EIEF WORKING PAPER SERIES

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 22/05

April 2022

JAQ of All Trades: Job Mismatch, Firm Productivity and Managerial Quality

By

Luca Coraggio
(University of Naples Federico II)

Marco Pagano
(University of Naples Federico II and EIEF)

Annalisa Scognamiglio
(University of Naples Federico II)

Joacim Tåg
(Research Institute of Industrial Economics (IFN))

JAQ of All Trades: Job Mismatch, Firm Productivity and Managerial Quality*

Luca Coraggio
University of Naples Federico II

Marco Pagano
University of Naples Federico II

Annalisa Scognamiglio
University of Naples Federico II

Joacim Tåg
Hanken School of Economics and IFN

April 2023

ABSTRACT

We present a novel measure of job-worker allocation quality (*JAQ*) by exploiting employer-employee data with machine learning techniques and validate it in various ways. Our measure correlates positively with earnings and negatively with separations over individual workers' careers. At firm level, it increases with competition, non-family firm status, workers' human capital and has a robust correlation with productivity. The quality of rank-and-file workers' job matches responds positively to improvements in management quality. *JAQ* can be constructed for any employer-employee data including workers' occupations, and used to explore research questions in organization and labor economics, as well as in corporate finance.

Keywords: jobs, workers, matching, mismatch, machine learning, productivity, management.
JEL Codes: D22, D23, D24, G34, J24, J31, J62, L22, L23, M12, M54.

*We are grateful for insightful suggestions and comments from Ramin Baghai, Kirill Borusyak, Vicente Cuñat, Morten Grindaker, Daniel Halvarsson, Alex Xi He, Hans Hvide, Simon Ek, Isil Erel, Anastassia Fedyk, Jessica Jeffers, Tullio Jappelli, Camelia Kuhnen, Mikael Lindahl, Topi Miettinen, Ahmed Ameya Prapan, Raffaele Saggio, Elia Sartori, Elena Simintzi, Rune Stenbacka, Marko Terviö and participants at the 2023 AFA Meetings, the 5th Bank of Italy-CEPR workshop on labour market policies and institutions, 2022 EFA Meetings, EEA-ESEM Congress and WFA Conference, 2021 Brucchi Luchino Workshop and ES Winter Meetings, the Labor and Finance Online Seminar, the Bristol-Exeter-Lancaster seminar, and at Ca' Foscari University of Venice, CSEF, EIEF, Hanken, Helsinki GSE, IFN, Swedish Conference in Economics, Ratio, UNC Kenan-Flagler Business School, and Örebro University. Marco Pagano acknowledges funding from the Einaudi Institute for Economics and Finance (EIEF) and the Italian Ministry for University and Research (MUR). Work on this paper by Annalisa Scognamiglio has been supported by the Modigliani Research Grant awarded by the UniCredit Foundation. Joacim Tåg thanks the Marianne and Marcus Wallenberg Foundation (2015.0048, 2020.0049), Torsten Söderbergs Stiftelse (E31/18, ET2/20), and Jan Wallanders och Tom Hedelius stiftelse samt Tore Browaldhs stiftelse (P22-0094) for financial support. E-mails: luca-coraggio@hotmail.com, pagano56@gmail.com, annalisa.sco@gmail.com, and joacim.tag@ifn.se.

1 Introduction

Matching workers to their best possible job is of paramount importance for firms and workers alike: hiring or firing the wrong people, or matching employees to the wrong task may sap a firm's productivity at least as much as it damages the employees' careers, by slowing down their skill acquisition process, reducing their wage growth and possibly forcing them to switch to another employer. In fact, the ability to match workers to the right jobs is a hallmark of a good manager, on a par with the set of human resources management skills that have come to be known as "managerial practices" and shown to contribute significantly to firm productivity (Bloom and Van Reenen, 2007, 2010; Bloom et al., 2013, 2019; Scur et al., 2021). However, managerial practices measured by existing research regarding human resource management focus on workers' incentives, and neglect choices regarding the allocation of workers to jobs. This shortcoming probably reflects the fact that the measurement of managerial practices has so far been based on replies to surveys regarding the way managers run their firms' operations, monitoring, incentives and targets, and it would be very difficult to use self-reported information to evaluate whether managers allocate workers to their best possible use within the firm.

This is where our paper comes in: we develop a novel measure of job allocation quality (*JAQ*, for short) by applying machine-learning (ML) techniques to administrative employer-employee data rather than building on responses to questionnaires, and validate our measure in a variety of ways, which also bear witness to its versatility. *JAQ* is built in four steps. First, we estimate via a ML algorithm how workers' characteristics map into jobs, using as a benchmark for match quality the allocation of workers to jobs in the most productive firms. This is similar in spirit to how Bloom and Van Reenen (2007) benchmark management practices against standards set by a leading management consulting firm and how Fredriksson et al. (2018) gauge mismatch among junior workers by the distance between their skills and those of senior workers. Second, we predict worker suitability for each job based on the function estimated via the ML algorithm. Third, we measure whether a worker's actual job coincides with her most suitable job (*eJAQ*). Fourth, we average the *eJAQ* measure across all the employees of each firm to construct our firm-level measure of match quality–*JAQ*.

To validate these measures, we start by testing whether workers' careers benefit from their job match quality, in line with predictions from labor economics. It is natural to expect employees to be increasingly assigned to more suitable jobs over their career, as managers learn about their characteristics (Fredriksson et al., 2018), and employees themselves fine-tune their skill set via on-the-job training (Guvenen et al., 2020). Also, insofar as better job allocation enhances productivity, workers can be expected to appropriate at least part of the gain in the form of higher wages. Both of these predictions find support in our data. The goodness of worker-job matches (*eJAQ*) rises significantly over the span of working lives, the largest gain occurring in the first few years: this accords with the intuition that learning is faster for junior workers, and their reallocation to more suitable jobs is easier than for senior employees (Farber and Gibbons, 1996). Moreover, workers allocated to their most suitable job are found to earn significantly more than mismatched workers with the same characteristics or with the same job, and to be less likely to switch to a new employer. Both of these findings dovetail with those reported by Fredriksson et al. (2018), despite the differences in methodology and sample used.

Next, we validate the firm-level measure of job allocation quality (*JAQ*) by showing that it correlates positively with market competition, non-family firm status, workers' human capital and, most importantly, with productivity, just as management practices do. In particular, we show that *JAQ* has a significant, sizeable and robust positive correlation with log value added and log sales per employee, even upon controlling for the firm-level variables generally associated with productivity (industry, capital and labor, ownership) and for the workers' characteristics used to predict *JAQ*. A possible concern about this finding is that it may be vitiated by circularity, as we first train the ML algorithm to assign workers to jobs based on data for the most productive firms, and then investigate whether *JAQ* correlates with firm productivity. The first counter to this criticism is that the correlation between *JAQ* and productivity is estimated dropping the observations used to train the ML algorithm. However, the correlation between *JAQ* and productivity may still arise because the assignment rule is estimated on the most productive firms, so that its estimation error may correlate by construction with firm productivity. To face this concern, we retrain our algorithm on a random sample of firms, and find that the resulting measure of job allocation quality still correlates positively and significantly with productivity. Moreover, we perform a placebo test where firms' actual productivity is replaced by a noise variable to dispel the concern that the relationship between *JAQ* and firm performance is purely mechanical.

Finally, we explore whether the allocation of rank-and-file workers responds positively to improvements in management quality. Upon constructing two distinct *JAQ* measures for rank-and-file workers and for managers, the former turns out to be positively and significantly correlated with the latter, as well as with the average experience of the firm's management team, even when only within-firm variation is exploited. Interestingly, the quality of rank-and-file workers' allocation rises significantly when turnover leads to an improvement of the allocation of management, which tends to occur in the wake of a deterioration in the allocation of rank-and-file workers. Conversely, when managerial turnover results in a worse assignment of managers, it is associated with a persistent disruption in the allocation of rank-and-file workers. These results persist even when the analysis is restricted to changes in management associated with the death of an incumbent manager, although in this subsample the estimates are imprecise due to the paucity of observations.

On the whole, our main contribution is to provide a new measure of mismatch in the labor market, which can be constructed for any country and time period where survey or administrative data on workers' characteristics and job titles exist, without requiring expensive targeted surveys (Bloom and Van Reenen, 2007; Bloom et al., 2019) or detailed expert evaluations of the skills required for each job, such as those contained in the O*Net data (Lise and Postel-Vinay, 2020; Guvenen et al., 2020). Our results illustrate that this measure lends itself to be effectively deployed in three distinct strands of research: the literature on firm productivity, labor research on match quality, and work in corporate finance on the importance of managers for firms.

As already mentioned, the measure that we propose complements the research on the role of managerial practices and of human capital for firm productivity (Scur et al., 2021). Most closely related are Ichniowski et al. (1997), Bloom and Van Reenen (2007), Cornwell et al. (2021), Bloom et al. (2013) and Bloom et al. (2019), who study how management practices relate to productivity, Bender et al. (2018), who investigate the relationship between productivity, management practices, and employee ability, and the study by Fox and Smeets (2011) on the role of workers' quality in explaining the dispersion in productivity. Our distinctive contribution here is to focus on managerial policies governing the allocation of workers to jobs within the firm, and relate job-worker mismatch to firm productivity. Our measure of mismatch is likely to be informative not only about the role of labor misallocation at the firm level, but also at higher level of aggregation, for example to shed light on how technological innovation and regulatory changes influence labor market efficiency and overall productivity. Our measure could also be used to investigate the "cleansing effect" of recessions, that is, how strong the reallocation of workers is during economic downturns (Bowlus, 1995; Baley et al., 2022).

As for the labor literature on mismatch, our work complements research on how workers match with firms (Jovanovic, 1979; Cahuc et al., 2006; Postel-Vinay and Robin, 2002; Chiappori and Salanié, 2016; Eeckhout and Kircher, 2018; Adenbaum, 2023; Pastorino, 2023) and with tasks (Perry et al., 2016; Lindenlaub, 2017; Deming and Kahn, 2018; Lise and Postel-Vinay, 2020; Guvenen et al., 2020; Ocampo, 2022), and on how managers match with firms (Terviö, 2008; Lippi and Schivardi, 2014; Benson et al., 2019; Bandiera et al., 2020). Perhaps most closely related to our study is Fredriksson et al. (2018), who investigate the impact of job mismatch on starting wages and subsequent labor market outcomes, measuring mismatch as the absolute distance between senior workers' and new hires' talent. Our measure of job assignment quality differs from that used in this study in two main respects: (i) we rely on a ML-estimated function rather than on average characteristics of senior workers to determine the efficient allocation of workers across jobs;[1] (ii) by the same token, our method applies just as well to the allocation of senior workers as to that of junior ones, while the other method only applies to junior ones. The latter point is key to

[1] Other papers in the labor economics literature that rely on ML techniques include papers on the measurement of worker specialization (Ek, 2022) and on internal labor markets and hierarchies (Huitfeldt et al., 2021).

evaluate how job assignment quality correlates with firm productivity, which requires measuring job assignment quality for *all* the employees of each firm. More broadly, our new measure of mismatch in the labor market can be used to test the predictions of the above-mentioned search and matching models, and to assess labor market policies aimed at improving the efficiency of job matching and reducing unemployment, such as job search assistance programs or unemployment benefits. JAQ can also help investigate the role of match quality in determining wages and their dispersion, as well as in driving the process of job mobility, creation and destruction. Moreover, identifying the skills and characteristics associated with better job matches can inform policies and programs aimed at developing human capital, such as education and training initiatives.

Finally, our paper relates to the corporate finance literature, and particularly to an emerging strand of research that exploits ML to address research questions at the interface between labor and finance.[2] Examples include the appointment of board of directors (Erel et al., 2021), the screening of resumes in recruitment (Li et al., 2020), the measurement of corporate culture based on earnings call transcripts (Li et al., 2021), and the measurement of what managers do (Bandiera et al., 2020). Our results on the correlation between managerial quality and the match quality of rank-and-file workers directly relate to the research on managers' role in allocating human capital within the firm and in shaping workers' careers (Minni, 2023; Pastorino, 2023) and more generally for firm performance.[3] More broadly, our measure also has other potential uses in corporate finance. JAQ can be useful for investigating the role of human capital in private equity interventions and in mergers: for instance, it can provide a way to assess if worker reallocation from combining the workforce of two firms leads to higher productivity and lower costs. It can also shed light on the relationship between match quality and the financial returns on human capital investments, such as employee training, development programs, and recruitment strategies. Finally, incorporating match quality into firm valuation models can help better capture the intangible value of human capital, which may impact a firm's long-term prospects and value.

The road map reads as follows. The next section describes our data and Section 3 details how we construct *JAQ* and describes how it correlates with some firm characteristics. Section 4 relates *JAQ* to firm performance, and Section 5 explores the relationship between the quality of rank-and-file worker-job matches and the quality of management, especially in the wake of managerial turnover. The last section concludes.

[2]For surveys on how ML can be applied to economics research in general, see for instance Varian (2014), Mullainathan and Spiess (2017), or Abadie and Kasy (2019), Athey (2019).

[3]See, for instance, Bertrand and Schoar (2003), Malmendier and Tate (2005), Bennedsen et al. (2007), Malmendier and Tate (2009), Kaplan et al. (2012), Lazear et al. (2015), Mullins and Schoar (2016), Bandiera et al. (2018), and Bennedsen et al. (2020).

2 Data

To develop and estimate the *JAQ* measure proposed in this paper we use Swedish registry data. This data set is ideal for our purposes for at least two reasons. First, it allows us to observe for a relatively long period the entire population of workers and firms in Sweden, including a number of variables regarding workers' job histories, such as occupations and wages over their career. Second, despite their institutional differences, labor markets are surprisingly similar in their functioning in Scandinavian countries, Belgium, France, Germany, Italy, the Netherlands and the United States (Lazear and Shaw, 2009), which bodes well for the external validity of our results.

The bulk of our data come 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 1990–2010 interval but our analysis focuses on the 2001–10 interval since occupation information is not available prior to 2001. However, we draw on 1990–2000 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 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.

The firm-level variables drawn from LISA are firm age, 2-digit industry, size (measured by the number of employees), sales, and total assets, as well as ownership categories measured by indicators for the firm being a state-owned firm, a listed firm, or a family firm. Information on listed status is drawn from the Statistics Sweden's FRIDA database, and the indicator of family firm status is obtained by combining information on firm ownership from FRIDA with info on board members and CEOs from the Swedish Companies Registrations office and the multi-generational register on biological parent-child relationships. Following Keloharju et al. (2023), a family firm is defined as one managed or owned by at least two members of the same family.

We identify jobs based on international ISCO-88 (COM) 3-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 total, over our entire sample period over 90% of workers are sampled at least once.[4]

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 30 and 6,000. The lower bound is due to the sparsity of occupational information for firms with less than 30 employees: including these firms would introduce large noise in the estimation of the job-employee matching rule. These firms employ about 30% of the total reported workforce and on average report 2.9 employees per year. 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. These firms account for 20% of the total reported workforce, but are few: out of a total of 945,385 firms in the database, there are only 80 such firms, which drop to 11 in the industries retained in our analysis.

Our sample includes firms in three industries: (i) manufacturing; (ii) real estate, renting and business activities; and (iii) wholesale and retail, which include 62% of the firms and 70% of the employees present in the LISA database over our sample period (after applying the screen based on firm size), and thus employ far more workers than other industries.[5] Moreover, these industries feature the most complete and heterogeneous set of occupations in the Swedish economy: they include the greatest number of occupations, namely, 99%, 98% and 96% of the total 110 jobs, against a mean of 72% in other industries. The first two of these industries also feature a more diversified set of occupations than others, making the workers' assignment problem more relevant: the Herfindahl–Hirschman index measuring the concentration of occupations is 4% and 6% respec-

[4]See Tåg (2013) and Tåg et al. (2016) for additional details and descriptive statistics on occupations and hierarchical structures within firms.

[5]The excluded industries are: agriculture, hunting, forestry and fishing; mining; utilities; construction; hotels and restaurants, transport, storage, and communications; financial intermediation; public administration and defence; education; health and social work; other service activities.

tively, compared to a mean of 14% in other industries.[6] After applying these filters, our sample comprises 9,023 firms, employing a total of 1,541,343 employees over the 2001-10 period.

3 Measuring Job Assignment Quality

Suppose that managers strive to allocate workers to jobs so as to maximize productivity, by picking a job assignment function that maps observable worker and firm characteristics to jobs within the firm. The allocation can vary also depending on location and thus on features of the local labor market.However, firms may deviate from the most efficient assignment rule, incurring in errors that reduce their productivity, because of managerial shortcomings and/or information and learning frictions. As noted by Bloom and Van Reenen (2007), firms may also choose to not implement the optimal management practices because doing so may be costly to the manager and the manager may be entrenched. As external circumstances change, firms may also not immediately adjust to the new optimal assignment rule due to adjustment costs.

In order to judge the job allocation quality (*JAQ*) of a firm it is necessary to estimate the rule that managers use to assign workers to jobs. In principle, this can be done using a random subsample of firms. However, insofar as the rule maximizes productivity, the firms that apply it most rigorously should deviate less often from it and thus feature the highest productivity. Hence, the rule can be observed with the least noise for the most productive firms.

Thus, a key feature of our measure is that it is generated by benchmarking against match quality in the most productive firms. As such, it is consistent with any model of labor market search and worker assignment to firms that predicts higher match quality in more productive firms, such as Moen (1997), Cahuc et al. (2006) and Postel-Vinay and Robin (2002). In these directed search models, more productive firms can afford to pay higher wages, thus attracting higher quality workers and producing better match outcomes. Such productivity advantage may stem from managers being better at solving the multidimensional skill mismatch problem due to the discrepancy between the portfolio of skills required by an occupation and the portfolio of abilities a worker has to acquire those skills, as in Guvenen et al. (2020).

Accordingly, we use a machine learning (ML) algorithm to estimate the rule using only observations that refer to firms in the top decile of the productivity distribution. The benchmark provided by this ML prediction enables us to measure how close the job allocation adopted by any given firm is to that predicted by the estimated rule. Of course, the rule estimated by the ML algorithm is bound to be efficient only on average: firms are likely to condition their job-worker matches on more information than that available to us in estimating the algorithm. Hence, some

[6]This does not apply to the wholesale and retail industry, whose concentration index is 18%.

of the observed firm-level deviations from the estimated rule may reflect firm-specific information not captured by the algorithm rather than firm-level errors in applying the optimal rule.

3.1 Mapping workers' characteristics to jobs via machine learning

In our framework, managers use the job assignment rule $J = g(X,Z)$ to identify the job J to which each worker is best assigned, based on workers' observable characteristics, X, and on firm's characteristics Z. We do not observe g, but we can recover it by estimating the conditional probabilities $P(J|X,Z)$ for firms that are likely to adhere most closely to the rule, i.e., the most productive firms. We do not impose any particular restriction or parametric form on g, and allow for the possibility that firms with different characteristics rely on different rules. Hence, the conditional probabilities to be estimated are denoted by $P(J|X,Z)$.

For computational reasons, the sample is broken up in various subsamples to train the algorithm: this significantly reduces the estimation time compared to that required to estimate the algorithm on the full sample.[7] The sample split is based on firms' characteristics: firms are sorted across the three industries described above and three size classes, resulting in 9 size-industry bins. The three size classes are based on firms' median number of employees, N, over the sample period: (i) small ($30 \leq N \leq 50$); (ii) medium ($51 < N \leq 250$); (iii) large firms ($N > 250$). The algorithm is estimated separately for the firms in each size-industry bin, taking into account that firms in different bins may use different rules to match their employees to occupations: for instance, larger firms typically have more layers in their hierarchy than smaller ones, and manufacturing firms have a greater variety of occupations than those in wholesale and retail trade. Hence, this approach amounts to estimating conditional probabilities $P_Z(J|X)$, for $Z \in 1,\dots 9$, where the firm's characteristics Z is a variable identifying the size-industry bin a firm belongs to.

Within each size-industry bin, we define the "learning sample" used to estimate the conditional probabilities $P_Z(J|X)$ as the subsample of firms in the top decile of the productivity distribution. More precisely, in order to include in the learning sample only firms that are consistently more productive, for each size-industry bin we (i) estimate a model of value added per employee with firm fixed effects and calendar year effects, (ii) consider the distribution of fixed effects for firms present in the 2010 subsample, and (iii) select firms belonging to the top decile of this distribution. We then use 2010 data for these firms to train our algorithm: being the last year available in our sample, it contains the longest job histories that can be exploited to learn how firms allocate employees to jobs. Using data for these firms, we estimate bin-specific conditional probabilities $\hat{P}_Z(J|X)$ to predict workers' allocation to jobs in remaining firms – referred to as the "main sample" – within the corresponding bin.

[7] Attempts to perform the estimation on the full sample exceeded the 1,000-hours limit to computation time set by the Statistics Sweden server.

Table 1 compares the characteristics of the workers included in the main sample and in the learning sample: the workers included in the latter earn higher wages, are more educated, and have longer tenure and fewer days of unemployment than workers included in the former sample. These differences are consistent with the fact that the learning sample includes more productive firms, where workers can be expected to feature more productive matches, hence experience fewer separations.

Insert Table 1 here

Despite these differences, the two samples are sufficiently similar as to have common support: this is shown by Figure 1, which displays the distributions of the predicted wages for workers in the two samples. For both samples, the predictions are obtained from wage regressions estimated on the main sample, whose explanatory variables are the worker characteristics included in the ML algorithm. The figure shows that the support of the two distributions overlaps considerably, even though the distribution of the learning sample places more weight on high predicted wages than that of the main sample. This evidence supports our assumption that the learning sample can be used to estimate an allocation rule that is relevant for workers in firms included in the main sample.

Insert Figure 1 here

Within the learning sample, we estimate the conditional probability of each occupation via the Random Forests algorithm (Breiman, 2001).[8] There are three advantages to using Random Forests (RF) in our setting: (i) they are among the best performing algorithms for classification (Zhang et al., 2017);[9] (ii) they feature few-to-none tuning hyperparameters, dramatically reducing total estimation time;[10] (iii) they easily handle multi-class classification problems and mixed-type characteristics (continuous and categorical), which are relevant in our data.[11]

As occupations are not all equally frequent in the sample, we adjust our estimation procedure by forming a balanced subsample via bootstrap, under-sampling more frequent occupations, and use this subsample to train a random forest with 50 trees. This is repeated 100 times and the

[8]As implemented by Robnik-Šikonja and Savicky (2020), with the R language.

[9]Although some of the measures we build rely on the full set of estimated conditional probabilities $P_Z(J|X)$, our main measure of job assignment quality relies solely on workers' classification into their most suitable jobs. Moreover, we use a bagging procedure for estimation, which significantly mitigates possible calibration issues related to the estimation of conditional probabilities (Wallace and Dahabreh, 2012).

[10]Hyperparameters are parameters set by the researcher to control the learning process, such as the number of trees and the number of features selected at each a node in random forest algorithms. Compared with other algorithms, such as neural networks, random forests require fewer parameters to be specified, making them relatively easier to tune. This reduces the overall estimation time due to the limited need to estimate multiple models in order to choose the best performing one.

[11]To deal with categorical variables with a high number of levels, we use the coding proposed by Micci-Barreca (2001).

results from the 100 random forests are averaged together—a strategy that combines ideas from EasyEnsemble proposed by Liu et al. (2008) and Balanced RF in Chen et al. (2004).

We evaluate the performance of our algorithm via an average of the F1 scores, computed across jobs (labeled as the macro F score in Sokolova and Lapalme (2009)), with weights equal to job frequencies to address the unbalancedness of the sample.[12] The average F1 score is computed via a stratified 5-fold cross-validation: the learning sample is randomly partitioned in 5 subsamples, where each subsample has the same job frequencies as the initial sample, and the algorithm is trained using 4 subsamples and tested on the remaining one; the procedure is repeated until all of the 5 subsamples are used as a test set, so as to obtain a total of 5 pairs of weighted F1-scores (where each pair refers to a training set and to the corresponding test set); finally, these 5 weighted F1-scores are averaged. The average of the resulting F1 scores is 78% when computed for the training set and 69% when computed for the test set. This performance is reassuring, considering that a random allocation of workers to jobs would at most achieve an average weighted F1-score of $2/(K+1)$, where K is the total number of jobs. Since the minimal number of jobs in our training and test sets is 38[13], the maximal weighted F1 score resulting from a random allocation of workers to jobs in our sample would be 5.1% at most.[14]

To characterize our algorithm, we explore the role that each worker characteristic plays in identifying the allocation of jobs across workers. To this purpose, we compute the explanatory power of each of the workers' features used in the random forest algorithm, i.e., its discriminatory power in the correct classification of the instances, as described in Robnik-Šikonja (2004) and Robnik-Šikonja and Savicky (2020). Figure 2 displays a box plot of this measure for all the features used in the ML algorithm, which are listed on the horizontal axis.

Insert Figure 2 here

On the whole, Figure 2 highlights that the type and level of education play the most important role in job allocation, far more than tenure, experience, occupation-specific and industry-specific

[12] The F1 score for a given class is computed as the harmonic mean of the estimator's precision and recall scores for such class. The precision score is defined as the ratio between the number of instances correctly identified as belonging to the class and the total number of instances that the estimator attributes to the class: it indicates the ability to estimate the class "precisely". The recall score is defined as the ratio between the number of instances correctly identified as belonging to the class and the total number of instances belonging to the class: it indicates the ability of the estimator to retrieve instances of that class.

[13] This is because in estimation strategy we split the full sample in size-industry bins. The minimal number of unique jobs in a bin is 38.

[14] This can be seen as follows. Denote job frequencies by π_k, $k = 1 \dots K$. If the algorithm were to assign workers to jobs at random with equal probability, the probability of assigning a worker to a given job is $1/K$. Hence, the precision and recall for class k, in large samples, are roughly π_k and $1/K$, respectively, so that the F1 score for class k is $2\pi_k/(K\pi_k+1)$, and the weighted F1 score is $2\sum_k \pi_k^2/(K\pi_k+1)$. The maximal value of this expression is $2/(K+1)$, which is achieved when $\pi_s = 1$ for some s and $\pi_k = 0$ for $s \neq k$.

experience, suggesting that generic human capital is more important than firm-specific one in job-worker matching. However, the numerous outliers that can be observed for occupation-specific and industry-specific experience indicate that match quality in some industry-size bins is sensitive to experience in a few specific jobs, such as computing professionals, legal professionals, writers and creative performing artists, as well as metal and mineral products machine operators, building finishers, and office clerks. The same applies to the geographic location of employees: while typically a worker's municipality appears to play little role in determining match quality, it is quite important to match workers who live in a few specific areas, such as Stockholm, Gothenburg and Malmö.

3.2 Job assignment quality at employee level

To predict the quality of worker-job matches in the main sample, we use the algorithm trained on the learning sample to construct an employee-level measure of job assignment quality (*eJAQ*). This measure equals 1 if the employee's job coincides with the most suitable one, i.e., the job to which the algorithm assigns the highest conditional probability for that worker, and 0 otherwise: formally, if $\hat{J}_i$ is the job predicted for worker i and J_i is the actual job held by that worker, then $eJAQ_i = \mathbf{1}_{\{J_i=\hat{J}_i\}}$. This indicator is the key building block of our measure of job assignment quality at the firm level (*JAQ*), which is simply obtained by averaging *eJAQ* across the employees of the same firm in a given year.

While the *eJAQ* measure has the benefit of simplicity, it has two shortcomings: first, it only captures changes in match quality if workers switch to (or away from) their best possible match, thus neglecting any intermediate change in match quality; second, it does not take into account that workers may feature different suitability to the best possible match, depending on the specialization of their skill set: for instance, the probability that a position as computing professional is the top job for an electronic engineer may be, say, 90%, while a worker with a high-school degree may be well suited to several jobs as machine operator with equal probability of 30%. To overcome both limitations, we also construct a continuous measure of employee-level job match quality (*pJAQ*) by estimating the probability that the algorithm assigns to the actual job held by a worker: formally, $pJAQ = P_Z(J_i|X_i)$ for worker i. This alternative measure is a gauge of a worker's fit for her actual job compared to other jobs that she might perform, and as such it captures the change in match quality associated with any job switch, as well as differences in the degree of specialization across workers. Indeed, the *pJAQ* measure ranges between zero and a worker-specific maximum, $\bar{p}_i$, defined as the highest predicted probability with which the algorithm assigns worker i to any job: formally, $\bar{p}_i = \max_{j \in \text{Jobs}} P_Z(j|X_i)$. Hence, the upper bound of the *pJAQ* measure is greater for workers with more specialized skill sets.[15]

In this section we investigate the validity of both *eJAQ* and *pJAQ* as measures of workers' job assignment quality. First, it is natural to expect that the likelihood of being assigned to a more suitable job increases along workers' careers, as managers learn about employees' characteristics (Fredriksson et al., 2018), and employees themselves adapt their skills via on-the-job training (Guvenen et al., 2020). Second, insofar as an improvement in job allocation generates productivity gains, these are likely to be partly appropriated by workers in the form of higher wages. Hence, one can expect wages to be positively related to *eJAQ*. Third, separations should be less likely for workers that are matched to their most suitable job, as found by Fredriksson et al. (2018).

All these predictions find support in our data. Figure 3 shows the binned scatter plot of *eJAQ* against labor market experience: the likelihood of being assigned to the job predicted by the ML algorithm increases with experience, as the goodness of worker-job matches rises significantly (from 35% to 57%) over the span of a 50-year working life. The largest gain (about 12 percentage points) occurs in the first 5 years of a worker's career: this accords with the intuition that learning is faster for junior workers, and that their reallocation to more suitable jobs is easier than for senior employees (Farber and Gibbons, 1996).

Insert Figure 3 here

Moreover, better matches between workers and jobs are systematically associated with higher compensation, suggesting that assigning workers to the right jobs brings about efficiency gains. This is shown in columns 1-4 of Table 2, where Panel A reports the estimates of the following earnings regression:

$$w_{it} = \alpha_j + \beta eJAQ_{it} + \gamma X_{it} + \delta Z_{f(i,t)} + \lambda_t + u_{it}, \tag{1}$$

where w_{it} is the logarithm of annual earnings of worker i in year t; α_j are job indicators; $eJAQ_{it}$ is a dummy variable that equals 1 if worker i is allocated to her most suitable job in year t, and 0 otherwise; X_{it} are all the workers' characteristics included in the ML algorithm; $Z_{f(i,t)}$ are the characteristics of the firm f that employs worker i in year t (e.g., 2-digit industry dummies, firm age, indicators for family firm, listed company, presence of a human resources manager), and λ_t are year dummies. Panel B shows the estimates of the same specification, simply replacing $eJAQ_{it}$ with $pJAQ_{it}$.

Insert Table 2 here

[15] If one wants a measure that does not capture such difference in specialization across workers, and is merely a continuous counterpart of the *eJAQ* metric, one can normalize the *pJAQ* measure by $\bar{p}_i$. We find that the results obtained when using this continuous metric (which we label *cJAQ*) are quite similar to those based on the *eJAQ* measure, and therefore we report them only as a robustness check in the Web Appendix of this paper.

Column 1 of Panel A reports the estimate of β in a version of equation (1) that includes only job and year effects and the machine learning variables. The resulting estimate is 0.026: a worker allocated to her most suitable job ($eJAQ_{it} = 1$) is estimated to earn 2.6% more than a mismatched worker with the same characteristics or with the same job ($eJAQ_{it} = 0$). The estimate of β remains unchanged upon controlling for 2-digit industry dummies and firm characteristics (column 2), while it decreases slightly upon considering only within-worker variation in $eJAQ_{it}$ (column 3), even controlling for unobserved heterogeneity across firms (column 4): the estimated β in a specification that includes worker, jobs and year effects is 1.9% and highly statistically significant. These findings are in line with the -2% estimate of the coefficient of job mismatch on wage growth regressions reported in column 1 of Table 7 in Fredriksson et al. (2018), despite the differences in methodology and sample used.

We also explore the correlation of *pJAQ* with labor earnings in Panel B of Table 2 to provide a robustness check of the results obtained using the *eJAQ* indicator with a continuous measure of workers' suitability to 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 0.047 coefficient estimate in column 1 indicates that a 10 percentage points increase in a worker's *pJAQ* (amounting to half of its standard deviation) is associated with a 0.47% increase in labor earnings. This effect is qualitatively similar and equally precisely estimated in the specification with industry fixed effects and firm-level controls shown in column 2, and in those with worker fixed effects (column 3) and both worker and firm fixed effects (column 4).[16]

Columns 5 through 8 show the relationship between the likelihood of an employer change from one year to the next and the two measures of job allocation quality, *eJAQ* and *pJAQ*, in Panel A and B respectively. Specifically, we estimate a version of equation (1) where the outcome variable is a separation indicator I_{it+1}, which equals 1 if worker i changes employer between year t and year $t+1$ and 0 otherwise.

Column 5 in Panel A shows that well-matched workers (i.e., those with $eJAQ_{it} = 1$) are 1.2 percentage points less likely to change employer than mismatched workers (i.e., those with $eJAQ_{it} = 0$) with the same characteristics. The coefficient is virtually unchanged upon adding firm controls and industry fixed effects (column 6). In the specification of column 7 we exploit only within-worker variation, namely, we ask how much less likely a given worker is to switch to a new employer when she goes form being mismatched to being well-matched: interestingly, in this case the coefficient increases in absolute value to 2.6 percentage points. In column 8 we also control for unobserved heterogeneity in turnover rates across firms: in this specification, the likelihood

[16] In the Web Appendix we also estimate these regressions replacing the *pJAQ* variable with the *cJAQ* measure, which is the continuous counterpart of the *JAQ* variable: the resulting estimates are very similar to those obtained for the *pJAQ* variable, especially for the specification shown in column 1.

of a separation reverts to being close to 1 percentage point lower for well-matched workers than for mismatched ones. Columns 5 through 8 in Panel B show that a 10 percentage points increase in *pJAQ* is associated with a reduction in the likelihood of changing firm between 1.63 and 0.7 percentage points.

As the quality of worker-job matches is positively and significantly associated with labor earnings, it is worth asking which types of jobs are more often assigned to the wrong workers according to our algorithm, thus forgoing attainable increases in labor earnings. The upper panel of Figure 4 shows the percentages of instances in which workers fail to be allocated to their most suitable job in the main sample, averaging such percentages within each of the following six job classes: 1) managers, 2) professionals, 3) technicians and clerks, 4) skilled manual workers, 5) machine operators and assemblers, and 6) elementary occupations.[17] Thus, for each job class, the corresponding bar in the figure indicates the frequency of cases in which a worker holding a job in that class should have been allocated to a different job according to our algorithm. The graph shows that the frequency of mismatches is quite uniform across job classes, except for a slightly lower value for professionals (40%) and a considerably larger value for elementary occupations (64%): in the remaining classes, mismatches range from 54% for managers and 52% for technicians and clerks to 48% for skilled manual workers and 51% for machine operators and assemblers.

Insert Figure 4 here

The greater frequency of mismatches for elementary occupations may be due to two concomitant reasons: first, these are low-skill jobs and as such they do not require very specific worker profiles, so that job-worker mismatches may arise easily than for other occupations; second, fewer workers hold these jobs, so that fewer observations inform their allocation rule. Indeed, elementary occupations account for a relatively small fraction of jobs in the economy (6%), not dissimilar from that of managers (7%), while the bulk of workers hold jobs in intermediate classes, as shown in the lower panel of Figure 4. Hence, the absolute frequencies of mismatches in the extreme job classes is much lower than in the intermediate ones: the inefficiency arising from the misallocation in the two extreme classes is mitigated by their relatively lower size.

3.3 Job assignment quality at firm level

The next step in the analysis is to average $eJAQ_{it}$ for all the employees i of any firm f in a given year t: we refer to the resulting firm-level measure of job allocation quality as *JAQ*. By the same token, we average the *pJAQ* measure across the employees of each firm, to produce a firm-level

[17] Skilled manual workers comprise service and shop sales workers, skilled agricultural and fishery workers, and craft and related trade workers.

continuous metric of job-workers match quality. As our approach builds on the assumption that firms differ in their ability to assign workers to jobs, we expect to observe heterogeneity in both of these variables across firms.

Insert Figure 5 here

The top panel of Figure 5 shows the kernel density estimate of firm-level *JAQ* for firms in the main sample and in the learning sample. The bottom panel shows the two corresponding densities for *pJAQ*. As expected, the density of both match quality measures in the main sample assigns greater probability mass to lower values than the corresponding density for the learning sample. Moreover, the dispersion in *JAQ* across firms in the main sample exceeds that in the learning sample. This is as expected, for two reasons. First, the learning sample is used to train the ML algorithm at the core of our match quality measures, so that by construction this sample features a better fit between firms' observed choices and the estimated allocation rule. Second, our learning sample is formed by firms in the top productivity decile: insofar as their higher productivity results from fewer mistakes in applying the most efficient allocation rule, they should feature more concentrated *JAQ* than firms in the main sample. In the limit, if there were no noise in the estimation procedure, the learning sample should feature no dispersion in *JAQ* (i.e., we should observe $JAQ = 1$ for all firms), while there should be dispersion in *JAQ* in the main sample, reflecting deviations from the allocation rule estimated on the learning sample.

Instead, *pJAQ* is more concentrated around low values for the main sample than it is around high values for the learning sample. Again, this is for two reasons. First, firms in the main sample may assign workers to less suitable jobs because they deviate more often from the allocation rule estimated on the learning sample. Second, being based on the learning sample, the algorithm tends to predict jobs' conditional probabilities with lower confidence in the main sample, lowering the probability assigned to the most suitable job. Indeed, it turns out that on average the algorithm places a 28% probability on the most suitable job for employees of main-sample firms, against 50% for employees of learning-sample firms.

3.4 How does job assignment quality vary across firms?

The quality of management practices–defined as managers' ability to monitor performance, set targets and incentivize employees–has been shown to be consistently higher in firms facing harsher product market competition, those run by non-family managers, and those with a better educated workforce (Bloom and Van Reenen, 2007, 2010). These correlations have been respectively interpreted as reflecting the selection and incentive effects of competition, the inefficiencies stemming from dynastic succession in control, and the ability of better managed firms to attract more skilled employees.

It is reasonable to expect similar correlations between these characteristics and the measures of job allocation quality presented in Section 3.3: product market competition can be expected to focus managers' attention on matching employees to the most suitable jobs; family management is likely to have a greater tendency to promote family and friends rather than the most deserving candidates; finally, more educated workers may seek jobs in firms where they can expect to be correctly assigned, especially in view of the evidence in Section 3.2 that better matches are associated with higher earnings. The estimates shown in Table 3 are consistent with all three predictions.

Insert Table 3 here

The first four columns of the table present regressions of measures of job allocation quality on the Lerner index of market competition. The dependent variable is *JAQ* in columns 1 and 2 and *pJAQ* in columns 3 and 4. The Lerner index is defined for each firm as 1 – profits/sales, lagged by two years to remove any potential contemporaneous feedback, and averaged across all firms in the same 2-digit industry excluding the firm itself. All specifications include year and industry dummies (where industries are manufacturing, real estate, renting and business activities, and wholesale and retail). The specifications of the even-numbered columns include the following additional controls: the share of employees with a college degree, log employment, log capital, log firm age, indicator for listed firms, years of managerial experience averaged over employees in the firm. In all the specifications, firms operating in more competitive markets turn out to allocate their employees more closely in line with the estimated allocation rule, according to both our measures.

The last four columns of the table present regressions of measures of job allocation quality on a family firm dummy, which is constructed on the basis of family relations among major shareholders (called owners by the tax authorities) and directors.[18] For each owner and director in a firm, we calculate the number of other family members who are directors or owners in the company. A company is a family firm if at least two family members are owners or board members, or at least one owner and one director comes from the same family. The estimates show that family firms feature significantly lower job allocation quality in most specifications: based on the estimates in columns 7 and 8, in family firms the probability that an employee is matched to his/her most suitable job is between 1.4 and 2.9 percentage points lower than in non-family ones.

Finally, the coefficient of the share of employees with a college degree is positive and significantly different from zero in all the specifications where this variable is included among the explanatory variables: a 10-percent increase in this measure of employees' human capital is associated with about a 1 percentage point increase in the probability of a suitable job assignment. This correlation can either be seen as suggesting that better job-worker matching attracts more qualified

[18] An individual's family comprises his parents, grandparents, children, grandchildren, siblings, and partner(s). A partner is the person with whom the individual has a child.

employees or as indicating that managers pay more attention to the job assignment of employees with a college degree, or both.

4 Job Assignment Quality and Firm Performance

This section explores how the heterogeneity in *JAQ* and in *pJAQ* correlates with firm performance, as measured by sales per employee, value added per employee, and operating return on assets (OROA): we wish to determine whether these two measures capture meaningful variation in the quality of workforce allocation, rather than just statistical noise or firm heterogeneity in productivity. Our exercise parallels the approach used by Bloom et al. (2019) to validate their measure of structured management practices, by investigating their correlation with various indicators of firm performance.

4.1 Descriptive evidence

Figure 6 shows that firm-level productivity correlates positively with both *JAQ* and *pJAQ*. The figure shows partial regression plots of value added per employee against these two job-worker match quality variables, conditioning on year effects and 2-digit industry effects. The two top panels refer to main-sample firms, and the bottom two to learning-sample ones. The left-side panels of the figure show how value added per employee correlates with *JAQ*, and the right-side ones how it correlates with *pJAQ*. A positive relationship is evident in the two top graphs, providing preliminary evidence that main-sample firms tend to feature higher productivity insofar as they allocate employees more closely along the rule estimated from the learning sample.

Insert Figure 6 here

Specifically, the positive correlation with *JAQ* indicates that firms where workers are more often allocated to their most suitable job are more productive than others. Moreover, the positive correlation with *pJAQ* suggests that the specialization of a firm's workforce also plays a role. As explained in Section 3.2, a firm's employees may feature high *pJAQ* not only if they are well assigned within the set of jobs they can possibly hold (p_i close to $\bar{p}_i$), but also if they are highly specialized ($\bar{p}_i$ close to 1), in the sense that their characteristics make them highly suitable to a specific job profile.[19]

The two lower panels of Figure 6 instead show that no correlation between productivity and either *JAQ* or *pJAQ* emerges for firms in the learning sample. This is to be expected, as for these

[19]Indeed, we find that firms' productivity is positively and significantly correlated with firm-average $\bar{p}_i$, which can be regarded as a measure of specialization of its workforce, even after controlling for workers' characteristics.

firms variation in measures of match quality should only reflect sampling variability stemming from random deviations from the estimated allocation rule. This can be easily illustrated by considering an extreme example: if firms in the learning set were to adhere perfectly to a common deterministic allocation rule, then *JAQ* would equal 1 for all of them, and would feature no relationship with productivity. To the extent that the variation in *JAQ* detected in the learning sample reflects firms' random deviations from the same allocation rule, one would not expect it to feature a systematic relationship with firm productivity.

4.2 Regression analysis

Table 4 explores further the firm-level correlation between productivity (as well as profitability) and *JAQ*, controlling for other determinants of productivity. All the specifications presented in the table include year dummies and municipality dummies: the first control for aggregate movements in productivity, the second for productivity differentials across locations. The latter may arise not only from location-related technological advantages, 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 4, column 1 reports the OLS estimates of a regression of log sales per employee on *JAQ* including only year dummies. We find a highly significant coefficient of 0.37, implying that a 10-percentage-point increase in *JAQ* is associated with a 3.7% increase in sales per employee. Equivalently, a one-standard-deviation increase in *JAQ* (0.32) is associated with a 11.8% increase in sales per employee. To put this estimate in perspective, Bloom et al. (2019) find that a one-standard-deviation increase in their management score is associated with a 26.2% increase in sales per employee. The difference in magnitude between the two estimates may reflect the fact that *JAQ* focuses on the gains stemming from the efficient allocation of employees, while the score constructed by Bloom et al. (2019) is a broader synthetic indicator of management practices. It may also reflect the fact that our sample excludes the most productive firms in the economy.

Insert Table 4 here

In column 2 of Table 4, 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 1.8%.

These results are not robust to the inclusion of firm fixed effects, possibly because of attenuation bias due to measurement error. However, we control for various possible sources of spurious correlation between productivity and our measures of job allocation quality, namely, omitted variables such as firm characteristics, differences in firms' occupation structures and in workers' quality across firms.

First, the correlation may reflect other firm characteristics such as their size, sector or input mix. However, this is not the case, as shown by the estimates in columns 4 and 5 of the table, which refer to specifications that control for 2-digit industry indicators, log number of employees, log capital, and the fraction of employees with at least a college degree. 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 in a different fashion: 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. To take this concern into account, Panel B of Table 4 reports the estimates of the following specification:

$$y_{ft} = \theta_0 + \theta_1 JAQ_{ft} + \theta_2 F_{jft} + \theta_3 Z_{ft} + \lambda_t + \gamma_h + u_{ft} \qquad (2)$$

where y_{ft} is log(sales/employees), value added per employee or operating return on assets, F_{jft} is the fraction of workers in firm f assigned to job j in year t; Z_{ft} are the characteristics of firm f in year t, namely their age, family firm status, state-owned status, listed firm status, an indicator for the presence of a human resources (HR) manager, its log number of employees and its log of total assets; λ_t are year dummies, and γ_h are 2-digit-industry dummies. In columns 1, 2 and 3 of Panel B this specification is estimated omitting the firm characteristics Z_{ft}, while in columns 4, 5 and 6 these are also included. The results 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.

A third source of spurious correlation is that firms with higher *JAQ* may feature higher-quality workers, irrespective of the job they are allocated to, thus creating a spurious correlation between *JAQ* and productivity. To address this concern, in Panel C of Table 4 we augment specification (2) with the workers' characteristics included in the machine learning algorithm, averaged across all workers employed in firm f in year t. In columns 1, 2 and 3 we control for year effects, occupation structure and workers' characteristics. Columns 3, 4, and 5 also add industry dummies and firm characteristics. The coefficient of *JAQ* remains positive and statistically significant also in these very conservative specifications, even though in some of them it drops further in magnitude.[20]

[20]The results reported in Table 4 are obtained using the main sample. Upon estimating the same specifications with

In almost all of the specifications shown in Table 4 profitability, as measured by operating return on assets, is not significantly correlated with our measure of efficient job allocation, as shown in columns 3 and 6 of the table. A possible interpretation of this finding is that in Swedish firms the productivity gains afforded by better job-worker matches in 2001-10 translated mostly into higher wages, rather than increases in firm profitability.

To check the robustness of these results, in Table 5 we repeat the estimation of the specifications shown in Table 4 upon measuring worker-job match quality by the firm-level average of *pJAQ*. The estimated coefficient of this variable is positive and significantly different from zero in all the specifications of the productivity regressions, but not of the profitability regressions, in line with the results of the previous table. The baseline estimates shown in columns 1 and 2 of Panel A imply that a 10 percentage points increase in firm-level suitability of workers to jobs is associated with a 10 percentage points increase in log sales per employee and a 7 percentage points increase in value added per employee. These results are qualitatively robust to the addition of other controls, even though they drop considerably in size.

Insert Table 5 here

4.3 A circularity issue?

One last concern is that the construction and validation of *JAQ* and *pJAQ* performed up to this point may be vitiated by circularity: as explained in Subsection 3.1, we train the ML algorithm to assign workers to jobs in firms from the top decile of the productivity distribution, and then check whether the measures thus obtained correlate with firms' productivity. The obvious counter to this criticism is that the correlation between these measures and productivity is tested on the main sample, and not on the learning sample used to train the algorithm, and indeed the correlation is present only for the main and not for the learning sample, as shown in Section 4. However, one may still fear that *JAQ* and *pJAQ* correlate positively with productivity for spurious reasons: as the assignment rule is estimated on the decile of the most productive firms (by value added per employee), its estimation error may correlate by construction with firm productivity, thus contaminating the regressions in Table 5.

To address this concern, we re-train our ML algorithm on a random subsample of firms, so as to calibrate the reference rule on the basis of the average firm in our sample, rather that top-productivity firms, and investigate whether the resulting measures of job assignment quality still correlate significantly with productivity across firms. Specifically, we redefine the learning sample used to train our ML algorithm as a 10% random draw of firms (in the same size-industry class)

the learning sample, no robust relationship between *JAQ* and productivity emerges, consistently with what is shown in Figure 6.

from our entire sample. In this way, we effectively calibrate the rule with which firms allocate employees to jobs on the basis of the behavior of the average firm in our sample. We refer to the resulting measure of job allocation quality as JAQ^R and to the corresponding measure of workers' suitability to their actual job as $pJAQ^R$, where the superscript is a mnemonic for "random". Both JAQ^R and $pJAQ^R$ turn out to correlate positively and significantly with productivity, as shown by the two partial regression plots of the log of value added per employee (controlling for year and industry effects) in the top panel of Figure 7. The regressions shown in Table IA2 in the Web Appendix show that these correlations are robust to the inclusion of firm and worker controls.

Insert Figure 7 here

It is worth comparing the graphs shown in the two top panels of Figure 7 with the corresponding graphs of Figure 6, based on the measures of job allocation quality calibrated on top-productivity firms. In the top panel of Figure 7 the positive correlation appears to be present especially for firms in the bottom and middle portion of the productivity distribution, rather than for the entire support of the distribution as in the top panel of Figure 6. As a result, the relationship between the two variables has an inverse-U shape. This is precisely as expected: since now the rule reflects the behavior of the average firm, the firms that adhere most closely to this rule (i.e., those with the highest value of JAQ^R and $pJAQ^R$) cannot hope to achieve more than an average productivity level. Still, adhering more closely to such an allocation rule is associated with productivity improvements for the typical firm in the sample, because it reduces firms' deviations from the estimated rule in allocating their employees. This applies in particular to firms in the lowest part of the support of the distribution, i.e., those that adhere the least to the estimated rule: for such firms, an increase in JAQ^R and $pJAQ^R$ is associated with a steep productivity gain.

One may still object that estimation error may create a mechanical correlation between our measures of job allocation quality and firm productivity: this would be the case if the selection of firms into the learn set based on productivity were to induce a correlation between the estimated job allocation quality and productivity. To address this concern, we perform a placebo test: we replace firms' actual productivity measure (log of value added per employee) with a noise variable, obtained by randomly reshuffling the original variable across firms, and use it in place of the original productivity measure to re-estimate the algorithm. That is, the learn set is now built using the top 10% firms in terms of the noise variable, as explained in Subsection 3.1, and is used to compute again the *JAQ* and *pJAQ* measures. By construction, the new productivity variable used for the placebo test has the same distribution as the original one, but is independent from the rest of the data.

Note that this placebo test leaves intact the relationship between employees' characteristics and task allocations, and only alters the selection of firms into the learn set. If the positive relationship between our measures of job assignment quality and firm productivity is indeed generated by this selection process, one should also expect to find a positive and significant correlation between them and the noise productivity measure in the placebo test for the main sample. If instead there is no mechanical relationship induced by the learn set selection process, no significant relationship should emerge. This is indeed what emerges from the bottom panels of Figure 7, which plot the results of the regression of the placebo productivity measure on *JAQ* (left-hand chart) and on *pJAQ* (right-hand chart). The lack of correlation between these two variables contrasts with the positive and significant correlation obtained in the main estimation strategy and shown in the top panel of Figure 6.

5 Impact of Manager Quality on Job-Worker Matches

The results presented so far are consistent with our ML algorithm capturing a best-practice rule to allocate workers to jobs, whose adoption is correlated with higher firm-level productivity. Why don't all firms in our sample follow such a best-practice rule? As workers' hiring, assignment to jobs and promotions are typically management decisions, it is natural to inquire whether workers' assignment to jobs is systematically related to managerial quality in our data. This immediately begs another question, namely, how to measure managerial quality based on the observed characteristics of managers. In line with the approach of this paper, a synthetic measure of a firm's managerial quality should be the frequency with which they are assigned to their managerial task in the most productive firms. Another, simpler measure of the quality of firm's managers is their average work experience in managerial positions.

Hence, to investigate this issue, for each firm and date we split *JAQ* into its two components, one measuring the quality of rank-and-file employees' assignment to jobs (*R&F-JAQ*) and the other measuring the quality of managers' allocation to their respective jobs (*M-JAQ*). The first is the average *eJAQ* for all workers that hold non-managerial positions in a given firm at a given date, while the latter is the average *eJAQ* for the corresponding firm's managers. Next, we investigate the firm-level relationship between these two variables, as well as *R&F-JAQ* and managerial experience.

Table 6 presents the results of the corresponding regressions, which are based on data from 2003 to 2010: data for 2001 and 2002 are omitted in order to enable the *JAQ* measure to condition on at least two years of experience for all workers. In columns 1 to 3 *M-JAQ* refers both to top managers (CEOs and firm directors) and to middle managers, whereas in columns 4 to 6 they only refer to top managers. Columns 1 and 4 display results from baseline regressions whose dependent variable is the job allocation quality of rank-and-file employees (*R&F-JAQ*), and whose explanatory variable is the allocation quality of managers (*M-JAQ*), including only year effects.

The correlation is positive and significantly different from zero in both regressions: 10 percentage points increase in the quality of managers' allocation is associated with an increase in the quality of rank-and-file workers' allocation ranging between 2 and 1.3 percentage points, depending on the specification. When the quality of managers' allocation refers only to the firm's top management, the coefficient approximately halves in size, indicating that middle management is also important for the correct allocation of workers to their jobs.

Insert Table 6 here

The specifications shown in columns 2 and 5 also include firm fixed effects and the average experience of the firms' managers (Manager exp), and those shown in columns 3 and 6 additionally include industry fixed effects, municipality fixed effects 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). In both of them, managerial experience appears to contribute positively and significantly to *R&F-JAQ*, but the coefficient of *M-JAQ* remains large and precisely estimated. Importantly, these regressions are based only on within-firm variation in the relevant variables: they indicate that a firm's improvements in allocating of rank-and-file employees to jobs tend to occur when the firm improves its management's quality and experience.

These findings beg the question whether managerial quality and expertise, by improving the matching of workers to jobs, contributes to account for the firm-level productivity differentials analyzed in Section 4. Table 7 shows that indeed this is the case: both the log of value added per employee and the log of sales per employee are positively and significantly correlated with the quality of managers' allocation, irrespective of whether the specification only contains year and municipality dummies (columns 1 and 2) or also includes industry dummies. Indeed, the coefficient of the *M-JAQ* variable in column 2 of this table (0.14) is almost as large as that of the *JAQ* variable in the corresponding column (0.18) in panel A of Table 4, suggesting that managerial quality accounts for most of the correlation between worker-job matching and productivity. The specifications in columns 5 and 6 suggest that not only the allocation quality of managers but also their average experience contributes to account for firms' productivity. Hence, the evidence is consistent with the view that managerial quality and experience, via their effects on the matching of rank-and-file workers to jobs, contribute to the observed productivity differentials between firms.

Since it is natural to expect improvements in managerial quality and experience to result from the hiring of better managers and/or the dismissal of incompetent ones, our next step is to test whether the allocation of rank-and-file workers improves upon incumbent managers being replaced with more suitable ones, and worsens upon them being replaced with less suitable ones. To perform this test, the first step is to measure the change in managers' quality associated with their turnover, relative to the counterfactual level of managerial quality associated with no turnover.

Letting τ denote a year in which managerial turnover occurs in a given firm (meaning that at least one of its managers changes), we measure the concomitant change in managerial quality, denoted by $\Delta M\text{-}JAQ_\tau$, as the difference between the average quality of the firm's new management team, $M\text{-}JAQ_\tau$, and the weighted average of the mean quality of retained managers in year τ and the mean quality of dismissed managers in year $\tau-1$. Formally, the change in managerial quality associated with turnover is

$$\Delta M\text{-}JAQ_\tau = M\text{-}JAQ_\tau - \frac{\sum_{i=1}^{N_\tau^r} eJAQ_{i\tau}^r + \sum_{j=1}^{N_{\tau-1}^d} eJAQ_{j\tau-1}^d}{N_\tau^r + N_{\tau-1}^d}, \tag{3}$$

where $eJAQ_{i\tau}^r$ denotes the quality of retained manager i in year τ, $eJAQ_{j\tau-1}^d$ denotes the quality of dismissed manager j in year $\tau-1$, N_τ^r the number of managers retained in year τ and $N_{\tau-1}^d$ the number of managers dismissed in year $\tau-1$. The fractional term in expression (3) measures the counterfactual level of managerial quality in the absence of managerial turnover, which is based on the assumption that the average quality of dismissed managers would have remained the same if they had not been dismissed. Importantly, this measure is designed so as to only track changes in the quality of the managerial team associated with changes in its composition: it disregards the change in the average quality of retained managers between years τ and $\tau-1$, as this change would occur irrespective of managerial turnover. Indeed, $\Delta M\text{-}JAQ_\tau$ is zero by construction if no managers are dismissed ($N_{\tau-1}^d = 0$) and no managers are hired.

We then define a "positive turnover event" to occur for a given firm in year τ if in that year expression (3) turns positive for the first time for that firm, and this rise in managerial quality is persistent over time, i.e., is never subsequently reversed, or more than reversed. Symmetrically, a "negative turnover event" occurs in year τ if in that year expression (3) turns negative, and this drop in managerial quality is persistent over time. This is done to purge the event of interest from the confounding effects of sequences of transitory changes in managerial quality associated with turnover. In our data, 1,267 firms (16.5% of the total) experience positive turnover events, 2,691 (35%) experience negative ones, and the remaining 3,720 (48.5%) experience none.

Our final step is to investigate whether such positive and negative managerial turnover events are associated with significant changes in the allocation quality of rank-and-file workers. To this purpose, we estimate the parameters of the treatment effects of these managerial turnover events on *R&F-JAQ*, exploiting variation in treatment timing. To estimate the dynamic treatment effects of interest, we employ the estimator proposed in Callaway and Sant'Anna (2021). This estimator bypasses the pitfalls related to the interpretation of the TWFE estimators – see for instance de Chaisemartin and D'Haultfœuille (2020), Goodman-Bacon (2021), Borusyak et al. (2021), Athey and Imbens (2022), Sun and Abraham (2021), and Baker et al. (2022). It is particularly well-suited to our setting because it focuses on recovering treatment effect dynamics with variation in the timing of the treatment. Figure 8 show the estimated dynamic treatment effects on rank-and-file workers around managerial turnover events, respectively associated with an increase (left) or a decrease (right) in the *JAQ* of the relevant firm's management.

Insert Figure 8 here

The chart on the left shows that replacement of incumbent managers with better ones tends to occur in the wake of sharp and statistically significant deterioration in the allocation of rank-and-file workers to jobs (by about 5 percentage points on average), and are followed by a significant improvement over the subsequent five years, starting at 5 percentage points at the time of the event, and eventually vanishing. Conversely, the chart on the right indicates that replacement of incumbent managers with worse ones tend to occur in firms starting from a normal level of rank-and-file workers allocation quality, but are associated with a strong, persistent and statistically significant deterioration in the allocation of rank-and-file workers–by over 10 percentage points in the first three years, subsequently reduced by half. Overall, this evidence suggests that persistent changes in managerial quality are an important driver of changes in workers' allocation, and therefore – for better or for worse – of organizational change within firms. To address concerns regarding causality, we repeat the estimation using only the 350 managerial turnover events associated with the death of the incumbent management, and find results that are qualitatively similar, although the effects are imprecisely estimated due to the paucity of observations (see Figure IA1 in the web appendix). On the whole, this evidence rhymes well with the finding in Bender et al. (2018) that firms with better management have workers with higher human capital.

In principle, the organizational changes brought about by new management may consist mainly of reallocating existing employees to different tasks or of changing the composition of the firm's workforce via new hires and/or dismissals. Moreover, the reliance on one or the other of these strategies may differ depending on whether the new managerial team is better or worse than the preexisting one according to our metric. To investigate this point, we partition the impact of managerial turnover events on the allocation quality of rank-and-file employees into a component associated with turnover (hires and fires) and one reflecting the allocation quality of retained employees. For brevity, we refer to the former group as "movers" and to the latter as "stayers", and denote their respective job allocation quality as $R\&F\text{-}JAQ^m$ and $R\&F\text{-}JAQ^s$, where the job allocation quality of "movers" is defined as a residual ($R\&F\text{-}JAQ^m \equiv R\&F\text{-}JAQ - R\&F\text{-}JAQ^s$). hence, up to estimation error, the impact of managerial turnover events on these two components should sum to their effect on the firm's overall job allocation quality *R&F-JAQ*, used as dependent variable in the estimates of Figure 8. For each of these three variables, Table 8 shows the estimates obtained with the Callaway-Sant'Anna method at the event time (i.e., the time-0 parameter) of the treatment effects of the positive and negative managerial turnover events, respectively.

Insert Table 8 here

While the estimates reported in the first column of Table 8 show that positive and negative managerial turnover events respectively trigger a 5-percentage-points rise and a 11-percentage-points drop in the overall allocation quality of rank-and-file workers, the second column shows that for negative turnover events most of this drop (9 percentage points) is accounted by a worse allocation of "stayers" (only 2 percentage points being attributable to new hires and/or dismissals). In contrast, positive managerial events have little impact on the allocation of "stayers", while they are associated with a significant improvement in the the job allocation quality of "movers" (3 percentage points). These effects may not only reflect management-initiated human resources policies but also the change in the attractiveness of firms triggered by new management: for instance, firms taken over by low-quality managers may lose their best employees and fail to attract new ones, and thus be forced to reshuffle the remaining employees into positions for which they are not well qualified; conversely, firms taken over by high-quality managers need not engage in much internal reshuffling but are able to attract new workers from the labor market, possibly replacing worse incumbent employees. Whether or not such interpretation is warranted, negative and positive changes in managerial quality appear to affect job-worker matches via different channels: worse managers mainly increase the misallocation of pre-existing employees, while better managers improve the firm's workforce via employee turnover.

6 Conclusions

This paper proposes a novel measure of job-worker allocation quality (*JAQ*) by combining employer-employee data with machine learning techniques and validate it by exploring its correlation with workers' wages over their careers, firm performance, and with managerial turnover.

Over individual workers' careers, our measure correlates positively with earnings and negatively with separations. At firm level, it increases with competition, non-family firm status, workers' human capital and has a robust correlation with productivity. The quality of rank-and-file workers' job matches responds positively to improvements in management quality.

Our evidence shows that workers earn significantly more as they are better allocated to jobs over their careers, and that workers better matched to their jobs are less likely to switch to a new employer. Job allocation quality is found to vary systematically across firms: companies that operate in more competitive markets, those that are not family-managed and those with a more educated workforce do a better job at matching their employees to jobs. Most importantly, firm productivity correlates robustly with our measure of job-worker allocation quality.

Hence, our measure correlates with key firm characteristics in the same way as management practices do, suggesting that it measures a hitherto unmeasured dimension of management's ability. Indeed, we find that the quality of management plays a key role in the efficient assignment of workers to jobs: rank-and-file workers' allocation improves significantly when managerial turnover leads to better assigned and more experienced managers, while the opposite occurs when turnover leads to lower-quality management.

The measure proposed in this paper can be constructed for any linked employer-employee data that include workers' occupations, without requiring either expensive surveys or detailed expert evaluations of the skills required for each job, and can be applied to explore a vast range of research questions in organizational economics, in labor economics and in corporate finance. Its promise is already witnessed by the worker-level and firm-level evidence provided by this paper.

References

Abadie, Alberto and Maximilian Kasy, "Choosing Among Regularized Estimators in Empirical Economics: The Risk of Machine Learning," *The Review of Economics and Statistics*, 2019, *101* (5), 743–762.

Adenbaum, Jacob, "Endogenous Firm Structure and Worker Specialization," Technical Report, University of Edinburgh 2023.

Athey, Susan, "The Impact of Machine Learning on Economics," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019, pp. 507–547.

_ **and Guido W. Imbens**, "Design-based analysis in Difference-In-Differences settings with staggered adoption," *Journal of Econometrics*, 2022, *226* (1), 62–79. Annals Issue in Honor of Gary Chamberlain.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang, "How Much Should We Trust Staggered Difference-in-Differences Estimates?," *Journal of Financial Economics*, 2022, *144* (2), 370–395.

Baley, Isaac, Ana Figueiredo, and Robert Ulbricht, "Mismatch Cycles," *Journal of Political Economy*, 2022, *130* (11), 2943–2984.

Bandiera, Oriana, Andrea Prat, Stephen Hansen, and Raffaella Sadun, "CEO Behavior and Firm Performance," *Journal of Political Economy*, 2020, *128* (4), 1325–1369.

_ **, Renata Lemos, Andrea Prat, and Raffaella Sadun**, "Managing the Family Firm: Evidence from CEOs at Work," *The Review of Financial Studies*, 2018, *31* (5), 1605–1653.

Bender, Stefan, Nicholas Bloom, David Card, John Van Reenen, and Stefanie Wolter, "Management Practices, Workforce Selection, and Productivity," *Journal of Labor Economics*, 2018, *36* (S1), S371–S409.

Bennedsen, Morten, Francisco Pérez-González, and Daniel Wolfenzon, "Do CEOs Matter? Evidence from Hospitalization Events," *The Journal of Finance*, 2020, *75* (4), 1877–1911.

_ **, Kasper Meisner Nielsen, Francisco Pe Rez-Gonza Lez, and Daniel Wolfenzon**, "Inside the Family Firm: The Role of Families in Succession Decisions and Performance," *Quarterly Journal of Economics*, 2007, *122* (2), 647–691.

Benson, Alan, Danielle Li, and Kelly Shue, "Promotions and the Peter Principle," *The Quarterly Journal of Economics*, 2019, *134* (4), 2085–2134.

Bertrand, M. and A. Schoar, "Managing with Style: The Effect of Managers on Firm Policies," *The Quarterly Journal of Economics*, 2003, *118* (4), 1169–1208.

Bloom, N. and J. Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries," *The Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

Bloom, Nicholas and John Van Reenen, "Why Do Management Practices Differ across Firms and Countries?," *Journal of Economic Perspectives*, 2010, *24* (1), 203–24.

__ , **Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

__ , **Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen**, "What Drives Differences in Management Practices?," *American Economic Review*, 2019, *109* (5), 1648–1683.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting Event Study Designs: Robust and Efficient Estimation," Technical Report, arXiv:2108.12419 2021.

Bowlus, Audra J., "Matching Workers and Jobs: Cyclical Fluctuations in Match Quality," *Journal of Labor Economics*, 1995, *13* (2), 335–350.

Breiman, Leo, "Random Forests," *Machine learning*, 2001, *45* (1), 5–32.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 2006, *74* (2), 323–364.

Callaway, Brantly and Pedro H.C. Sant'Anna, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230. Themed Issue: Treatment Effect 1.

Chen, Chao, Andy Liaw, and Leo Breiman, "Using Random Forest to Learn Imbalanced Data," Technical Report No. 666, University of California, Berkeley 2004.

Chiappori, Pierre-André and Bernard Salanié, "The Econometrics of Matching Models," *Journal of Economic Literature*, 2016, *54* (3), 832–861.

Cornwell, Christopher, Ian M. Schmutte, and Daniela Scur, "Building a Productive Workforce: The Role of Structured Management Practices," *Management Science*, 2021, *67* (12), 7308–7321.

de Chaisemartin, Clément and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Deming, David and Lisa B. Kahn, "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals," *Journal of Labor Economics*, 2018, *36* (S1), S337–S369.

Eeckhout, Jan and Philipp Kircher, "Assortative Matching With Large Firms," *Econometrica*, 2018, *86* (1), 85–132.

Ek, Simon, "Worker Specialization and the Consequences of the Decline in Routine Work," 2022.

Erel, Isil, Léa H Stern, Chenhao Tan, and Michael S Weisbach, "Selecting Directors Using Machine Learning," *The Review of Financial Studies*, 2021, *34* (7), 3226–3264.

Farber, H. S. and R. Gibbons, "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

Fox, Jeremy T. and Valérie Smeets, "Does Input Quality Drive Measured Differences in Firm Productivity?," *International Economic Review*, 2011, *52* (4), 961–989.

Fredriksson, Peter, Lena Hensvik, and Oskar Nordström Skans, "Mismatch of Talent: Evidence on Match Quality, Entry Wages, and Job Mobility," *American Economic Review*, 2018, *108* (11), 3303–3338.

Goodman-Bacon, Andrew, "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 2021, *225* (2), 254–277. Themed Issue: Treatment Effect 1.

Guvenen, Fatih, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer, "Multidimensional Skill Mismatch," *American Economic Journal: Macroeconomics*, 2020, *12* (1), 210–244.

Huitfeldt, Ingrid, Andreas R Kostøl, Jan Nimczik, and Andrea Weber, "Internal Labor Markets: A Worker Flow Approach," *IZA DP No. 14637*, 2021.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *The American Economic Review*, 1997, *87* (3), 291–313.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 1979, *87* (5), 972–990.

Kaplan, Steven N., Mark M. Klebanov, and Morten Sorensen, "Which CEO Characteristics and Abilities Matter?," *The Journal of Finance*, 2012, *67* (3), 973–1007.

Keloharju, Matti, Samuli Knüpfer, and Joacim Tåg, "CEO Health," *The Leadership Quarterly*, 2023, p. 101672.

Lazear, Edward P. and Kathryn L. Shaw, "Wage Structure, Raises and Mobility: An Introduction to International Comparisons of the Structure of Wages Within and Across Firms," in "The Structure of Wages: An International Comparison," University of Chicago Press, 2009, pp. 1–57.

_ , _ , and Christopher T. Stanton, "The Value of Bosses," *Journal of Labor Economics*, 2015, *33* (4), 823–861.

Li, Danielle, Lindsey Raymond, and Peter Bergman, "Hiring as Exploration," Technical Report, NBER Working Paper 27736 2020.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, "Measuring Corporate Culture Using Machine Learning," *The Review of Financial Studies*, 2021, *34* (7), 3265–3315.

Lindenlaub, Ilse, "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 2017.

Lippi, Francesco and Fabiano Schivardi, "Corporate Control and Executive Selection," *Quantitative Economics*, 2014, *5* (2), 417–456.

Lise, Jeremy and Fabien Postel-Vinay, "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 2020, *110* (8), 2328–2376.

Liu, Xu-Ying, Jianxin Wu, and Zhi-Hua Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2008, *39* (2), 539–550.

Malmendier, Ulrike and Geoffrey Tate, "CEO Overconfidence and Corporate Investment," *The Journal of Finance*, 2005, *60* (6), 2661–2700.

_ **and** _ , "Superstar CEOs," *Quarterly Journal of Economics*, 2009, *124* (4), 1593–1638.

Micci-Barreca, Daniele, "A Preprocessing Scheme for High-Cardinality Categorical Attributes in Classification and Prediction Problems," *ACM SIGKDD Explorations Newsletter*, 2001, *3* (1), 27–32.

Minni, Virginia, "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs," Technical Report, London School of Economics 2023.

Moen, Espen R., "Competitive Search Equilibrium," *Journal of Political Economy*, 1997, *105* (2), 385–411.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, 2017, *31* (2), 87–106.

Mullins, William and Antoinette Schoar, "How do CEOs see their roles? Management philosophies and styles in family and non-family firms," *Journal of Financial Economics*, 2016, *119* (1), 24–43.

Ocampo, Sergio, "A Task Based Theory of Occupations," Technical Report, University of Western Ontario, Centre for Human Capital and Productivity (CHCP) Working Papers 2022.

Pastorino, Elena, "Careers in Firms: The Role of Learning and Human Capital," *Journal of Political Economy*, 2023.

Perry, Anja, Simon Wiederhold, and Daniela Ackermann-Piek, "How Can Skill Mismatch be Measured? New Approaches with PIAAC," *Methods, Data, Analyses*, 2016, *8* (2), 137–74.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Robnik-Šikonja, Marko, "Improving Random Forests," in J.-F. Boulicaut et al., eds., *ECML 2004*, Vol. 3201 of *LNAI* Springer 2004, p. 359–370.

_ **and Petr Savicky**, *CORElearn: Classification, Regression and Feature Evaluation* 2020. R package version 1.54.2.

Scur, Daniela, Raffaella Sadun, John Van Reenen, Renata Lemos, and Nicholas Bloom, "The World Management Survey at 18: Lessons and the Way Forward," *Oxford Review of Economic Policy*, 2021, *37* (2), 231–258.

Sokolova, Marina and Guy Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information processing & management*, 2009, *45* (4), 427–437.

Sun, Liyang and Sarah Abraham, "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 2021, *225* (2), 175–199. Themed Issue: Treatment Effect 1.

Terviö, Marko, "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, 2008, *98* (3), 642–668.

Tåg, Joacim, "Production Hierarchies in Sweden," *Economics Letters*, 2013, *121* (2), 210–213.

_ , Thomas Åstebro, and Peter Thompson, "Hierarchies and Entrepreneurship," *European Economic Review*, 2016, *89*, 129–147.

Varian, Hal R., "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 2014, *28* (2), 3–28.

Wallace, Byron C and Issa J Dahabreh, "Class Probability Estimates Are Unreliable for Imbalanced Data (and How to Fix Them)," in "2012 IEEE 12th international conference on data mining" IEEE 2012, pp. 695–704.

Zhang, Chongsheng, Changchang Liu, Xiangliang Zhang, and George Almpanidis, "An Up-to-Date Comparison of State-of-the-Art Classification Algorithms," *Expert Systems with Applications*, 2017, *82*, 128–150.

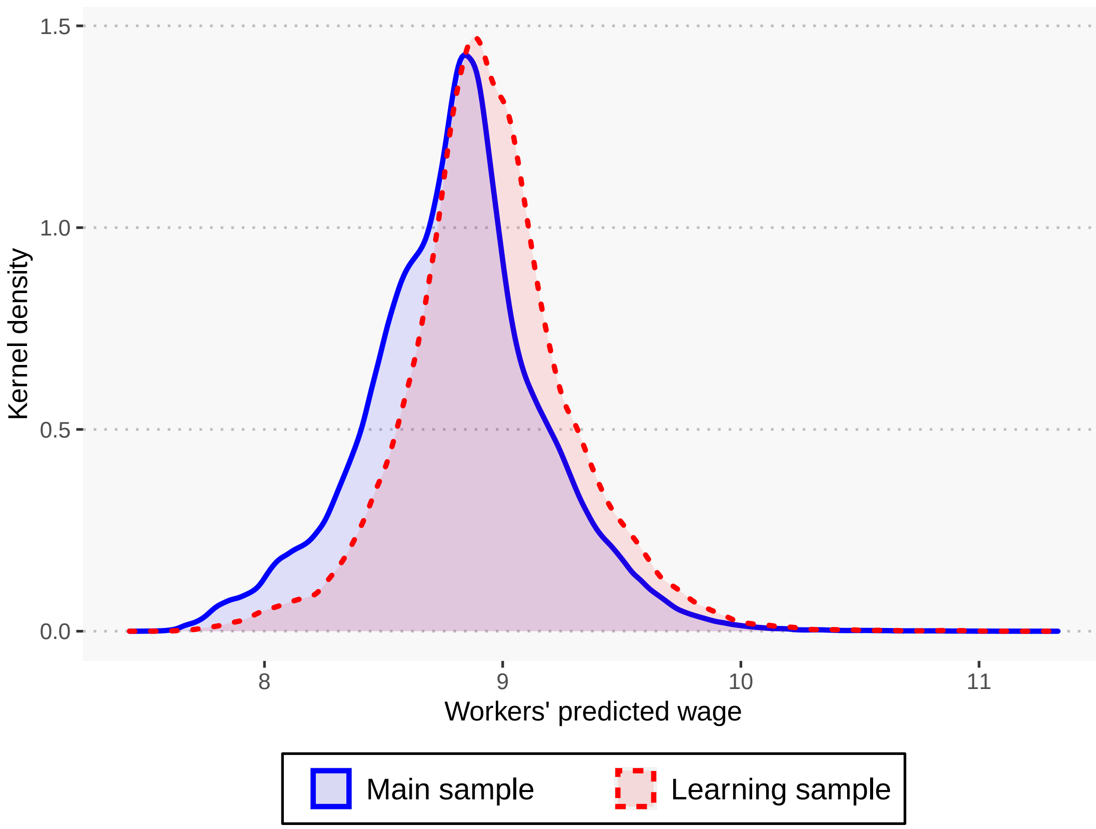


Figure 1: Common support of worker characteristics in the main and the learning samples
This figure shows the distributions of the predicted wages for workers in the learning sample (red line) and the main sample (blue line). For both samples, the predictions are obtained from wage regressions estimated on the main sample using as explanatory variables the worker characteristics included in the ML algorithm. These are 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. The figure shows that the support of the two distributions almost perfectly overlaps.

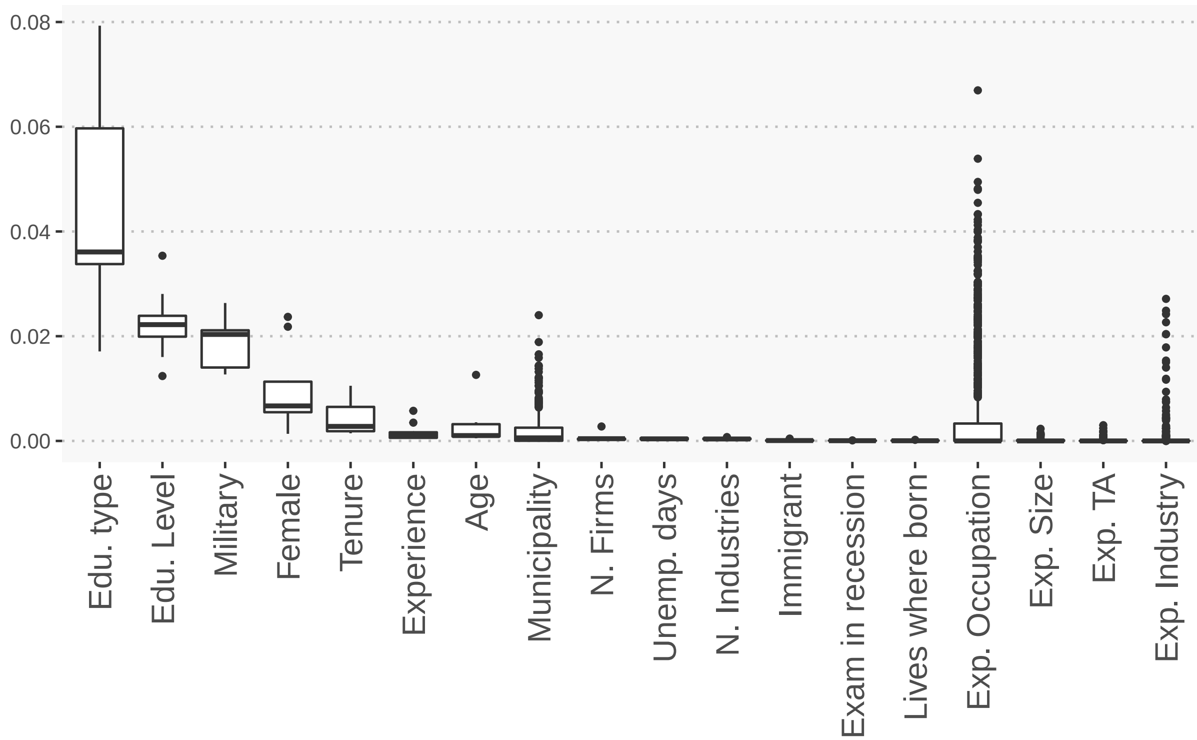


Figure 2: Importance of workers' features in the random forest algorithm, by size-industry bins
The graphs plots the maximum explanatory power of all the workers' features used in the random forest algorithm. Features are listed on the horizontal axis, and the importance of each feature—defined as in Robnik-Šikonja (2004) and Robnik-Šikonja and Savicky (2020)—measures its discriminatory power in the correct classification of the instances. Some features are aggregated under a single label: "Exp. Occupation" aggregates the years of experience in each occupation, "Exp. Industry" those in each industry, "Exp. TA" those in firms with given total assets, and "Exp. Size" those in firms with given number of employees. "Edu. Type" aggregates features related to education specialization. "Tenure" is the years of employment in the current firm, "Municipality" codes the worker's residence, "Female" and "Experience" are the worker's gender and years of experience, "N. Industries" and "N. Firms" the number of industries and firms where a worker was employed, "Unemp. days" the number of unemployment days. "Military", "Immigrant" and "Lives where born" are dummy variables indicating whether the worker performed military service, is an immigrant and lives in his/her birthplace, respectively.

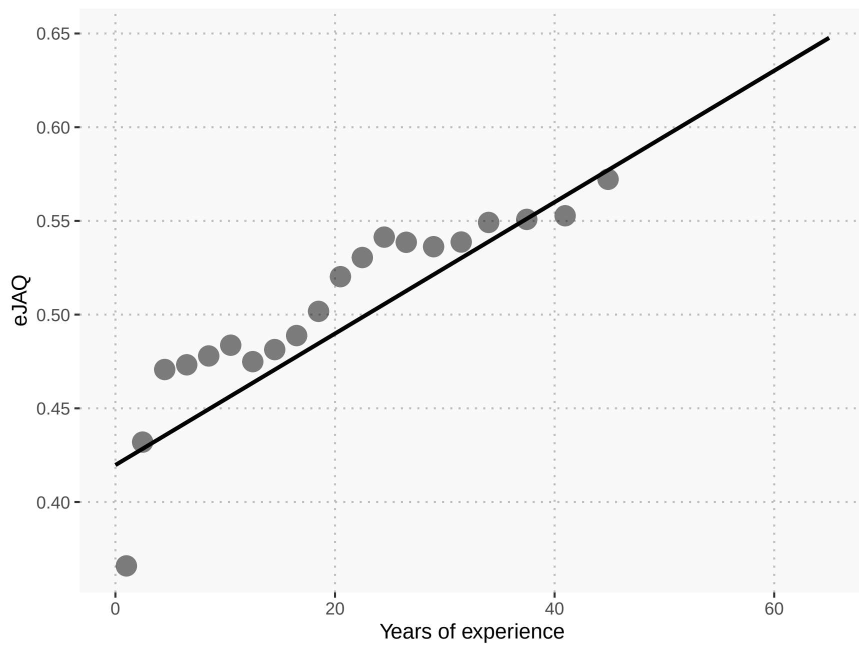


Figure 3: Worker-level job allocation quality (*eJAQ*) by labor market experience
This figure shows the binned scatter plot of an indicator for being assigned to the job predicted by the ML algorithm (*eJAQ*) against years of labor market experience.

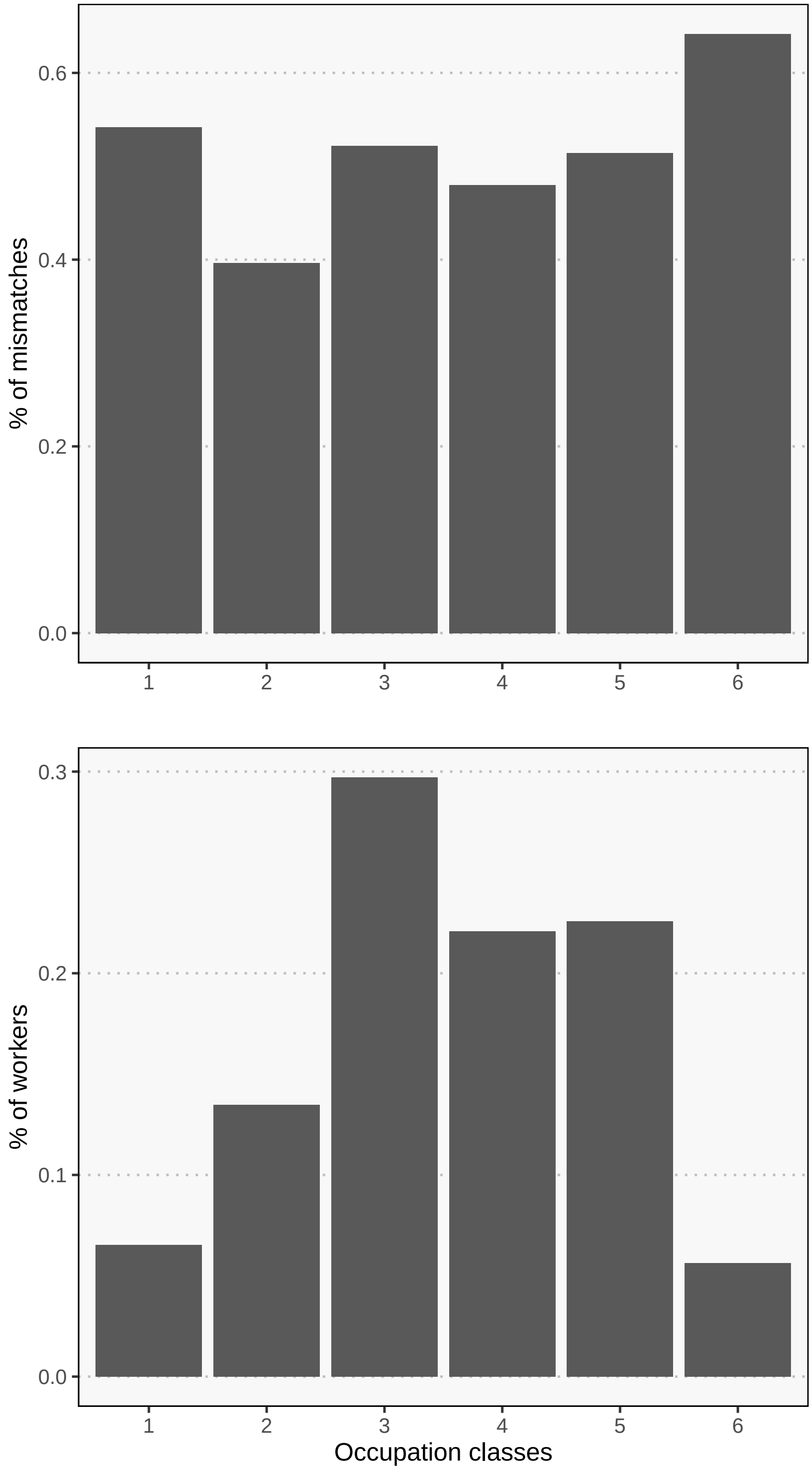
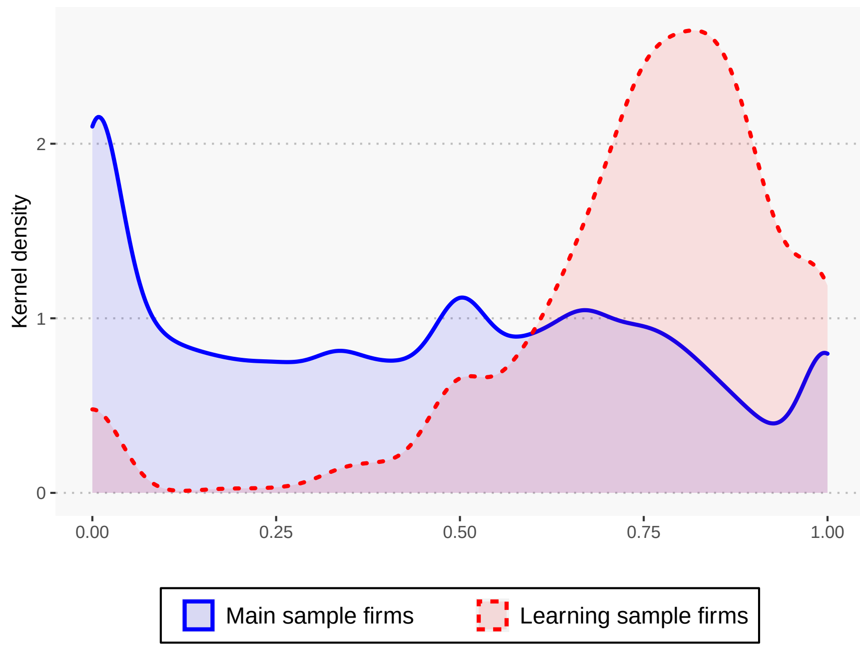


Figure 4: Distribution of mismatches and workers by occupation classes
The top graph shows the percentage of mismatches in each occupation class in the main sample. A mismatch occurs when an employee's observed job differs from the job predicted by the estimated allocation rule. The bottom graph shows the percentage of workers by occupation classes in the main sample. Occupation classes are defined as follows: 1) managers, 2) professionals, 3) technicians and clerks, 4) skilled manual workers, 5) machine operators and assemblers, and 6) elementary occupations.

(a) *eJAQ*



(b) *pJAQ*

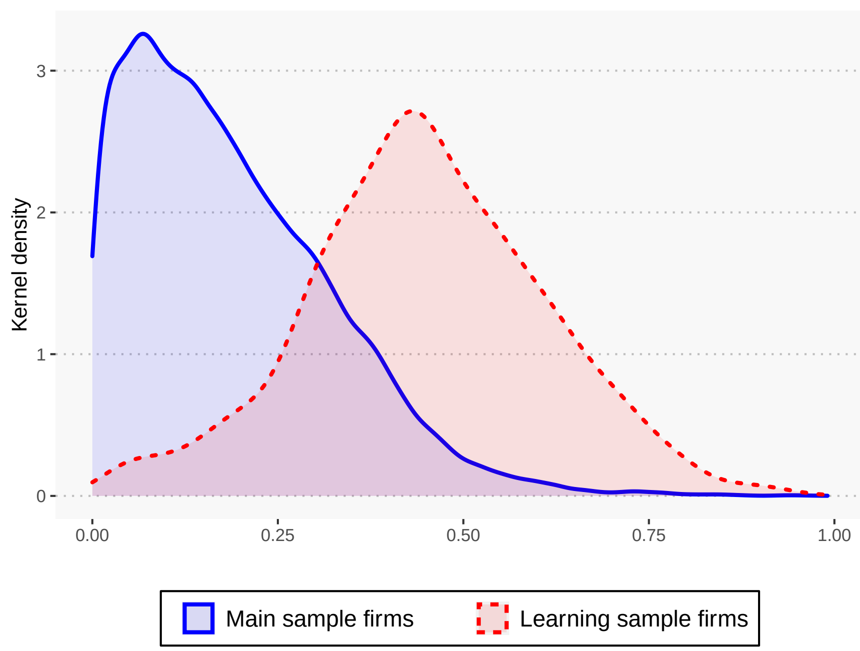
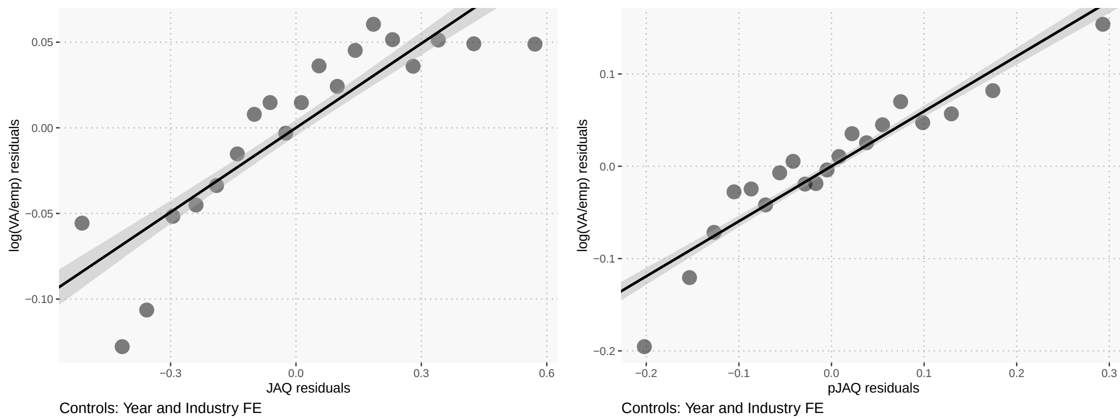


Figure 5: The distribution of ***JAQ***
The upper panel of this figure shows the kernel density estimate of *JAQ* for firms in the main sample (blue line) and in the learning sample (red line). The lower panel presents the corresponding kernel density estimates for *pJAQ*.

(a) Main sample



(b) Learning sample

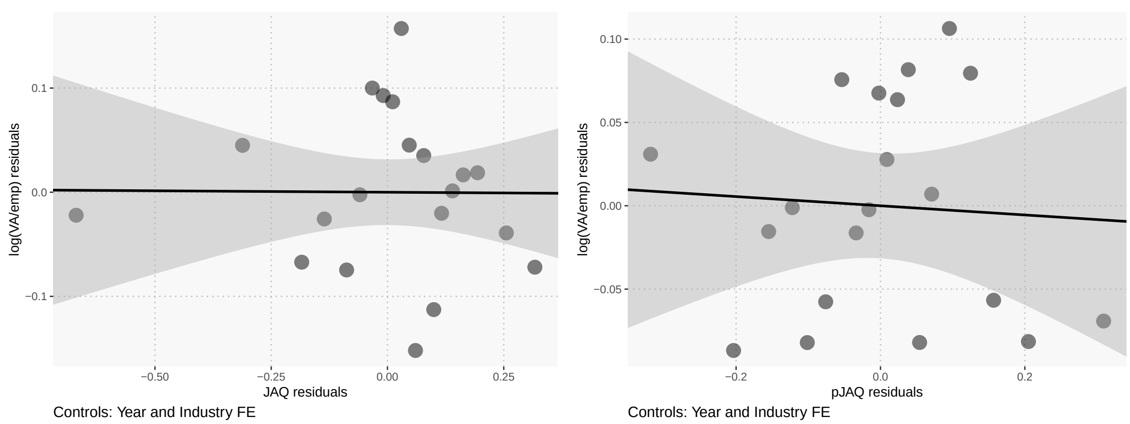
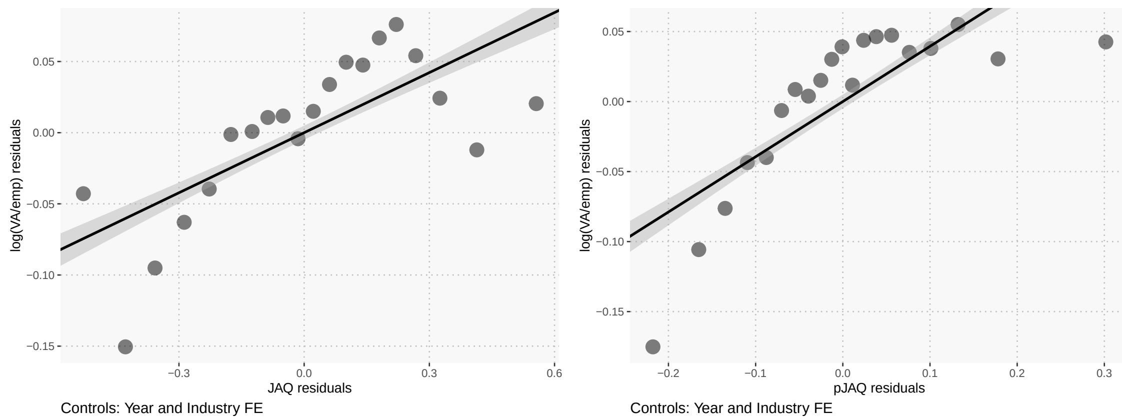


Figure 6: Correlation between productivity and job allocation quality
The figure shows binned scatter plots of productivity, as measured by log value added per employee, against job allocation quality, as measured by *JAQ* in the left charts and by *pJAQ* in the right charts, in each case controlling for year and industry fixed effects. The two top charts refer to the main sample and the two bottom ones to the learning sample. The regression lines are shown together with the respective 95% confidence intervals (shaded area). The points shown in the graphs represent the residuals of the partial regression plots and are computed as follows: residuals are first split into 20 equal-sized bins on the horizontal axis; points in a bin are represented with a unique point, with coordinates given by the average of the coordinates of the points in that bin. The regression line shown fits the residuals and not the binned points.

(a) Random firms set



(b) Placebo test

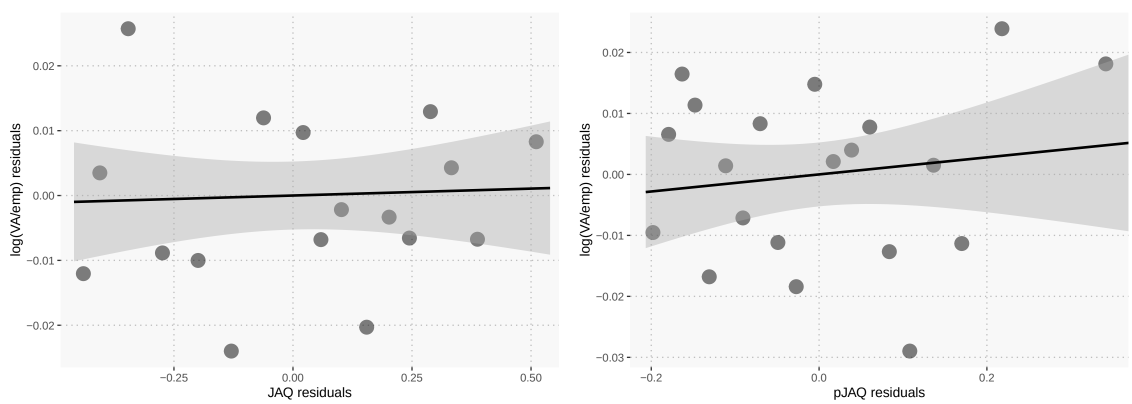


Figure 7: Addressing the circularity issue
Panel (a) in this figure shows partial regression plots of productivity, as measured by log value added per employee, against job allocation quality, as measured by JAQ^R in the left-hand chart and by $pJAQ^R$ in the right-hand chart. JAQ^R and $pJAQ^R$ are measures of job allocation quality and workers' suitability to their actual job respectively obtained by estimating a ML algorithm on a 10% randomly chosen sample of firms, controlling for year and industry fixed effects. Panel (b) shows partial regression plots of placebo productivity on *JAQ* in the left-hand chart and *pJAQ* on the right-hand chart, controlling for a constant. Placebo productivity is a noise variable obtained by randomly reshuffling the original log value added per employee across firms. The regression lines are shown together with the respective 95% confidence intervals (shaded area). The points shown in the two graphs represent the residuals of the partial regression plots and are computed as follows: residuals are first split into 20 equal-sized bins on the horizontal axis; points in a bin are represented with a unique point, with coordinates determined averaging the coordinates of the points in that bin. The regression line shown is fitted on the residuals and not on binned points.

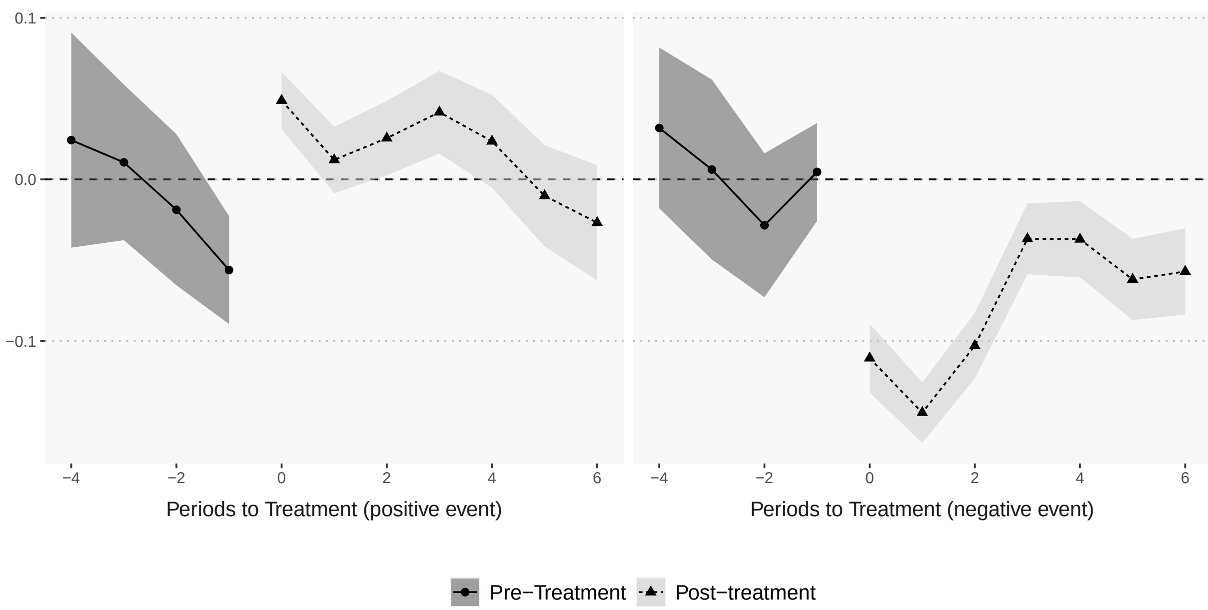


Figure 8: Response of rank-and-file workers' *JAQ* to positive (left) or negative (right) managerial turnover events
The figure shows the behavior of the *JAQ* of rank-and-file workers around managerial turnover events, respectively associated with a persistent increase in the *JAQ* of the relevant firm's management (left panel) and with a persistent decrease in the *JAQ* of the relevant firm's management (right panel). The event study coefficients are estimated using the method by Callaway and Sant'Anna (2021).

Table 1: Descriptive statistics

This table reports the summary statistics of the individuals included in the main sample and in the learning sample. Our total sample includes firms active at some point between 2001 and 2010, reporting a yearly median number of employees between 30 and 6000, and positive total assets and sales. Since information about a worker's specific occupation is not always available, we restrict the sample to firms with at least 10 workers for whom we do observe the current occupation. The main sample contains 5,901,551 observations at the individual level and the learning sample 66,684 observations.

	Mean	P50	P10	P25	P75	P90	SD
Panel A: Main sample							
Labor income (TSEK 2019)	351.991	324.377	168.806	258.118	409.305	541.715	197.704
University degree	0.13	0.00	0.00	0.00	0.00	1.00	0.34
Age	40.64	40.00	25.00	31.00	50.00	58.00	11.96
Female	0.34	0.00	0.00	0.00	1.00	1.00	0.47
Immigrant	0.13	0.00	0.00	0.00	0.00	1.00	0.34
Mobility (lives where born)	0.65	1.00	0.00	0.00	1.00	1.00	0.48
Labor market experience	19.66	19.00	3.00	9.00	29.00	39.00	12.84
Tenure	5.29	4.00	0.00	1.00	8.00	13.00	5.14
# industries worked in	2.28	2.00	1.00	1.00	3.00	4.00	1.26
# jobs held	2.29	3.00	1.00	2.00	4.00	5.00	1.69
# unemployment days since '92	181.90	0.00	0.00	0.00	224.00	599.00	338.90
Panel B: Learning sample							
Labor income (TSEK 2019)	474.627	415.484	270.307	333.533	535.107	724.454	301.928
University degree	0.22	0.00	0.00	0.00	0.00	1.00	0.41
Age	43.24	43.00	29.00	35.00	52.00	58.00	10.88
Female	0.31	0.00	0.00	0.00	1.00	1.00	0.46
Immigrant	0.14	0.00	0.00	0.00	0.00	1.00	0.35
Mobility (lives where born)	0.62	1.00	0.00	0.00	1.00	1.00	0.49
Labor market experience	21.36	21.00	5.00	11.00	31.00	39.00	12.28
Tenure	7.81	6.00	0.00	2.00	12.00	20.00	6.54
# industries worked in	2.71	3.00	1.00	1.00	4.00	5.00	1.49
# jobs held	3.29	3.00	1.00	2.00	4.00	6.00	1.92
# unemployment days since '92	164.37	0.00	0.00	0.00	192.00	530.00	321.09

Table 2: Labor earnings, separations and match quality

This table displays the relationship between the log of labor earnings and match quality in columns (1) to (4), and between a separation indicator (which equals 1 if a worker changes employer between time $t-1$ and t and 0 otherwise) and match quality in columns (5) to (8). Match quality is measured by *eJAQ* in Panel A and by *pJAQ* in Panel B for workers in the main sample. The worker controls (used in the ML algorithm) are 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. The firm controls are firm age, size (measured by the number of employees), sales, and total assets, as well as ownership categories measured by indicators for the firm being a state-owned firm, a listed firm, or a family firm. Standard errors clustered at worker level are shown in parentheses.

	Log(labor earnings)				Separation indicator			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
eJAQ	0.026	0.026	0.019	0.020	-0.012	-0.011	-0.026	-0.009
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.002)
Panel B								
pJAQ	0.047	0.054	0.043	0.053	-0.071	-0.070	-0.163	-0.083
	(0.001)	(0.001)	(0.001)	(0.007)	(0.001)	(0.001)	(0.002)	(0.012)
Year and job FE	✓	✓	✓	✓	✓	✓	✓	✓
Worker controls	✓	✓			✓	✓		
Industry FE		✓				✓		
Firm controls		✓				✓		
Worker FE			✓	✓			✓	✓
Firm FE				✓				✓
Observations	5,901,551	5,901,551	5,901,551	5,526,718	4,484,975	4,481,150	4,484,975	4,262,039

Table 3: Job allocation quality, market competition and firm ownership

This table shows regressions of measures of job allocation quality (*JAQ*) and of workers' suitability to their actual job (*pJAQ*) on the Lerner index of market competition in columns 1 to 4, and on a family firm status dummy in columns 5 to 8. The Lerner index for each firm is defined as $1 -$ profits/sales lagged by 2 years and averaged across all firms in the same 2-digit industry, excluding the firm itself. All specifications include year and industry dummies (where industries are manufacturing, real estate, renting and business activities, and wholesale and retail). The specifications of the even-numbered columns control for the share of employees with a college degree, and include the following additional controls (whose coefficients are not shown for brevity): log employment, log capital, log firm age, indicator for listed firms, years of managerial experience averaged over employees in the firm.

	JAQ		*pJAQ*		*JAQ*		*pJAQ*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lerner index	0.003	0.003	0.003	0.002				
(2-year lagged)	(0.000)	(0.000)	(0.000)	(0.000)				
Family firm					-0.021	-0.002	-0.029	-0.014
					(0.007)	(0.007)	(0.003)	(0.003)
Share emp. w/ college		0.057		0.108		0.108		0.112
		(0.019)		(0.010)		(0.016)		(0.009)
Year dummies	✓	✓	✓	✓	✓	✓	✓	✓
Industry dummies		✓		✓		✓		✓
Firm controls		✓		✓		✓		✓
Observations	33,254	33,254	33,254	33,254	48,116	47,350	48,116	47,350
No. Firms	6,269	6,269	6,269	6,269	7,875	7,763	7,875	7,763
y Mean	0.507	0.507	0.222	0.222	0.433	0.434	0.188	0.188
y St. Dev.	0.300	0.300	0.136	0.136	0.320	0.319	0.137	0.137

Table 4: Firm performance and *JAQ*

This table displays results from regressions on the association between productivity and *JAQ*. Panel A refers to our baseline specification. The results in Panel B 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 (HR) manager, its log number of employees and its log of total assets). Panel C adds controls for worker characteristics (listed in the notes to Table 2). Standard errors clustered at firm level are shown in parentheses.

	Log(sales/emp) (1)	Log(VA/emp) (2)	OROA (3)	Log(sales/emp) (4)	Log(VA/emp) (5)	OROA (6)
Panel A						
JAQ	0.374	0.180	-0.008	0.095	0.072	0.003
	(0.022)	(0.014)	(0.005)	(0.013)	(0.010)	(0.005)
log(cap/emp)				0.414	0.237	-0.020
				(0.012)	(0.009)	(0.002)
log(emp)				0.003	-0.004	-0.003
				(0.007)	(0.005)	(0.002)
Share emp w/ college				0.110	0.338	0.013
				(0.031)	(0.022)	(0.010)
Industry dummies				✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Panel B						
JAQ	0.165	0.113	0.002	0.078	0.052	0.006
	(0.017)	(0.011)	(0.005)	(0.013)	(0.010)	(0.005)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Firm controls				✓	✓	✓
Panel C						
JAQ	0.085	0.057	0.002	0.046	0.032	0.004
	(0.017)	(0.012)	(0.005)	(0.014)	(0.011)	(0.005)
Industry dummies				✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Workers X	✓	✓	✓	✓	✓	✓
Firm Z				✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Observations	48,116	47,743	48,116	48,116	47,743	48,116
No. firms	7,875	7,827	7,875	7,875	7,827	7,875
LHS mean	7.306	6.140	0.079	7.306	6.140	0.079
LHS SD	0.793	0.534	0.179	0.793	0.534	0.179

Table 5: Firm performance and worker suitability

This table displays results from regressions on the relationship between productivity and firm-level *pJAQ*, defined as the average of worker-level *pJAQ* for a given firm. Panel A refers to our baseline specification. The results in Panel B 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 (HR) manager, its log number of employees and its log of total assets). Panel C adds controls for worker characteristics (listed in the notes to Table 2). Standard errors clustered at firm level are shown in parentheses and three stars denote statistical significance at the one percent level.

	Log(sales/emp)	Log(VA/emp)	OROA	Log(sales/emp)	Log(VA/emp)	OROA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
pJAQ	0.975	0.698	-0.007	0.277	0.248	0.020
	(0.059)	(0.034)	(0.013)	(0.037)	(0.026)	(0.013)
log(cap/emp)				0.412	0.235	-0.020
				(0.012)	(0.009)	(0.002)
log(emp)				0.004	-0.003	-0.003
				(0.007)	(0.005)	(0.002)
Share emp w/ college				0.092	0.321	0.012
				(0.031)	(0.022)	(0.010)
Industry dummies				✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Panel B						
pJAQ	0.510	0.361	0.017	0.239	0.178	0.030
	(0.046)	(0.030)	(0.013)	(0.037)	(0.027)	(0.013)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Size-industry FE	✓	✓	✓	✓	✓	✓
Firm controls				✓	✓	✓
Panel C						
pJAQ	0.237	0.176	0.024	0.158	0.127	0.030
	(0.051)	(0.034)	(0.016)	(0.043)	(0.031)	(0.016)
Industry dummies				✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Workers X	✓	✓	✓	✓	✓	✓
Firm Z				✓	✓	✓
Size-industry FE	✓	✓	✓	✓	✓	✓
Observations	48,116	47,743	48,116	48,116	47,743	48,116
No. firms	7,875	7,827	7,875	7,875	7,827	7,875
LHS mean	7.306	6.140	0.079	7.306	6.140	0.079
LHS SD	0.793	0.534	0.179	0.793	0.534	0.179

Table 6: Role of management in the allocation quality of rank-and-file employees

This table displays results from regressions whose dependent variable is the job allocation quality of rank-and-file employees (*R&F-JAQ*) and whose explanatory variables are the allocation quality of managers (*M-JAQ*) and their experience in managerial jobs (Manager exp). In columns 1 to 3 Manager-*JAQ* refers both to top managers (CEOs and firm directors) and to middle managers, whereas in columns 4 to 6 they only refer to top managers. The regressions are based on data from 2003 to 2010. All specifications include year fixed effects; those in columns 2, 3, 5 and 6 include firm fixed effects, and those in columns 3 and 6 include industry fixed effects, municipality fixed effects and firm controls (age, family firm status, state-owned status, listed firm status, an indicator for the presence of a human resources (HR) manager, its log number of employees and its log of total assets). Standard errors clustered at firm level are shown in parentheses.

	R&F-JAQ					
	(1)	(2)	(3)	(4)	(5)	(6)
M-JAQ	0.201	0.127	0.127	0.121	0.068	0.065
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Manager exp		0.018	0.017		0.008	0.008
		(0.002)	(0.002)		(0.002)	(0.002)
Industry FEs			✓			✓
Municipality FEs			✓			✓
Year FEs	✓	✓	✓	✓	✓	✓
Firm FEs		✓	✓		✓	✓
Firm controls			✓			✓
Observations	36,230	36,230	36,206	22,821	22,821	22,807
No. Firms	7,680	7,680	7,679	6,454	6,454	6,452

Table 7: Role of management in firm productivity

This table displays the estimated relationship between productivity and the quality of managers' allocation. Productivity is measured either as log sales per employee or log value added per employee. The regressions are based on data from 2003 to 2010. All specifications include municipality and year fixed effects; those in columns 3 to 6 include industry fixed effects. The specifications in columns 5 and 6 also control for experience in managerial jobs. Standard errors clustered at firm level are shown in parentheses.

	Log(Sales/emp) (1)	Log(VA/emp) (2)	Log(Sales/emp) (3)	Log(VA/emp) (4)	Log(Sales/emp) (5)	Log(VA/emp) (6)
M-JAQ	0.208	0.140	0.153	0.085	0.103	0.066
	(0.018)	(0.014)	(0.014)	(0.012)	(0.016)	(0.012)
Managers exp					0.030	0.012
					(0.004)	(0.004)
Industry FEs	No	No	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Municipality FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,971	35,823	35,971	35,823	35,971	35,823
No. Firms	7,592	7,559	7,592	7,559	7,592	7,559
y Mean	7.408	6.163	7.408	6.163	7.408	6.163
y St. Dev.	0.779	0.577	0.779	0.577	0.779	0.577

Table 8: Decomposition of managerial turnover impact on stayers

The table shows the impact of managerial turnover events on allocations of employees staying in the firm. The coefficients measure the estimated average treatment effect on treated at event time (standard errors in parentheses) obtained using the Callaway-Sant'Anna method, for positive and negative managerial turnover events, respectively. The first column shows the estimates obtained when the dependent variable is *R&F-JAQ*, i.e., the average allocation quality of rank-and-file employees. The second column shows the estimates obtained when the dependent variable is $R\&F\text{-}JAQ^{s}$, i.e., the fraction of correctly allocated employees among those retained by the firm when the event occurs ("stayers"). The third column shows the the estimates obtained when the dependent variable is $R\&F\text{-}JAQ^{m}$, i.e., the fraction of correctly allocated employees due to turnover of rank-and-files employees ("movers").

	R&F-JAQ	$R\&F\text{-}JAQ^{s}$	$R\&F\text{-}JAQ^{m}$
Positive event	0.0488 (0.0091)	0.0166 (0.0141)	0.0301 (0.0138)
Negative event	−0.1106 (0.0109)	−0.0934 (0.0135)	−0.0199 (0.0092)

“For Online Publication”

Internet Appendix for

JAQ of All Trades: Mismatch and Firm Productivity

Luca Coraggio

University of Naples Federico II

Marco Pagano

University of Naples Federico II

Annalisa Scognamiglio

University of Naples Federico II

Joacim Tåg

Hanken School of Economics and IFN

April 2023

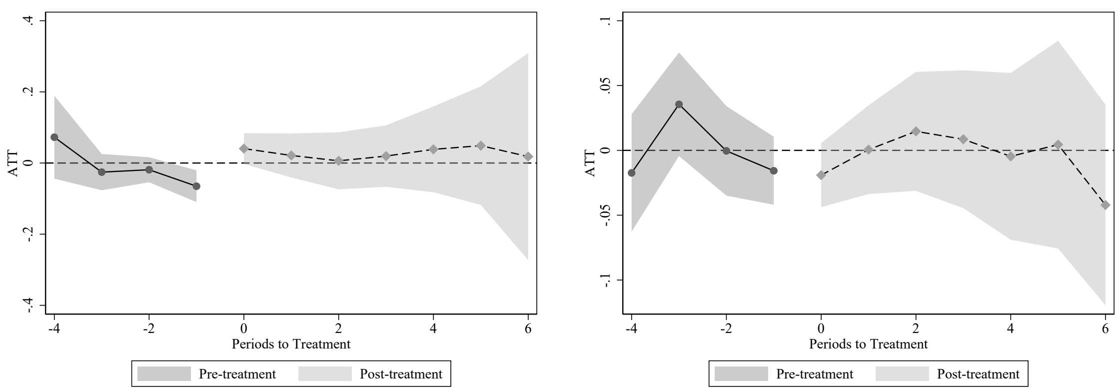


Figure IA1: Response of rank-and-file workers' *JAQ* to positive (left) or negative (right) managers' death events
The figure shows the event study estimated with the method by Callaway and Sant'Anna (2021) relating the JAQ of rank-and-file workers with managerial turnover events associated with the death of (at least) one of the members of the incumbent management. The left panel refers to positive events, i.e., events associated with an increase in managerial JAQ, whereas the right panel refers to negative events, i.e., those associated with a decrease in managerial JAQ.

Table IA1: Firm performance and *cJAQ*

The table reports the estimated relationship between productivity and *cJAQ*. Panel A refers to our baseline specification. The results in Panel B 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 (HR) manager, its log number of employees and its log of total assets). Panel C adds controls for worker characteristics (listed in the notes to Table 2). Standard errors clustered at firm level are shown in parentheses.

	Log(sales/emp)	Log(VA/emp)	OROA	Log(sales/emp)	Log(VA/emp)	OROA
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
cJAQ	0.578	0.253	-0.011	0.156	0.085	0.003
	(0.027)	(0.015)	(0.005)	(0.016)	(0.012)	(0.006)
log(cap/emp)				0.410	0.236	-0.020
				(0.012)	(0.009)	(0.002)
log(emp)				0.002	-0.004	-0.003
				(0.007)	(0.005)	(0.002)
Share emp w/ college				0.095	0.332	0.013
				(0.031)	(0.022)	(0.011)
Industry dummies				✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Panel B						
cJAQ	0.246	0.147	0.003	0.130	0.067	0.007
	(0.021)	(0.014)	(0.005)	(0.016)	(0.012)	(0.006)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Firm controls				✓	✓	✓
Panel C						
cJAQ	0.159	0.088	0.001	0.086	0.042	0.006
	(0.021)	(0.015)	(0.006)	(0.017)	(0.013)	(0.006)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Workers X	✓	✓	✓	✓	✓	✓
Firm Z				✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Observations	48,116	47,743	48,116	48,116	47,743	48,116
No. Firms	7,875	7,827	7,875	7,875	7,827	7,875
y Mean	7.306	6.140	0.079	7.306	6.140	0.079
y St. Dev.	0.793	0.534	0.179	0.793	0.534	0.179

Table IA2: JAQ^R **and productivity**

The table reports the estimated relationship between productivity and the three firm-level measures of job allocation quality, when the ML algorithm is estimated on a random subsample of firms: JAQ^R (panel A), $pJAQ^R$ (panel B) and $cJAQ^R$ (panel C). The panels replicate panel C from Table 4, Table 5, and Table IA1, respectively. The regressions include controls for firms' occupation structure (the fraction of workers in firm f assigned to job j in year t), firm controls (age, family firm status, state-owned status, listed firm status, an indicator for the presence of a human resources (HR) manager, its log number of employees and its log of total assets), and worker characteristics (listed in the notes to Table 2). Standard errors clustered at firm level are shown in parentheses.

	Log(sales/emp) (1)	Log(VA/emp) (2)	OROA (3)	Log(sales/emp) (4)	Log(VA/emp) (5)	OROA (6)
Panel A						
JAQ^R	0.077	0.044	0.004	0.058	0.035	0.006
	(0.017)	(0.013)	(0.005)	(0.013)	(0.011)	(0.005)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Workers X	✓	✓	✓	✓	✓	✓
Firm Z				✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Panel B						
$pJAQ^R$	0.153	0.083	0.023	0.154	0.088	0.028
	(0.049)	(0.036)	(0.018)	(0.039)	(0.031)	(0.018)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Workers X	✓	✓	✓	✓	✓	✓
Firm Z				✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Panel C						
$cJAQ^R$	0.115	0.060	0.004	0.073	0.037	0.008
	(0.022)	(0.016)	(0.006)	(0.017)	(0.014)	(0.006)
Industry dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓
Municipality dummies	✓	✓	✓	✓	✓	✓
Occupations	✓	✓	✓	✓	✓	✓
Workers X	✓	✓	✓	✓	✓	✓
Firm Z				✓	✓	✓
Size-industry bin dummies	✓	✓	✓	✓	✓	✓
Observations	48,116	47,743	48,116	48,116	47,743	48,116
No. Firms	7,875	7,827	7,875	7,875	7,827	7,875
y Mean	7.306	6.140	0.079	7.306	6.140	0.079
y St. Dev.	0.793	0.534	0.179	0.793	0.534	0.179