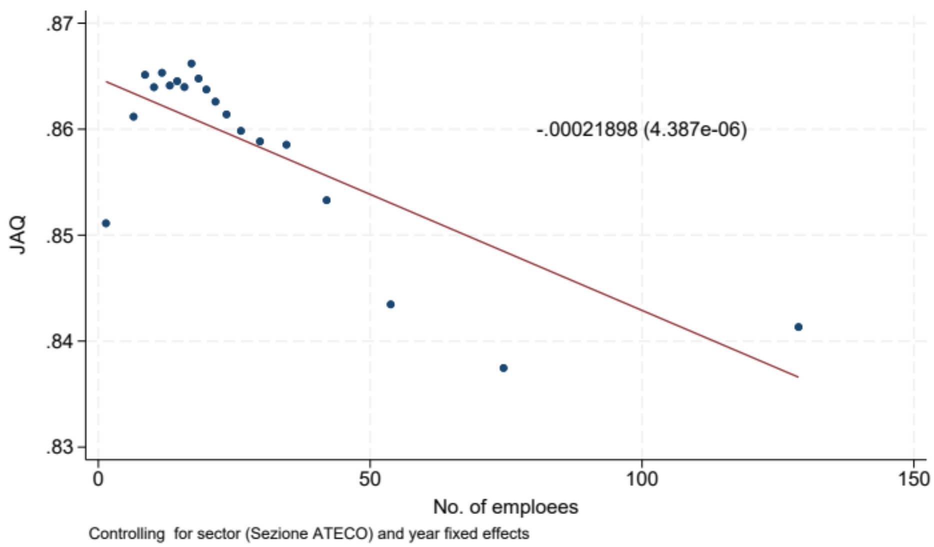

Figure 3.4 - Firm-level fraction of well-matched employees (*JAQ*) and number of employees

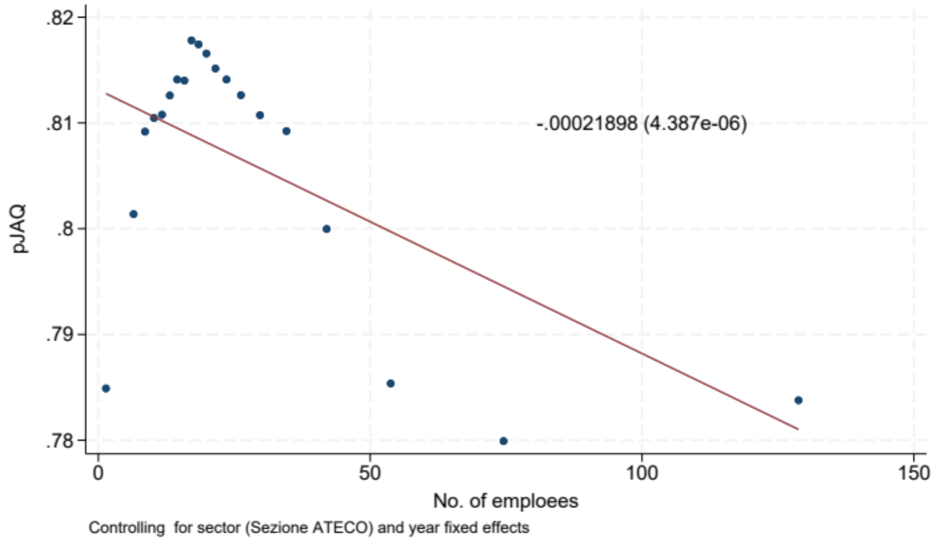

Figure 3.5 - Firm-level average match quality (p*JAQ*) and number of employees

Figures 3.6 and 3.7 investigate the correlation between firm-level match quality and employees' education, measured by the firm's share of employees with at least a college degree. In the INPS database, the education level is classified as follows: no education is coded as 00; elementary school as 10; junior high school as 20; professional diplomas and senior high school diplomas as 30 to 60; college degrees, master's degrees, and doctoral degrees as 70 to 90. The fraction of educated employees used in the following graphs equals the fraction of employees with an education level of at least 70. The relationship is quite interesting, as it sharply differs from that observed for the other three countries: initially, match quality correlates positively with the fraction of highly educated employees, but then it declines, resulting in an overall negative correlation. This may indicate an overall inability of Italian firms to put highly educated workers to their best use, as also suggested by much evidence regarding the outflow of Italian graduates to other countries.

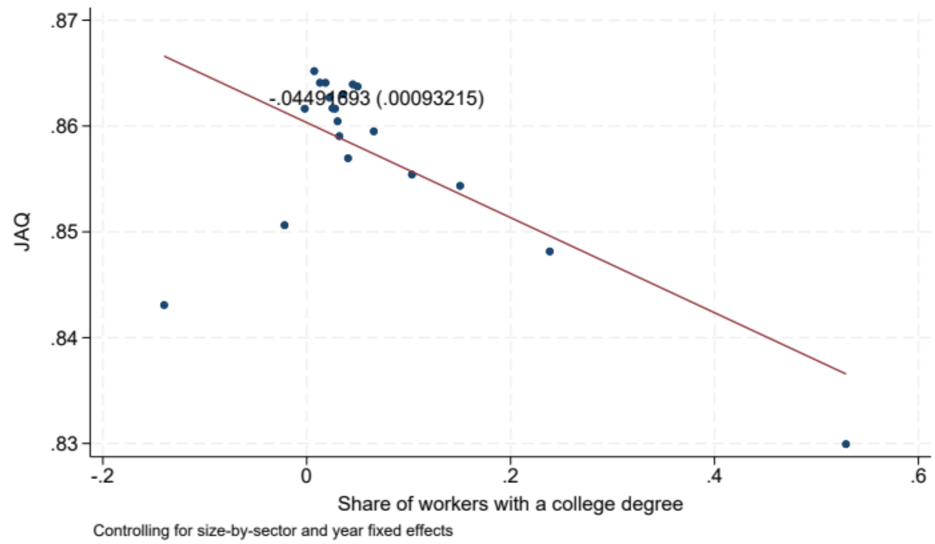

Figure 3.6. Firm-level fraction of well-matched employees (*JAQ*) and fraction of employees with at least a college degree

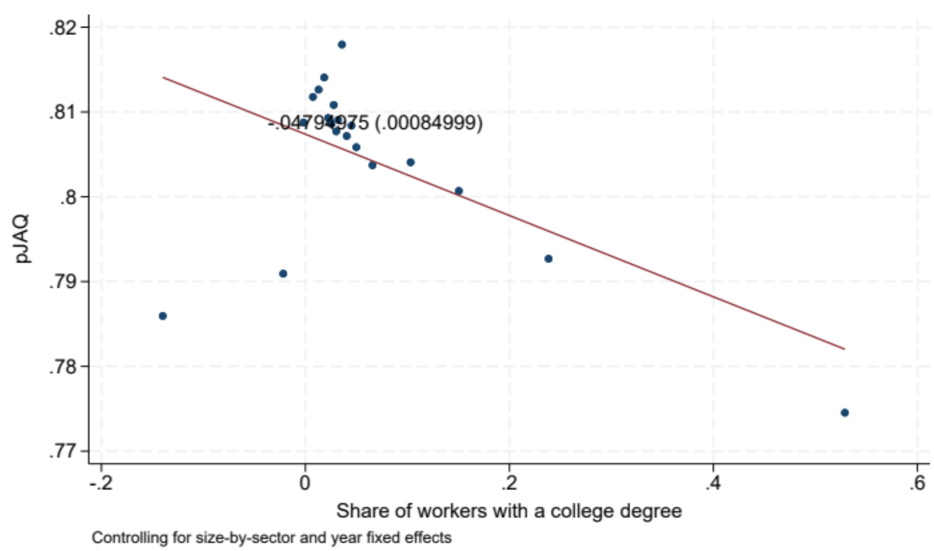

Figure 3.7 - Firm-level average match quality (p*JAQ*) and fraction of employees with at least a college degree

Figures 3.8 and 3.9 show a positive and significant correlation between firm-level match quality measures and average worker experience. This pattern is as expected, as more experienced workers are easier to allocate, and similar to that observed in Swedish and Portuguese data.

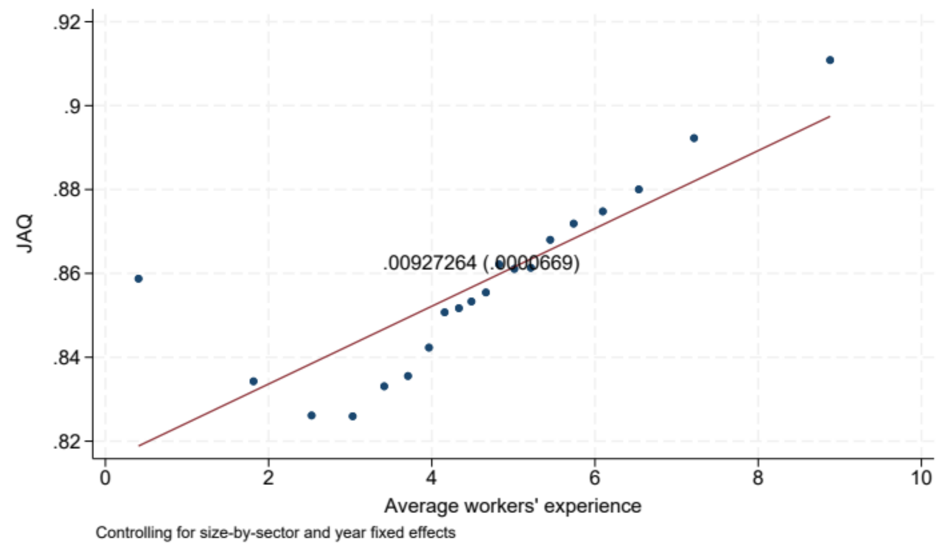

Figure 3.8. Firm-level fraction of well-matched employees (*JAQ*) and average employees' experience

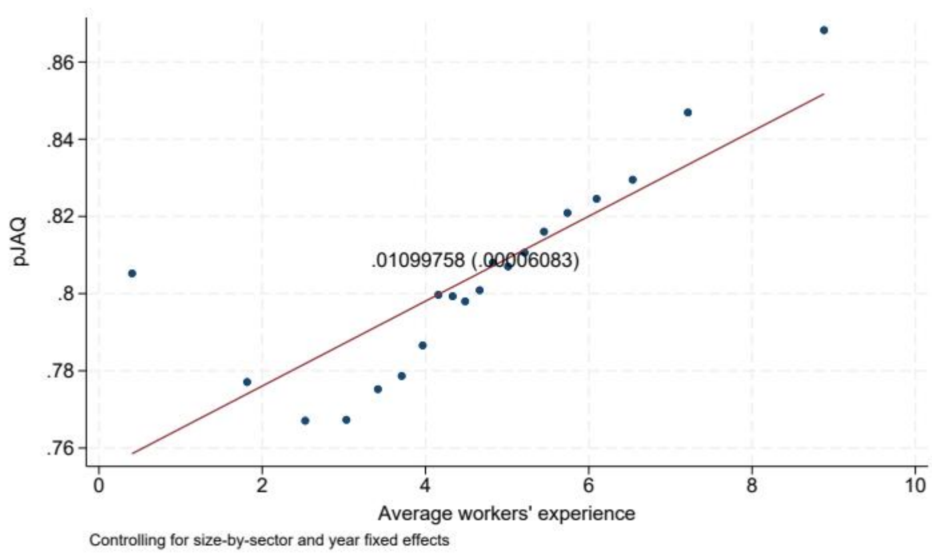

Figure 3.9 - Firm-level average match quality (p*JAQ*) and average employees' experience

Figures 3.10 and 3.11 show that firm-level match quality is also strongly and positively correlated with job tenure in the relevant, again consistently with expectations as well as with evidence for Sweden and Portugal.

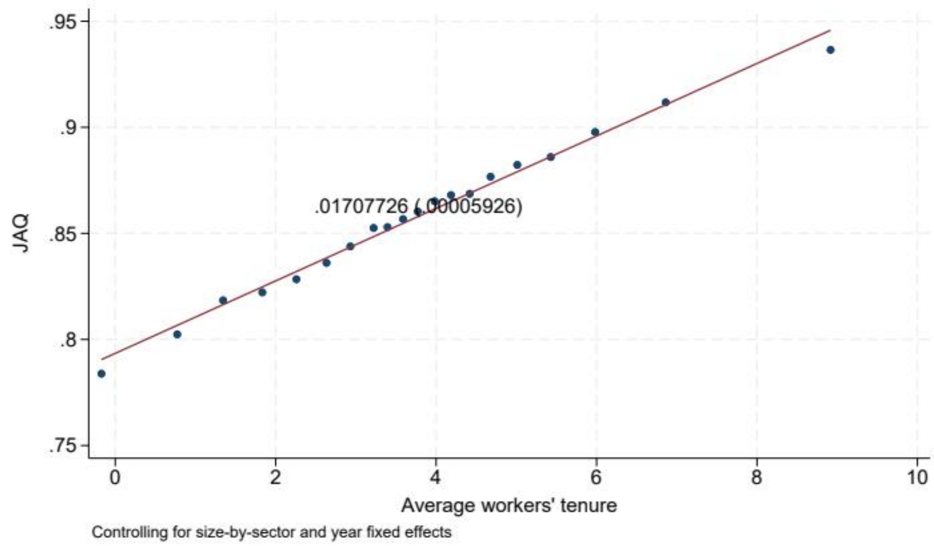

Figure 3.10. Firm-level fraction of well-matched employees (*JAQ*) and average tenure of employees

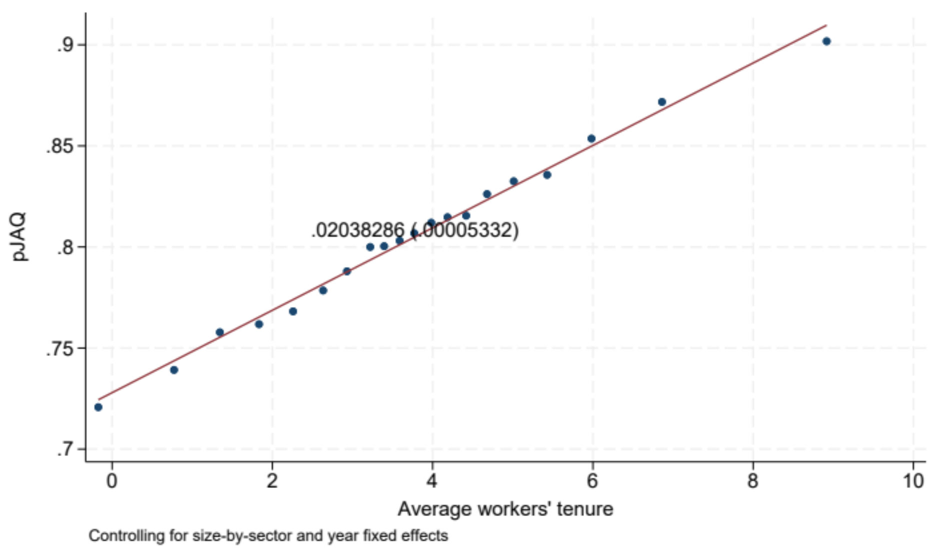

Figure 3.11 - Firm-level average match quality (p*JAQ*) and average tenure of employees

3.3.6 Match quality and firm-level outcomes

Figures 3.12 and 3.13 illustrate that, in the Italian data, firm-level match quality is positively correlated firm productivity, measured by value-added per employee, as in Sweden and Portugal.

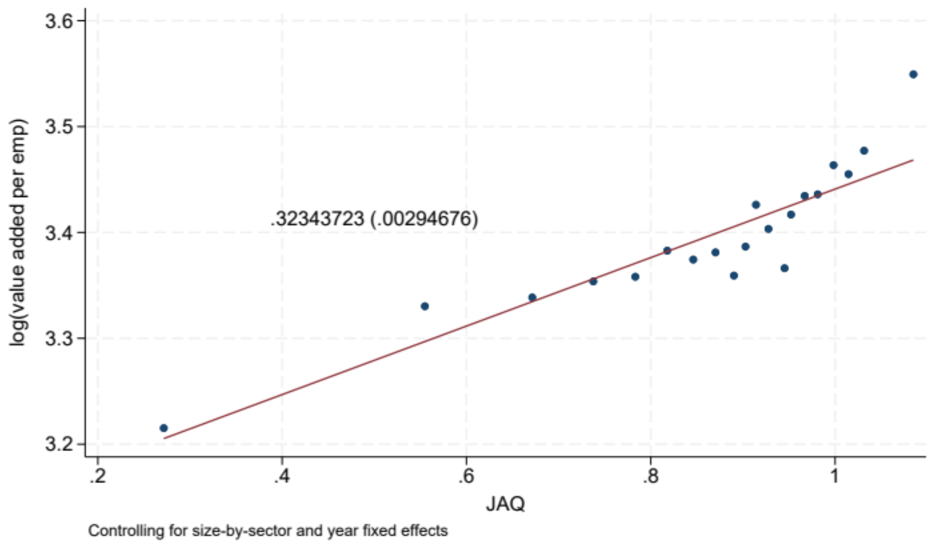

Figure 3.12 - Firm-level fraction of well-matched employees (*JAQ*) and firm productivity (logarithm of value added per employee)

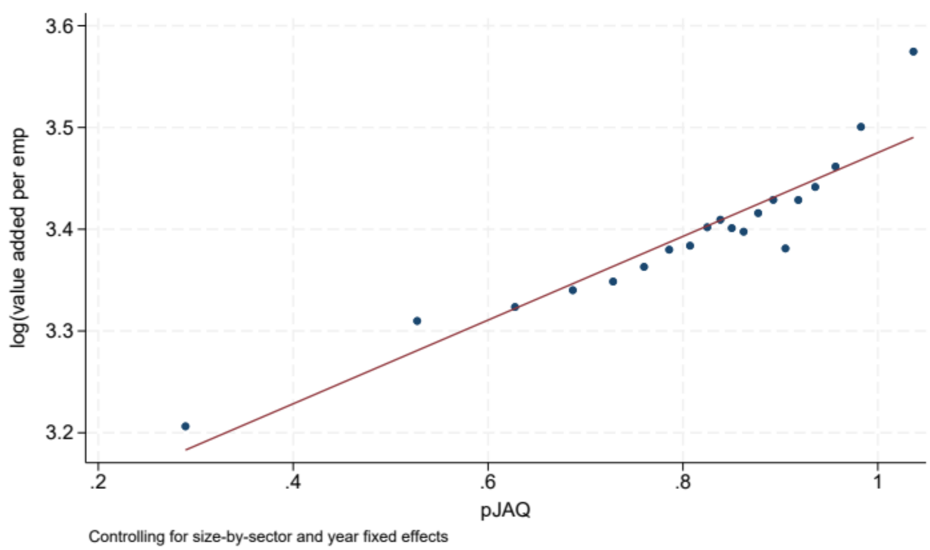

Figure 3.13. Firm-level average match quality (p*JAQ*) and firm productivity (logarithm of value added per employee)

A positive and significant relationship exists also between JAQ and firm profitability (ROA).

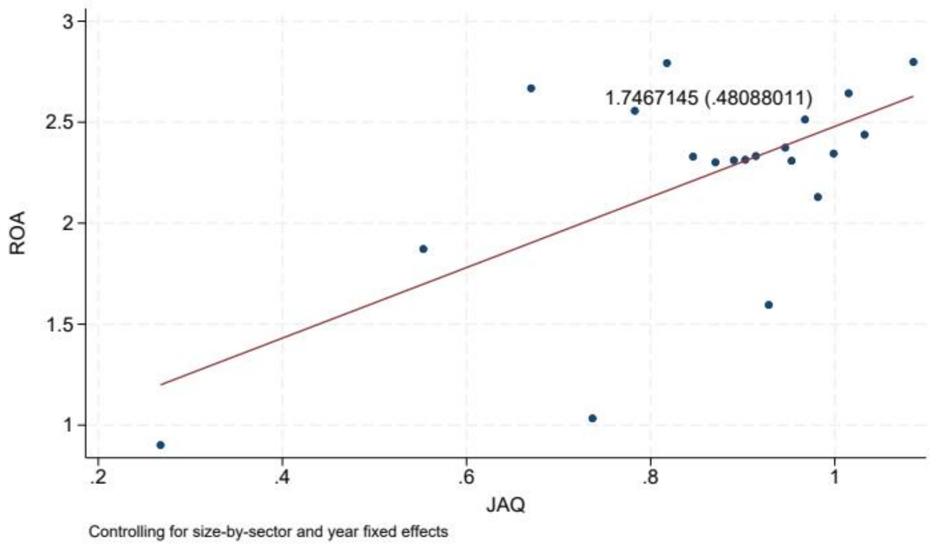

Figure 3.14. Firm-level fraction of well-matched employees (*JAQ*) and firm profitability (ROA)

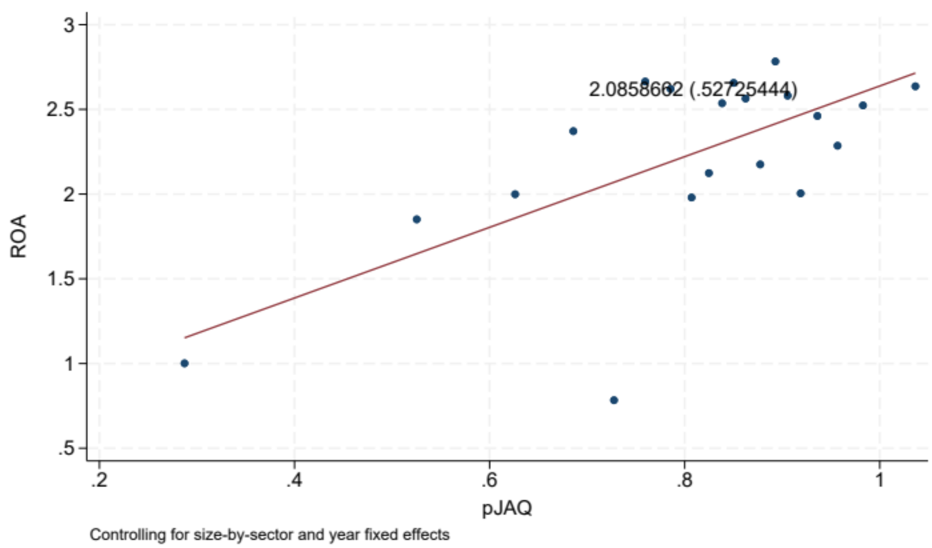

Figure 3.15 - Firm-level average match quality (p*JAQ*) and firm profitability (ROA)

Table 3.3 explores the relationships between match quality and both firm productivity and for profitability via panel regression, estimated on a 10% random sample of the firms present in the INPS database, controlling for other determinants of productivity and profitability. All specifications presented in the table include year and sector dummies. Hence, the relationship between productivity and JAQ captured by our estimates is not driven by differences in the availability of workers or labor market conditions over time or across sectors.

In Panel A of Table 3.3, column 1 reports the OLS estimates from a regression of log sales per employee on JAQ, where the coefficient on JAQ is significant and equal to 0.321, indicating that a 10-percentage-point increase in JAQ is associated with a 3.21% increase in sales per employee. In column 2 of Table 3.3, the dependent variable is the log of value added per employee. The coefficient of *JAQ* is again positive and significant: a 10-percentage-point increase in JAQ is associated with a 2.83% increase in value added per employee, identical to the estimate for Sweden using the same specification. This qualitative result is robust to the inclusion of industry indicators, log number of employees, log capital, and the fraction of employees with at least a college degree among the regressors, as shown by columns 4 and 5 of the table. The estimated coefficients of JAQ in columns 4 and 5 decrease in magnitude but remain positive and significantly different from zero.

In panels B, C, and D of the table, we control for various possible sources of omitted variable bias, namely, firm characteristics, differences in firms' occupation structures, and differences in workers' quality across firms. In the specifications of Panel B we control for firms' occupation structure and for firm controls. The specifications of Panel C also control for workers' characteristics used in the machine learning algorithm, averaged across all workers employed in firm *f* in year *t*. Those shown in Panel D also include firm and year fixed effects, as well as interactions between an indicator for firm size, industry, and year effects, so that, effectively, the coefficients are estimated within each of the industry-size bins used to calculate the machine-learning algorithm.

The results shown in Panels B, C and D are qualitatively similar to those in Panel A: the estimated coefficients of *JAQ* drop in magnitude but remain positive and statistically significant in columns 1, 2, 4, and 5. The result that productivity correlates positively and significantly with JAQ is robust even to the inclusion of firm and year fixed effects, as shown in Panel D.

In almost all specifications shown in Table 3.3, profitability, as measured by operating return on assets (OROA), is either not significantly related or negatively correlated with JAQ, as shown in columns 3 and 6, consistent with the estimates for Sweden.

Table 3.3 - Firm productivity, profitability, and fraction of well-matched employees (*JAQ*)

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A	log(sales/emp)	log(va/emp)	OROA	log(sales/emp)	log(va/emp)	OROA
JAQ	0.321***	0.283***	-0.661	0.098***	0.140***	-1.149*
	(0.028)	(0.020)	(0.585)	(0.016)	(0.014)	(0.765)
log(K/L)				0.633***	0.417***	1.969***
				(0.006)	(0.005)	(0.595)
log(L)				0.032	-0.013***	1.600***
				(0.006)	(0.005)	(0.330)
Sh. w/ college				0.127***	0.282***	-2.230*
				(0.030)	(0.021)	(1.361)
Sector	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
PANEL B						
JAQ	0.195***	0.191***	-1.187*	0.055***	0.080***	-1.496**
	(0.026)	(0.019)	(0.662)	(0.017)	(0.014)	(0.760)
Industry	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Occ.	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.3 – continued

PANEL C						
JAQ	0.115***	0.110***	-0.787	0.047***	0.058***	-1.021
	(0.026)	(0.018)	(0.786)	(0.017)	(0.014)	(0.822)
Group-by-yr FEs	Yes	Yes	Yes	Yes	Yes	Yes
Occupations	Yes	Yes	Yes	Yes	Yes	Yes
Workers X	Yes	Yes	Yes	Yes	Yes	Yes
Firm X	No	No	No	Yes	Yes	Yes
PANEL D						
JAQ				0.034**	0.038***	-1.326**
				(0.015)	(0.013)	(0.587)
Group-by-yr	No	No	No	Yes	Yes	Yes
Firm and yr FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	214712	209528	215499	214709	209525	215494
No. Firms	21598	21522	21606	21597	21521	21606
y Mean	4.711	3.528	4.100	4.712	3.528	4.102
y St. Dev.	1.145	0.818	49.575	1.145	0.818	49.570

3.3.7 Time series patterns in match quality

Figures 3.16 and 3.17 show the time series of aggregate match quality in Italy from 2005 to 2018, obtained by averaging *JAQ* and *pJAQ* across firms for each year of the sample. In both figures, the aggregate match quality is significantly larger in the learning sample than in the main sample, as expected. In fact, the gap between them widens over time, average match quality remaining stable in the learning sample, and declining steeply since 2008 in the main sample.

Like Portugal, Italy was severely affected by the 2010-11 sovereign debt crisis and the subsequent stagnation (2012-14). Again, like Portugal, it experienced a double-dip recession, with the economy recording a -5.31% growth rate in 2009 and unemployment reaching an all-time high of over 13% in late 2014. Unlike Portugal, however, in the following years Italy did not stage a robust recovery: growth resumed in 2015 but remained sluggish, peaking at 1.6% in 2017 before slowing to 0.83% in 2018. From 2004 to 2018, Italy experienced stagnant or near-zero productivity growth. The average annual growth rate for labor productivity between 1995 and 2017 was only 0.4%, compared to a 1.2% EU average for the pre-crisis period.

The severe shock due to the sovereign debt crisis and its aftermath, coupled with the dismal long-term performance of the economy in terms of growth and productivity, may be at the root of the increasing mismatch problem highlighted by the downward trends of *JAQ* and *eJAQ* in the Italian economy, which stand in sharp contrast with those in the other three countries analyzed in this paper. In turn, this increasing mismatch may have contributed to the rise in unemployment, as found by Şahin et al. (2014). It is also tempting to relate this time-series pattern to the finding in Figures 3.6 and 3.7 that firm-level match quality is negatively correlated with employees' education, suggesting that Italian firms are not able to put their most educated employees to their best use.

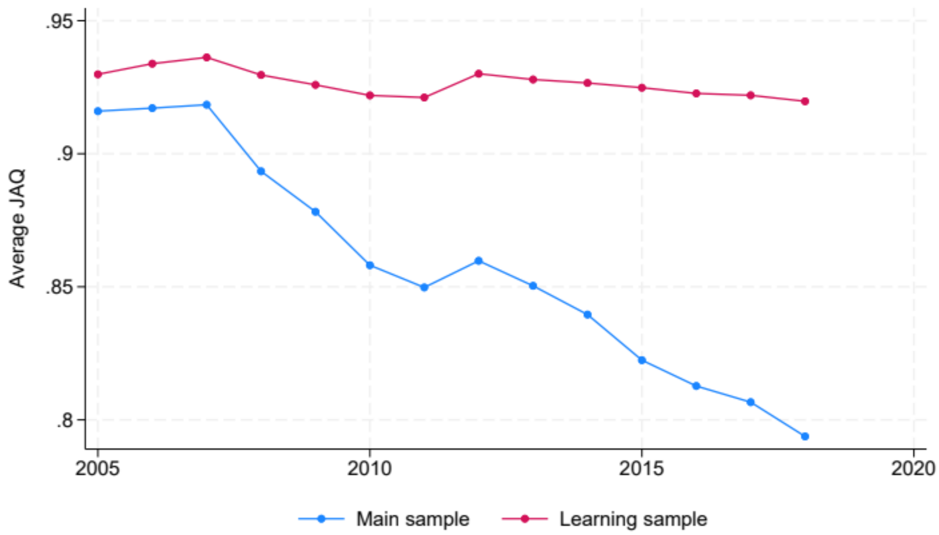

Figure 3.16. Aggregate fraction of well-matched employees (*JAQ*), in the main and in the learning samples, 2005-2012

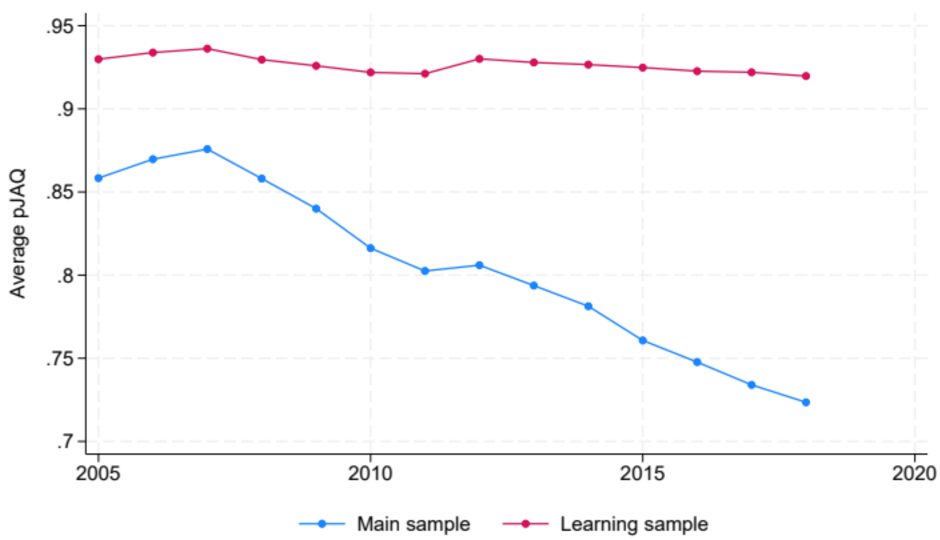

Figure 3.17. Aggregate average match quality (p*JAQ*), in the main and in the learning samples, 2005-2012

3.4 The Netherlands

This section describes the data and presents results on our worker-level and firm-level measures of match quality, obtained from administrative worker-firm matched data for the Netherlands.

3.4.1 Data sources

Our primary data source is the matched employer-employee administrative dataset covering the entire Dutch workforce, maintained by the National Statistical Office (*Centraal Bureau voor de Statistiek, or CBS*). Our data cover the years 2010-2022.

The data include information on workers' wages, earnings, hours worked, and demographic information, such as gender and age. We can also access detailed information on individuals' education for almost 90% of our sample, which we use to identify college graduates. Our analysis focuses on workers in the private sector. For these workers, we can access employers' balance sheet data, which we mainly use to compute measures of profitability, productivity, and firm size, as well as their industry, using the 2008 Dutch Standard Business Classification (*Standaard Bedrijfsindeling*, SBI 2008), which corresponds to the European Union 2-digit NACE, revision 2.

The occupation is not available for all workers but only for a subsample corresponding to the intersection of the CBS data with the CBS' National Survey on Working Conditions (*Nationale Enquête Arbeidsomstandigheden*, henceforth NEA) and the Labor Force Survey (*Enquête beroepsbevolking*, henceforth EBB). The final sample includes about 25 thousand workers per year.

The Dutch data have two significant limitations. First, worker-level data are repeated cross-sections, rather than panel data, unlike those for the other three countries, which of course prevents us from exploiting within-worker variation in the regression analysis of the relationship between wages and individual match quality. Second, we only have access to a random sample of the population. Thus, we know the occupation only for a small fraction of each firm's workforce. Hence, we risk that our inferences regarding mismatches and firm-level outcomes may be based on a small fraction of the firm's workforce. To address this concern, we sort firms by the fraction of their workers included in either survey and select the 2,000 firm-years with the highest fraction. More conservative or restrictive cutoffs lead to similar inferences.

3.4.2 Descriptive statistics

Table 4.1 presents descriptive statistics for our sample. Probably because of the much smaller sample size, both *eJAQ* and *pJAQ* are significantly lower than those estimated in the other countries, and exhibit more limited variation. The average worker earns about €43,000 and is 43 years old. Half of the workers have a college degree, and slightly less than half (48%) are women. We report statistics on experience and tenure even though, as discussed above, they are severely censored.

Table 4.1 - Job mismatch and worker characteristics

Variable	Mean	P50	P10	P25	P75	P90	SD
epJAQ	0.08	0.06	0.03	0.05	0.09	0.13	0.04
eJAQ	0.11	0.09	0.00	0.04	0.16	0.23	0.10
Labor Income	43,043	39,069	9,280	23,374	56,022	76,258	32,030
Log Labor Income	10.35	10.57	9.14	10.06	10.93	11.24	0.98
College Degree	0.50	1	0	0	1	1	0.50
Age	43.30	44	24	32	55	61	13.35
Female	0.48	0	0	0	1	1	0.50
Experience	7.26	8	3	5	10	11	3.20
Tenure	5.11	3.00	0.00	1.00	8.00	13.00	5.24

Table 4.2 reports the summary statistics after aggregating our mismatch measures at the firm level. We report statistics on key outcomes of interest, including measures of productivity (value added per employee and sales per employee) and profitability (ROA, defined as net earnings scaled by assets). We also report descriptive statistics for control variables, such as capital intensity (the ratio of assets to employees) and firm size, measured as the total number of employees. The average firm has 1,310 employees, showing that, compared to the other countries, the Dutch sample is skewed towards larger firms. Also at the firm level, mismatch measures (*JAQ* and *pJAQ*) tend to be lower than those

estimated in other countries, such as in Sweden. One possible explanation is the larger average firm size of Dutch firms, which may suggest the need for a finer grid of size bins to estimate the benchmark top-productivity firms in an alternative implementation of the machine-learning algorithm. For this deliverable, we have preferred to rely on a consistent version across countries rather than fine-tuning it to each dataset's characteristics.

Table 4.2 - Job mismatch and firm characteristics

Variable	Mean	P50	P10	P25	P75	P90	SD
pJAQ	0.07	0.06	0.03	0.04	0.09	0.13	0.05
JAQ	0.10	0.08	0.00	0.00	0.15	0.25	0.11
Log(Sales/Empl.)	3.61	4.46	-0.07	1.87	5.48	6.23	2.49
Log(VA/Empl.)	2.64	3.54	-0.75	0.88	4.42	4.83	2.28
ROA	0.05	0.04	-0.06	0.00	0.09	0.18	0.56
# Employees	1,310	899	502	656	1,568	2,778	1,055
Avg. College Degree	0.46	0.47	0.07	0.23	0.70	0.84	0.28
Avg. Age	42.98	44.20	33.80	39.73	47.55	50.09	6.64
Experience	7.26	8	3	5	10	11	3.20
Log(K/L)	3.45	3.94	-0.69	1.44	5.43	6.71	2.82

3.4.3 Match quality and worker characteristics

Since the Dutch data are repeated cross-sections rather than panel data, investigating the relationships between match quality and worker characteristics at the individual worker level would not add much to the evidence presented at the firm level, unlike the other three countries, where worker panel data are available. Hence, we defer the analysis of the relationships between match quality and worker characteristics to Section 3.4.5, where we shall investigate them at the firm level.

3.4.4 Match quality and worker-level wages

Table 4.3 presents regression results displaying the relationship between mismatch and wages, measured by (the logarithm of) workers' annual labor income. Panels A and B report results respectively obtained using *eJAQ* and *epJAQ* as the key regressor of interest. Standard errors are clustered at the firm level.

Table 4.3 - Match quality and workers' labor income

	Outcome variable: log(annual labor income)			
	(1)	(2)	(3)	(4)
Panel A				
eJAQ	0.386***	0.255***	0.328***	0.306***
	(0.145)	(0.110)	(0.110)	(0.080)
Panel B				
epJAQ	0.773***	0.563***	0.978***	0.955***
	(0.368)	(0.288)	(0.309)	(0.228)
N	46,282	46,282	46,179	46,177
Year Fes	No	No	No	No
Yr.×Ind. ×Size	No	Yes	Yes	Yes
Firm Controls	No	No	Yes	Yes
Worker Controls	No	No	Yes	Yes
Occupation FEs	No	No	No	Yes

Column 1 shows the estimates from simple bivariate regressions, that is, with no controls. In both panels, the coefficient of the measure of match quality is positive, large and precisely estimated, being significantly different from zero at the 1% level. Column 2 reports estimates from a specification that includes year-industry-size fixed effects: the results are largely unaffected.

While the cross-sectional nature of the Dutch data prevents the inclusion of firm or worker fixed effects, the specification in column 3 includes an extensive set of worker and firm controls, namely workers' age, tenure, a college dummy, the logarithm of firm size (measured by the number of employees), and the logarithm of the firm's asset-to-labor ratio. Again, the results are similar and, if anything, coefficient estimates rise slightly in magnitude.

Finally, the specification in column 4 includes occupation fixed effects. The estimated coefficients of *eJAQ* and *epJAQ* are 0.306 and 0.955, respectively. While these coefficients are larger than those estimated in other samples, such as the Swedish one, we also note that in the Dutch data these match quality metrics exhibit much lower variation: for example, their standard deviations are 0.10 and 0.04, respectively, about one-fifth compared to Sweden. Thus, the sample variation in wages explained by variation in worker-level match quality is of the same order of magnitude.

3.4.5 Match quality and firm-level characteristics

This section presents descriptive evidence on the relationship between firm-level match quality, as measured by *JAQ* and *pJAQ*, and firm-level characteristics. We present graphical evidence in the form of bin scatter plots, namely quantile plots of firm characteristics on the vertical axis and measures of match quality on the horizontal axis, always controlling for year fixed effects and size-class-by-industry fixed effects. We start with descriptive evidence on the relationships between *JAQ* (and *pJAQ*) and firm-level averages of worker characteristics, and subsequently turn to descriptive evidence regarding their relationship with firm size.

Figures 4.1, 4.2, 4.3 and 4.4 show that match quality is on the whole negatively correlated with both average age and average tenure, unlike what was found for the other three countries. However, upon closer inspection, both relationships may be better modeled by non-linear functions. Specifically, in Figures 4.1 and 4.2, firm-level average match quality is essentially invariant over the first 5 years of tenure and is a decreasing function of tenure only beyond that. Similarly, in Figures 4.3 and 4.4, average match quality is increasing in average employee age up to the 35-40 age bracket and becomes decreasing for larger values of average employee age. A qualitatively similar pattern is present in Swedish and Portuguese data as well, except that in those countries the relationship with employees' age becomes negative for ages above 50 rather than above the 35-40 bracket. In the case of tenure, the difference with Portugal and Sweden may be partly explained also by the fact that Dutch tenure data are severely censored because they start in 2010.

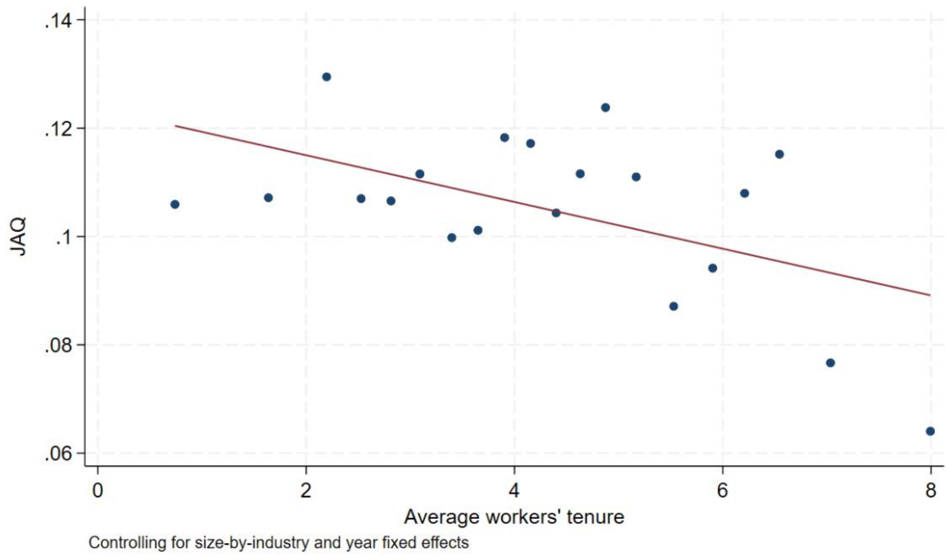

Figure 4.1. Firm-level fraction of well-matched employees (*JAQ*) and workers' tenure

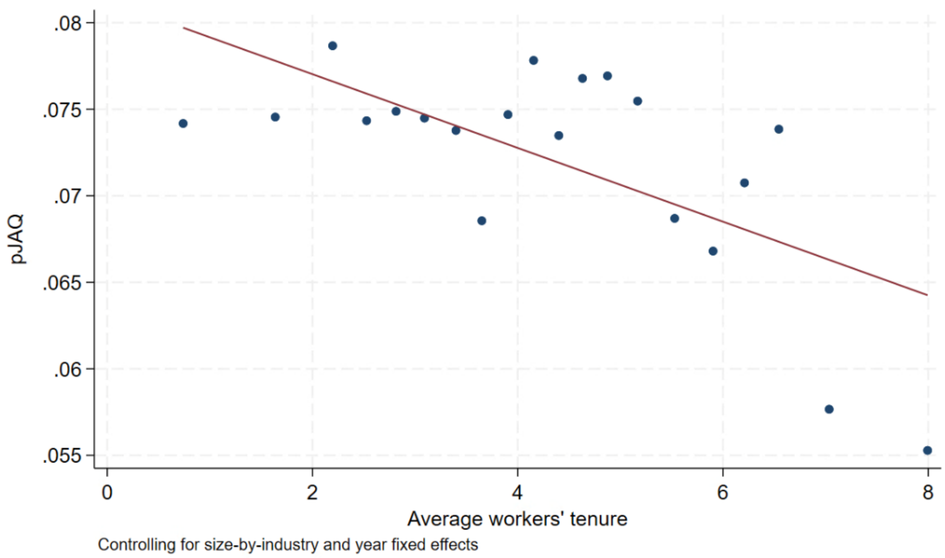

Figure 4.2. Firm-level average match quality (p*JAQ*) and workers' tenure

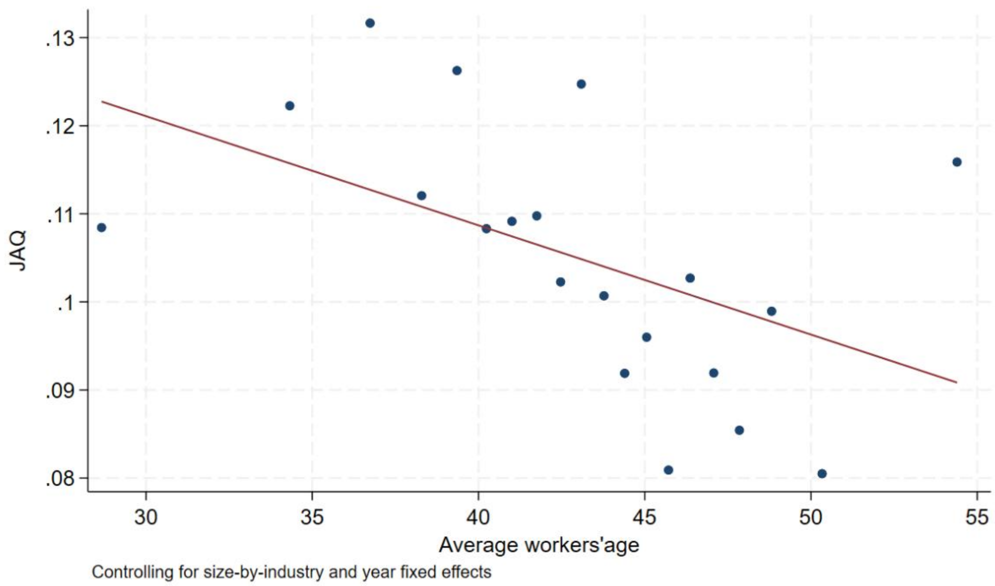

Figure 4.3. Firm-level fraction of well-matched employees (*JAQ*) and workers' age

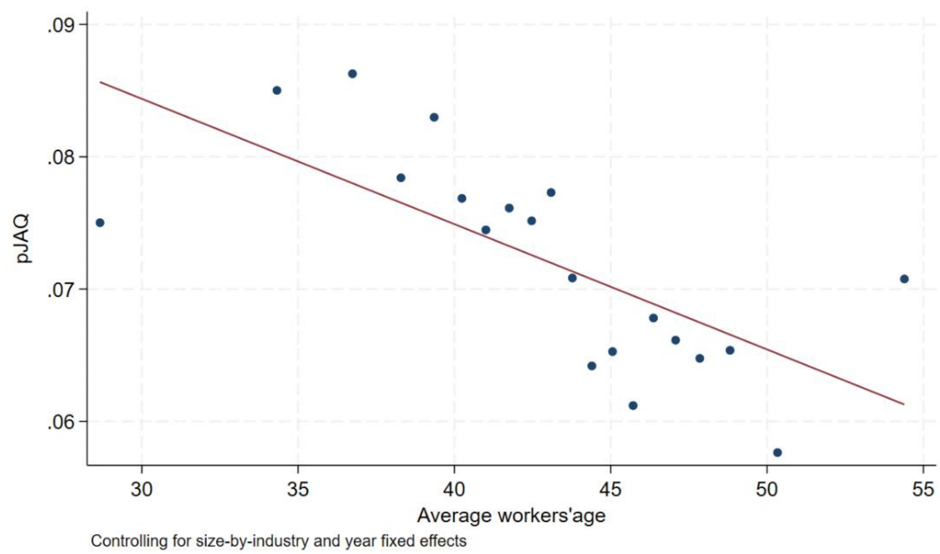

Figure 4.4. Firm-level average match quality (p*JAQ*) and workers' age

Figures 4.5 and 4.6 show that match quality has a strong positive relationship with employees' education, as measured by the firm's fraction of workers with a college degree, consistent with evidence from Sweden and Portugal. This highlights that the Italian evidence on the relationship between match quality and employees' education can be regarded as an exception.

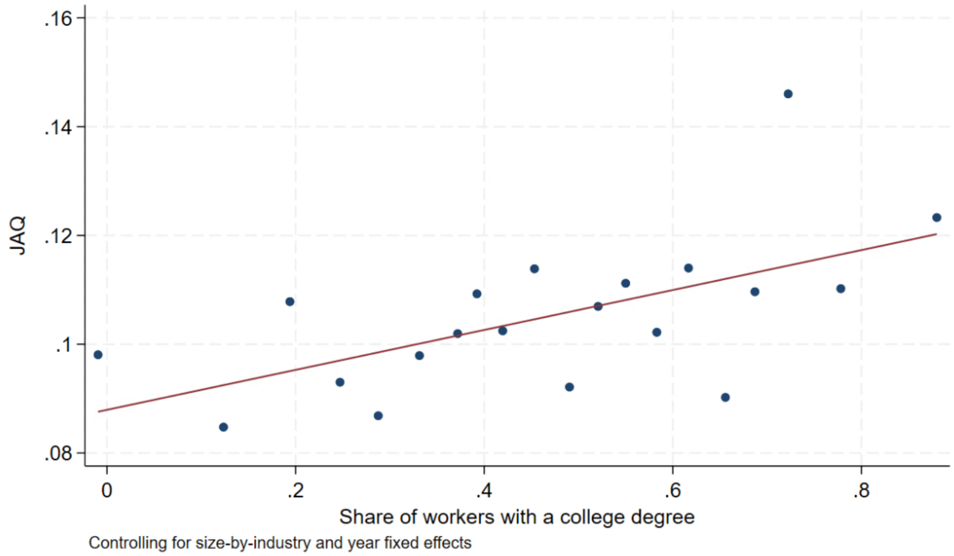

Figure 4.5. Firm-level fraction of well-matched employees (*JAQ*) and workers' education

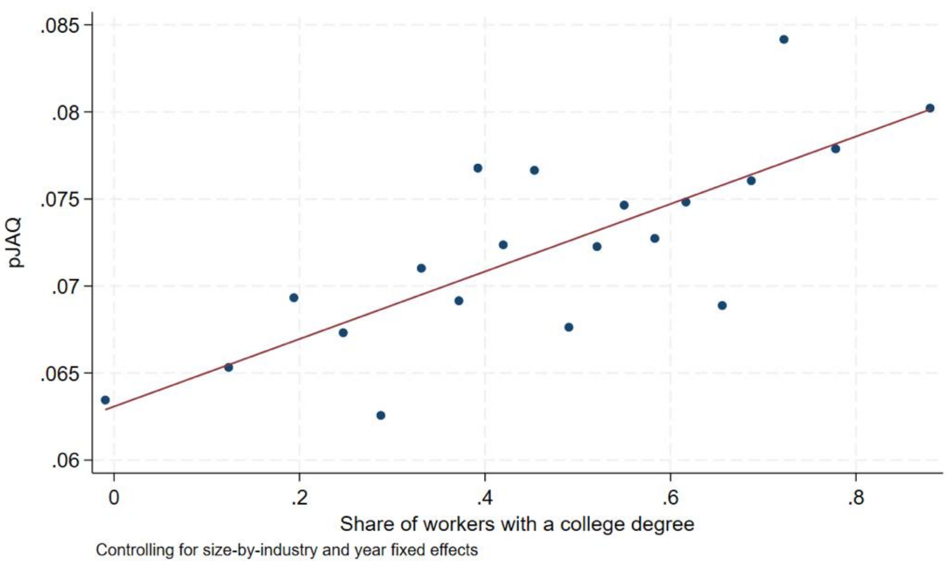

Figure 4.6. Firm-level average match quality (p*JAQ*) and workers' education

Figures 4.7 and 4.8 show the relationship between match quality and firm size, measured by the logarithm of the number of employees. Unlike the corresponding findings for Sweden and Portugal, and similarly to those for Italy, match quality appears to be negatively related to firm size, but the relationship is weak.

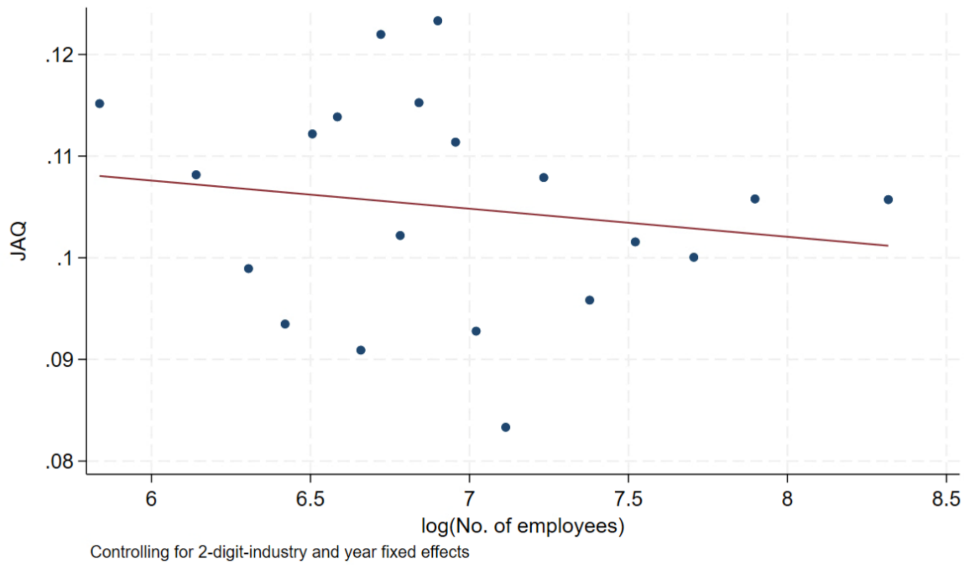

Figure 4.7. Firm-level fraction of well-matched employees (*JAQ*) and number of employees

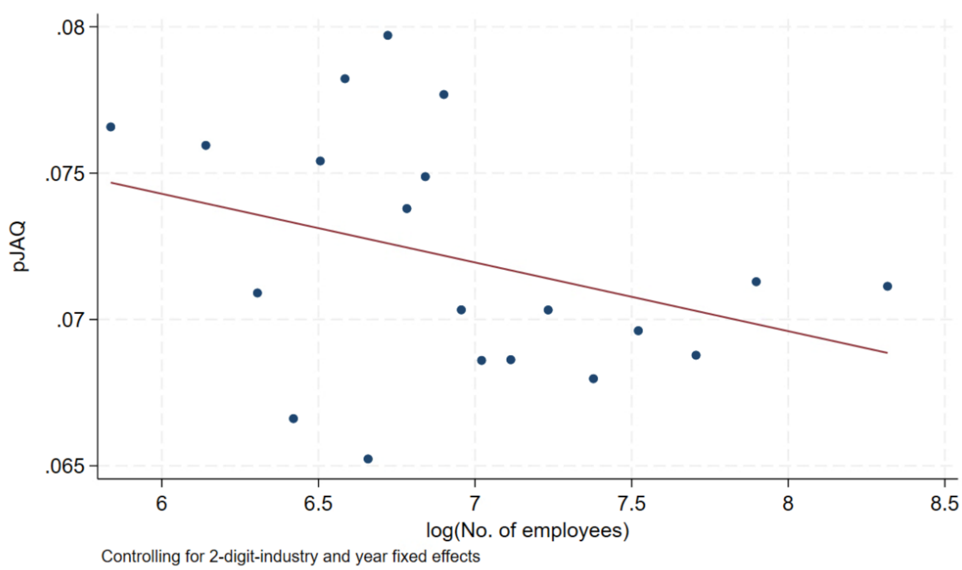

Figure 4.8. Firm-level average match quality (p*JAQ*) and number of employees

The variety of results that we observe across countries regarding the bivariate relationship between match quality and firm size may depend on the fact that greater firm size is associated with factors that enable firms to achieve better matching of their employees to jobs, but also factors that play the opposite role: on one hand, larger firms may afford to bear the fixed costs of high-quality managers and sophisticated personnel policies; on the other, they also feature a greater variety of occupations, hence a more complex job assignment problem than smaller firms.

3.4.6 Match quality and firm-level outcomes

As for key outcomes of interest, Figures 4.8 and 4.9 reveal a strong, positive relationship of firm-level measures of match quality with productivity, as measured by the logarithm of value added per worker, and Figures 4.10 and 4.11 show a weak, statistically insignificant relationship with profitability, as measured by return on assets (ROA). As will be seen from the regression analysis, the positive relationship with productivity also emerges when sales per worker are used as the productivity measure.

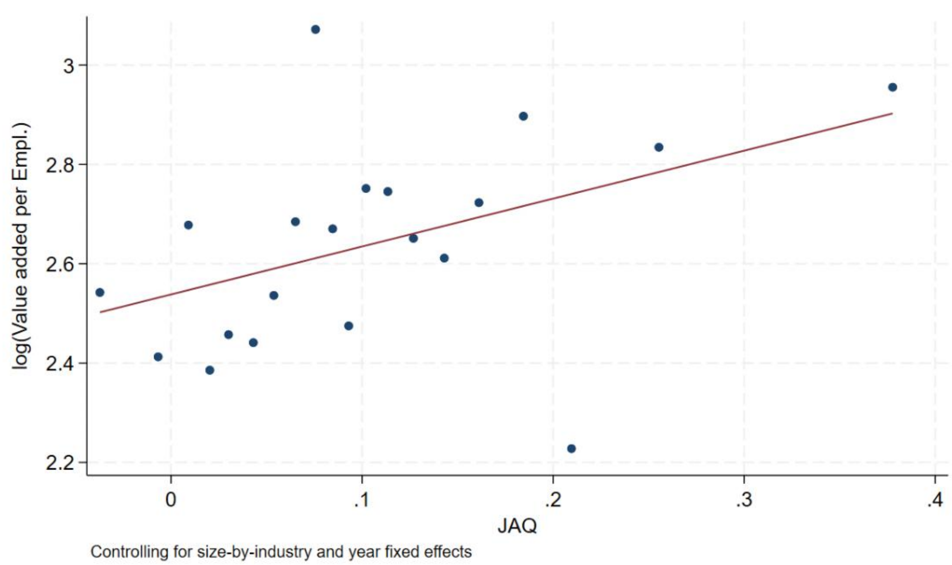

Figure 4.9. Firm-level fraction of well-matched employees (*JAQ*) and firm productivity

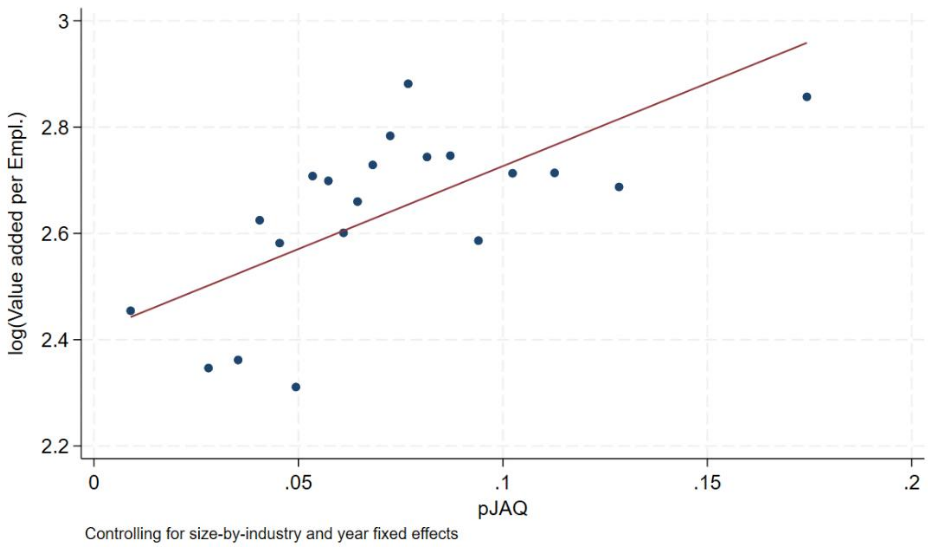

Figure 4.10. Firm-level average match quality (p*JAQ*) and firm productivity

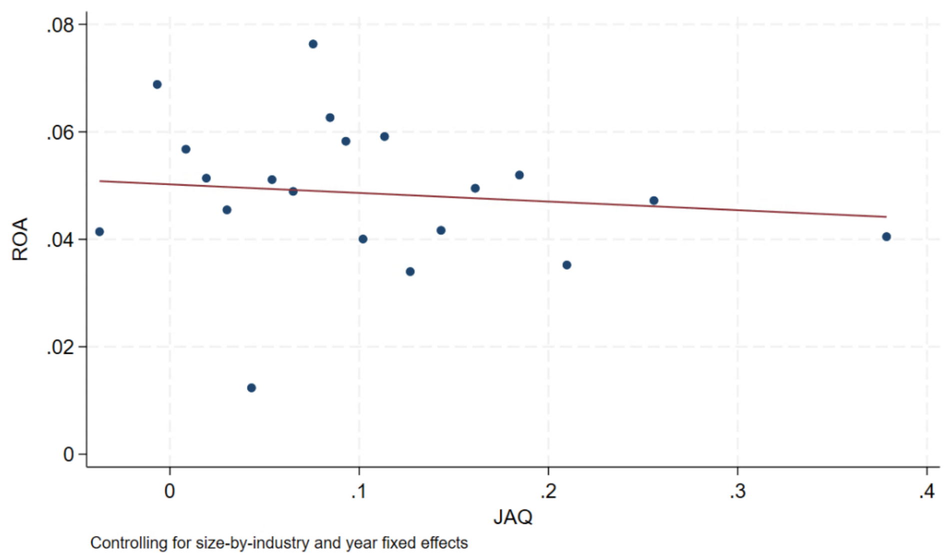

Figure 4.11. Firm-level fraction of well-matched employees (*JAQ*) and firm profitability

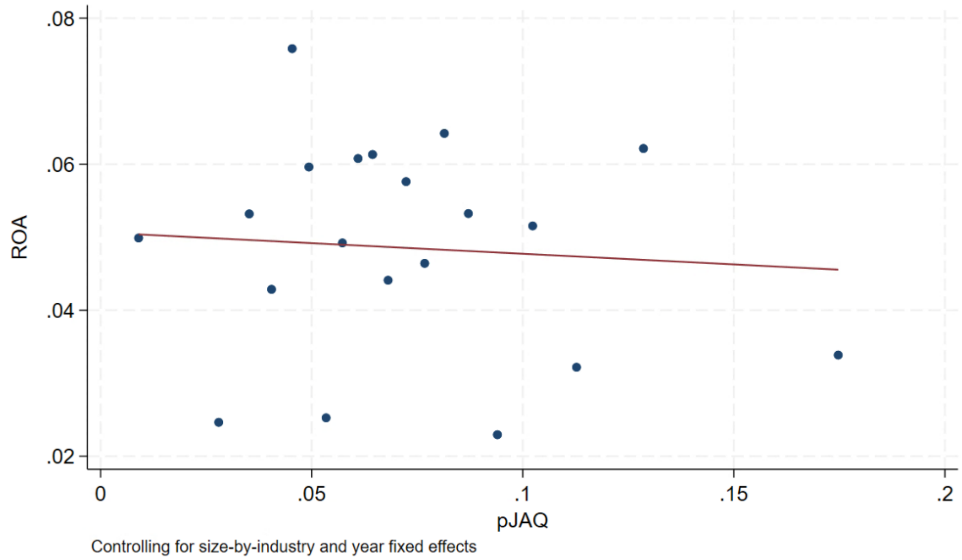

Figure 4.12. Firm-level average match quality (p*JAQ*) and firm profitability

The two tables below report regressions of measures or productivity or profitability with *JAQ* (Table 4.4) or *pJAQ* (Table 4.5). Our measures of productivity are the logarithm of sales per employee or value added per employee. In contrast, we use ROA (net earnings scaled by total assets) as the measure of profitability. In columns 1-3 of Panel A, we control only for year fixed effects and do not detect any meaningful correlation. In columns 4-5 of Panel A, where we include year-industry-size fixed effects, the relationship with productivity becomes positive and significant, whereas that with profitability remains not significantly different from zero, consistent with the evidence for Sweden.

In Panel B, we control for an extensive set of worker-level and firm-level characteristics, namely average age, average tenure, the fraction of college graduates, the logarithm of firm size, and the logarithm of capital intensity. The relationship between the two measures of match quality and productivity remains positive and significant, though it weakens slightly in the most conservative specifications when JAQ is used as the key regressor.

Table 4.4 - Firm productivity, profitability, and fraction of well-matched employees (*JAQ*)

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A	log(sales/ emp)	log(va/emp)	ROA	log(sales/em p)	log(va/emp)	ROA
JAQ	-0.478	0.216	0.017	1.026***	1.034***	-0.111
	(0.594)	(0.573)	(0.112)	(0.386)	(0.395)	(0.169)
Year	Yes	Yes	Yes	No	No	No
Yr.×Ind. ×Size	No	No	No	Yes	Yes	Yes
PANEL B						
JAQ	1.016**	1.073***	-0.128	0.420*	0.583*	-0.134
	(0.395)	(0.406)	(0.177)	(0.246)	(0.316)	(0.165)
Age	-0.005	0.002	0.002	-0.014	-0.005	0.001
	(0.007)	(0.007)	(0.002)	(0.004)	(0.005)	(0.002)
Tenure	0.058	0.068	-0.012	0.025	0.043	-0.011
	(0.029)	(0.028)	(0.008)	(0.018)	(0.022)	(0.007)
Frac. Coll.	0.466	0.431	0.025	-0.144	-0.095	0.011
	(0.207)	(0.209)	(0.050)	(0.125)	(0.153)	(0.044)
Log(Size)				-0.124	-0.139	-0.021
				(0.049)	(0.059)	(0.015)
Log(K/L)				0.650	1.623	0.014
				(0.023)	(0.437)	(0.021)
Year	Yes	Yes	Yes	No	No	No
Yr.×Ind. ×Size	No	No	No	Yes	Yes	Yes

Table 4.5 - Firm productivity, profitability, and average match quality (p*JAQ*)

PANEL A	(1) log(sales/emp)	(2) log(va/emp)	(3) ROA	(4) log(sales/emp)	(5) log(va/emp)	(6) ROA
pJAQ	-2.715	-0.342	0.309	3.318***	3.548***	-0.018
	(1.614)	(1.539)	(0.275)	(1.161)	(1.153)	(0.448)
Year	Yes	Yes	Yes	No	No	No
Yr.×Ind. ×Size	No	No	No	Yes	Yes	Yes
PANEL B						
pJAQ	3.293***	3.752***	-0.055	1.729**	2.461***	-0.068
	(1.202)	(1.196)	(0.477)	(0.729)	(0.883)	(0.441)
Age	-0.003	0.004	0.002	-0.014	-0.005	0.001
	(0.007)	(0.007)	(0.002)	(0.004)	(0.005)	(0.002)
Tenure	0.060	0.070	-0.012	0.025	0.043	-0.011
	(0.029)	(0.028)	(0.008)	(0.018)	(0.022)	(0.007)
Frac. Coll.	0.440	0.398	0.022	-0.144	-0.095	0.011
	(0.209)	(0.210)	(0.049)	(0.125)	(0.153)	(0.044)
Log(Size)				-0.124	-0.139	-0.021
				(0.049)	(0.059)	(0.015)
Log(K/L)				0.650	0.551	0.013
				(0.023)	(0.026)	(0.021)
Year	Yes	Yes	Yes	No	No	No
Yr.×Ind. ×Size	No	No	No	Yes	Yes	Yes

3.4.7 Time series patterns in match quality

For the Netherlands, the paucity of data makes aggregate measures of match quality quite noisy, so one should be cautious regarding their informativeness regarding the time-series behavior of aggregate match quality in the Dutch economy. Nevertheless, Figures 4.12 and 4.13 highlight that the Netherlands experienced a strong rebound of aggregate measures of match quality in the wake of the 2010-11 sovereign debt crisis and then a trend improvement in match quality, in sharp contrast to the absence of any trend in Sweden and the declines observed in Portugal and especially in Italy over the same time period.

These international differences in the trends of aggregate match quality may reflect the different exposure of the respective countries to the recession triggered by the euro-area sovereign debt crisis of 2010-11. The Netherlands had much lower exposure to the crisis, so that it featured fluctuating but generally positive macroeconomic performance over the subsequent decade, with steady growth (around 1-2.5%), marked by falling unemployment and strong export sectors, supported by innovation and high-tech industries. Sweden was largely immune to the euro-area sovereign debt crisis because it was outside the euro area. In contrast, the Portuguese and Italian economies suffered severe macroeconomic shocks and financial disruptions during the euro-area sovereign debt crisis, which likely destroyed productive job-worker matches and impaired their firms' ability to develop new ones for an extended period.

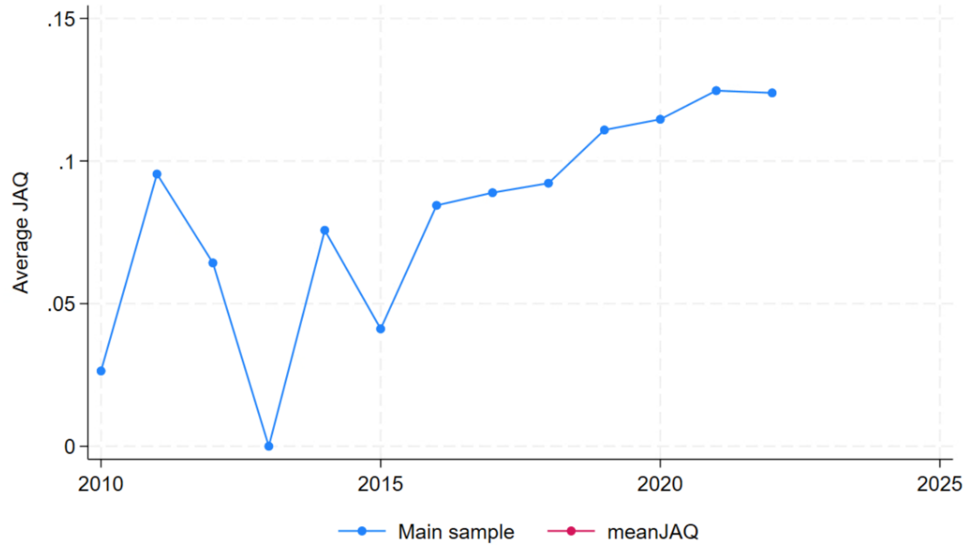

Figure 4.13. Aggregate fraction of well-matched employees (*JAQ*) in the main sample, 2010-22

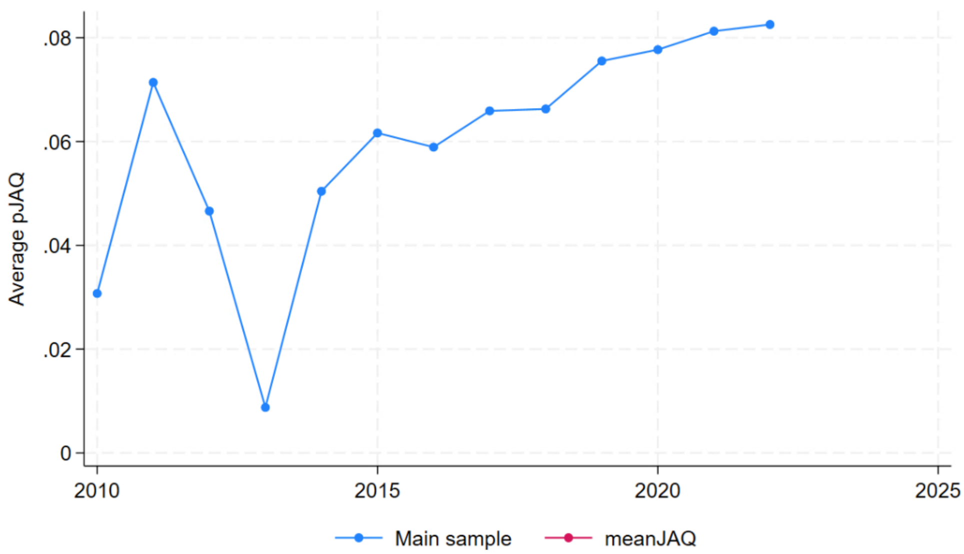

Figure 4.14. Aggregate average match quality (p*JAQ*) in the main sample, 2010-22

4. Conclusions

This deliverable has developed and applied a novel, machine–learning–based measure of job allocation quality (*JAQ*) to matched employer–employee administrative data from Sweden, Portugal, Italy, and the Netherlands. By inferring job–worker match quality from observed allocation patterns in highly productive firms, the methodology avoids reliance on externally imposed skill taxonomies or educational benchmarks and can be implemented using administrative data that are widely available across European countries. This makes the methodology particularly well-suited for longitudinal analysis and for detecting empirical regularities in job match quality across different countries, business cycle phases, and regulatory settings. The main value of the report lies in the substantive empirical findings that emerge consistently across very different institutional and labor market contexts: they are summarized in the following table.

Table 5. Key cross-country patterns

Early-career gains: Match quality improves sharply in the first 5–10 years of experience/tenure, then it plateaus.
Wage premiums: Better-matched workers earn 1.6–3.9% more, even after controlling for firm/worker fixed effects.
Firm-size threshold: In three of the four countries, match quality rises with firm size but plateaus or declines beyond 20–30 employees.
Role of workers' human capital: In three of the four countries, match quality rises with the fraction of employees with university degrees.
Productivity link: JAQ correlates positively with firm productivity (value added/employee) but weakly with profitability.
Cyclical sensitivity: Match quality declines during recessions, especially in countries more severely affected by the sovereign debt crisis.

A first robust finding is that job match quality improves sharply early in workers' careers and then plateaus. In all countries where individual job histories can be observed, match quality rises steeply during the first years of labor market experience and early job tenure, with much smaller gains thereafter. This pattern is remarkably similar across countries despite differences in education systems, labor market institutions, and employment protection regimes. It points to firms' learning about workers' skills and workers' learning about their comparative advantage as key drivers of match formation, suggesting that allocative efficiency gains are mostly realized early in careers.

A second robust result is that better job matches are systematically associated with higher wages at the individual level. Across countries with available earnings data, workers whose observed occupation is more closely aligned with their predicted suitability earn higher wages, even after controlling for observable characteristics, occupations, firms, and (in more demanding specifications) worker fixed effects. Although the magnitude of the estimated wage premium varies across countries and specifications, the direction and statistical significance of the relationship are stable. This provides strong validation that *JAQ* captures economically meaningful variation in match quality rather than mechanical correlations driven by worker or firm heterogeneity.

Third, firms with a larger fraction of highly educated workers (i.e., those with a university degree) feature better match quality in three of the four countries analyzed, suggesting that more meritocratic firms attract a more educated workforce. The exception is Italy, where the relationship is positive for low levels of the fraction of highly educated employees, but then turns sharply negative, resulting in an overall negative correlation.

Fourth, firms with higher job allocation quality are consistently more productive. At the firm level, *JAQ* is positively and significantly correlated with labor productivity, as measured by value added or sales per employee, in *all* the countries analyzed. This relationship is robust to controlling for firm size, industry, workforce composition, occupation structure, and fixed effects, indicating that efficient internal allocation of human capital is an independent determinant of firm performance. By contrast, the relationship between match quality and profitability is weaker and often statistically insignificant, suggesting that productivity gains from better matching are largely passed on to workers in the form of higher wages rather than accruing to firm owners.

A further result that holds across countries is that match quality is closely related to firm scale and organizational maturity. In all settings, match quality increases strongly with firm size up to a threshold—typically around 20–30 employees—after which additional gains level off. Older firms also tend to exhibit higher match quality. These patterns indicate the presence of fixed organizational costs in developing effective personnel management and job assignment practices, implying that small and young firms face structural disadvantages in efficiently allocating workers.

Alongside these robust findings, the analysis highlights dimensions along which results are more country-specific, most notably in the time-series behavior of job match quality, where differences align closely with countries' exposure to major macroeconomic shocks—especially the euro-area sovereign debt crisis. While aggregate JAQ is relatively stable over time in all four countries, its time-series pattern in economically meaningful ways. Sweden, which is outside the euro area, exhibits clearly pro-cyclical movements in match quality, with a marked decline during the global financial crisis and a subsequent recovery, consistent with its exposure to external demand shocks rather than to euro-area sovereign stress. By contrast, Portugal and Italy—both at the center of the euro-area sovereign crisis—display more persistent and dramatic dynamics in match quality during the crisis years, with some evidence of rapid reallocation during the post-crisis recovery in Portugal but not in Italy. In these countries, the tight credit conditions triggered by the crisis appear to have slowed the adjustment of worker–job matches, so that, at least for Italy, job mismatch became more structural during this period. The Netherlands occupies an intermediate position: although exposed to the euro-area crisis, it experienced less severe sovereign stress and maintained stronger labor market adjustment mechanisms, which is reflected in smoother and less persistent fluctuations in match quality. Taken together, these patterns indicate that the response of occupational mismatch to macroeconomic shocks depends crucially on the interaction between the nature of the shock and country-specific institutional and financial constraints.

Other country-specific patterns also emerge in the distribution of match quality across sectors and occupations, reflecting differences in industrial structure, task complexity, and occupational classification systems. Similarly, the relationship between match quality and demographic or contractual characteristics—such as gender, immigrant status, or contract type—is weaker and less consistent than suggested by traditional mismatch indicators. This indicates that allocative inefficiencies captured by *JAQ* are not primarily driven by observable demographic characteristics but by deeper features of career histories, task complexity, and firm-level assignment practices. For instance, it is striking that Italy is the only country that features both a distinctively negative trend in match quality and a decreasing relationship between match quality and the fraction of employees with tertiary education, suggesting the Italian firms' assignment practices tend to be unsuited to putting highly educated workers to their best possible use.

Taken together, the evidence in this report points to several broad conclusions. Job mismatch is neither random nor immutable: it declines systematically with experience, improves with firm capabilities, and is tightly linked to productivity at both the worker and firm levels. At the same time, much of the allocative inefficiency in labor markets appears to arise early in careers and within smaller or less established firms, where learning and organizational constraints are most binding. From a policy perspective, these findings suggest that interventions aimed at improving job matching may be most effective when they target early career stages, facilitate learning and mobility, and strengthen managerial and organizational capabilities within firms, rather than focusing exclusively on formal education mismatches. From a research perspective, the strong consistency of the core results across diverse European labor markets supports the external validity of the JAQ approach and highlights its potential as a scalable tool for monitoring skills mismatch.

TRAILS

Enabling data analytics for actions tackling skills shortages & mismatch

References

Baley, I., Figueiredo, A., & Ulbricht, R., 2022. Mismatch cycles. *Journal of Political Economy* 130(11), 2943–2984.

Bowlus, A. J., 1995. Matching workers and jobs: Cyclical fluctuations in match quality. *Journal of Labor Economics* 13(2), 335–350.

Coraggio, L., Langella, M., Miano, A., Pagano, M., Petterson, M. S., Pezone, V. and Scognamiglio, A., 2024. Report on research on skills mismatch. Deliverable D1.1, TRAILS Project.

Coraggio, L., Pagano, M., Scognamiglio, A., & Tåg, J., 2025. JAQ of all trades: Job mismatch, firm productivity and managerial quality. *Journal of Financial Economics*, Volume 164, 103992, February 2025.

Fredriksson, P., Hensvik, L., & Skans, O. N., 2018. Mismatch of talent: Evidence on match quality, entry wages, and job mobility. *American Economic Review* 108(11), 3303–3338.

Guvenen, F., Kuruscu, B., Tanaka, S., & Wiczer, D., 2020. Multidimensional skill mismatch. *American Economic Journal: Macroeconomics* 12(1), 210–244.

Oreopoulos, P., von Wachter, T., & Heisz, A., 2012. The short- and long-term career effects of graduating in a recession, *American Economic Journal: Applied Economics*, 4(1), 1-29.

Şahin, A., Song, J., Topa, G., & Violante, G. L., 2014. Mismatch unemployment. *American Economic Review*. 104(11), 3529-3564.

Enabling Data Analytics for Actions
Tackling Skills Shortages & Mismatch