DISCUSSION PAPER SERIES

DP17451
(v. 2022-12-14 15:27:07)

ROBOT ADOPTION, WORKER-FIRM SORTING AND WAGE INEQUALITY: EVIDENCE FROM ADMINISTRATIVE PANEL DATA

Ester Faia, Gianmarco Ottaviano and Saverio Spinella

LABOUR ECONOMICS AND MACROECONOMICS AND GROWTH

CEPR

ISSN 0265-8003

ROBOT ADOPTION, WORKER-FIRM SORTING AND WAGE INEQUALITY: EVIDENCE FROM ADMINISTRATIVE PANEL DATA

Ester Faia, Gianmarco Ottaviano and Saverio Spinella

Discussion Paper DP17451
First Published N/A
This Revision 08 July 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

ROBOT ADOPTION, WORKER-FIRM SORTING AND WAGE INEQUALITY: EVIDENCE FROM ADMINISTRATIVE PANEL DATA

Abstract

Leveraging the geographic dimension of a large administrative panel on employer- employee contracts, we study the impact of robot adoption on wage inequality through changes in worker-firm assortativity. Using recently developed methods to correctly and robustly estimate worker and firm unobserved characteristics, we find that robot adoption increases wage inequality by fostering both horizontal and vertical task specialization across firms. In local economies where robot penetration has been more pronounced, workers performing similar tasks have disproportionately clustered in the same firms ('segregation'). Moreover, such clustering has been characterized by the concentration of higher earners performing more complex tasks in firms paying higher wages ('sorting'). These firms are more productive and poach more aggressively. We rationalize these findings through a simple extension of a well-established class of models with two-sided heterogeneity, on-the-job search, rent sharing and employee Bertrand poaching. We conclude that our empirical findings reveal the presence of both 'routine-biased technological change' (RBTC), whereby new technology decreases the relative demand for workers in traditional routine tasks, and 'core-biased technological change' (CBTC), whereby new technology requires workers with specialized knowledge independently of their tasks being more or less routine intensive.

JEL Classification: N/A

Keywords: N/A

Ester Faia - faia@wiwi.uni-frankfurt.de
Goethe University Frankfurt and CEPR

Gianmarco Ottaviano - gianmarco.ottaviano@unibocconi.it
Bocconi University and CEPR

Saverio Spinella - saverio.spinella@unibocconi.it
Department of Economics, New York University

Robot Adoption, Worker-Firm Sorting and Wage Inequality: Evidence from Administrative Panel Data*

Ester Faia, Goethe University Frankfurt and CEPR

Gianmarco Ottaviano, Bocconi University and CEPR

Saverio Spinella, New York University

December 14, 2022

Abstract

Leveraging the geographic dimension of a large administrative panel on employer-employee contracts, we study the impact of robot adoption on wage inequality through changes in worker-firm assortativity. Using recently developed methods to correctly and robustly estimate worker and firm unobserved characteristics, we find that robot adoption increases wage inequality by fostering both horizontal and vertical task specialization across firms. In local economies where robot penetration has been more pronounced, workers performing similar tasks have disproportionately clustered in the same firms ('segregation'). Moreover, such clustering has been characterized by the concentration of higher earners performing more complex tasks in firms paying higher wages ('sorting'). These firms are more productive and poach more aggressively. We rationalize these findings through a simple extension of a well-established class of models with two-sided heterogeneity, on-the-job search, rent sharing and employee Bertrand poaching, where we allow robot adoption to strengthen the complementarities between firm and worker characteristics.

Keywords: robot adoption, worker-firm sorting, wage inequality, technological change, finite mixture models.

JEL codes: J22, J23, J31, J62, E21, D31.

*We are grateful to the Italian National Social Security Agency (INPS) for giving us access to its administrative social security dataset through Programma VisitInps. We thank Christian Bayer, Stéphane Bonhomme, Italo Colantone, Edoardo Di Porto, Marco Grazzi, David Hemous (discussant), Leo Kaas, Philipp Kircher, Salvatore Lattanzio, Ramsus Lentz, Ben Lochner, Paolo Naticchioni, Benjamin Schoefer, Vincenzo Scrutinio, Ludo Visschers, Basit Zafar as well as participants at seminars and conferences for useful comments. Gianmarco Ottaviano gratefully acknowledges financial support within the framework of the Achille and Giulia Boroli Chair in European Studies. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n 789049-MIMAT-ERC-2017-ADG). The views expressed here are those of the authors and do not represent in any manner the EU Commission. Contacts: faia@wiwi.uni-frankfurt.de, gianmarco.ottaviano@unibocconi.it, saverio.spinella@unibocconi.it.

1. Introduction

Wage inequality has grown significantly over the past decades in all industrialized countries. There is consensus that part of this growth is due to technological change. Following the rise in college enrollment Katz and Murphy (1992) have popularized the view that, by assigning large premia to education, skill-biased technological change raises wage inequality via vertical skill acquisition. Previous waves of technology improvements, however, did not produce the same effects, prompting to ask what is special about the current wave. Some authors (see Autor et al. (2003) among others) argue that, beyond vertical skill acquisition, recent technology adoption, mostly in the form of automation and digitalization, increases wage inequality by changing the relative market values of different tasks. While with skill biased technological change new technology complements high skill workers, with routine-biased technological change new technology decreases demand for workers in traditional routine tasks while creating additional demand for workers in new complex tasks.

We push these ideas a step further by arguing that, as automation and digitalization intensify, the efficient completion of related tasks increasingly requires human operators with specialized knowledge of automated systems involving specific algorithms, software and machines. The associated growing demand for specialized knowledge is conducive to a form of workers' specialization that increasingly matters above and beyond what would be needed by the high skill content of tasks or their routine intensity. In this respect, by fostering knowledge differentiation, automation and digitalization require from workers not only vertical but also horizontal skill specialization (Faia et al. (2020)). An important specific implication is that automation and digitalization should raise the assortativity between workers' specialized skills and firms' specific tasks with far-reaching ramifications for the evolution of the within- and between-firm components of wage inequality.

The aim of the present paper is to investigate that implication and its ramifications focusing on robot adoption, which entails both automation and digitalization, namely the automation of tasks and the acquisition of softwares to operate the corresponding robots. In doing so, we leverage a unique confidential dataset covering the universe of contracts between firms and workers across Italian local economies ('provinces') from 1983 to 2020, enriched with information on robot acquisition from the International Federation of Robotics (IFR). Italy is an interesting case. It is a G7 country with a large industrial sector that has generated an average value added of 390.51 billion U.S. dollar over the period 1990-2020. In 2020 the manufacturing value added of Italy was more than four times higher than the world average (408.41 vs 94.83 billion U.S. dollars).[1] Moreover, new technology adoption plays a crucial role for Italian industries. On a scale from 0 to 1, the World Bank digital adoption index for Italy equals 0.76 overall and 0.74 for business, largely above the corresponding world averages of 0.31 and 0.36 respectively.[2]

After introducing the dataset, we start our investigation by documenting the evolution of wage inequality in Italy during the period of observation. All measures we use reveal a sizeable increase in wage inequality: the 90-10 percentile ratio, the 75-25 percentile ratio and the variance all go up, by roughly 10%, 20% and 30% respectively. We then examine a simple wage variance decomposition across firms, occupations ('tasks') and sectors. We find that the between-firm component is more important than the within-firm component, with a cumulative increase of the former almost five times larger than the latter. Further refining the analysis, we also find that the between-firm, between-sector and between-task is the single most important component of overall wage variance. Accordingly wage dispersion has increased not only across firms, but also across tasks and sectors.

Clearly, these findings do not necessarily imply an increase in the assortativity between workers' specialized skills and firms' specific tasks as the evolution of match heterogeneity can be driven by firm characteristics, worker characteristics or their combination. Formally, after controlling for observables, between-firm wage inequality can be driven by the variance of unobserved firm characteristics, the variance of the average characteristics of the firm's workforce or the correlation between unobserved firm and worker characteristics ('sorting'). Moreover, the increase in assortativity during the period of observation may not be necessarily due to technological change. Those are ultimately empirical questions, which can be answered with a rich administrative dataset.

Our empirical strategy consists of two stages involving the estimation of how sorting evolves through time and the assessment of the causal effect of robot adoption on that evolution. Specifically, in the first stage we discuss the thorny problems one faces to separately identify sorting from firm and worker characteristics in matched

[1] See: https://www.theglobaleconomy.com/rankings/industry_value_added.

[2] See: https://www.worldbank.org/en/publication/wdr2016/Digital-Adoption-Index.

employer-employee wage data, and propose a solution based on Bonhomme et al. (2019). Classical reduced-form estimation with additive firm and worker fixed effects, albeit very versatile (Abowd et al. (1999)), neglects complementarity-induced non-linearities, and it is often affected by an identification bias as well as by an incidental parameter bias, which arises whenever many parameters are estimated with relatively few observations (Andrews et al. (2012)). This is particularly relevant in our case. The estimation of firm fixed effects requires the same firm to employ different workers, which is typically what we see in the data. Analogously, the estimation of worker fixed effects requires the same worker to work for different firms, which however happens only to workers who change employer ('movers'). As the number of observations is given by the sum of the numbers of firms and workers (which are themselves equal to the numbers of firm and worker fixed effects) plus the number of movers (who are typically relatively few), the incidental parameter bias is often referred to as the 'low mover bias' associated with lack of power in the estimation of worker fixed effects.[3]

To solve these problems, we follow the approach proposed by Bonhomme et al. (2019) to identify and estimate earnings distributions and worker composition on matched panel data allowing for two-sided worker-firm unobserved heterogeneity. To reduce the number of estimating parameters, they suggest to proceed in two steps. Firms are first partitioned into 'classes' by a dimension reduction method based on a machine learning ('k-means') algorithm. Estimation is then performed with firm class fixed effects rather than individual firm fixed effects. As for workers, their heterogeneity is captured through random fixed effects after reducing its dimensionality by approximating the workers' distribution via a finite support population density. The result is a finite-mixture specification that is estimated by maximum likelihood including interacted firm-class fixed effects to account for potential complementarity-induced non-linearities. This specification is used to collapse worker heterogeneity in a limited number of probabilistic 'types'.

An extension we make to Bonhomme et al. (2019) is due to the fact that, in order to study the effects of technological change on sorting, we need time varying estimates of the correlation between unobserved firm and worker characteristics, which themselves call for time varying estimates of those characteristics. This estimation requires a

[3]See Eeckhout and Kircher (2011) for a discussion of identification in assortative matching environments, and Andrews et al. (2012) for a test showing the sensitivity of two-way fixed effects estimates to the share of movers.

time window that is, on the one hand, wide enough to accommodate a large enough number of movers and, on the other hand, narrow enough to consider unobserved firm and worker characteristics as reasonably stable. Relying on Swedish matched employer-employee panel data from 2002 to 2004, Bonhomme et al. (2019) use a two-year window for their static model and a four-year window for their dynamic model. In the same vein, exploiting the longer time series dimension of our Italian data from 1983 to 2020, we obtain time varying estimates of unobserved firm and worker characteristics re-estimating them every second year over partially overlapping 4-year intervals.

Another extension we make to Bonhomme et al. (2019) stems from the fact that, in order to estimate sorting in local economies, unobserved firm and worker characteristics have to be themselves estimated at the local level. Assigning firm classes to local economies by the addresses of the firms they include is relatively straightforward as the k-means algorithm provides an exact partition of firms into such classes. This is not the case, however, for worker types as the same does not hold for the probabilistic types obtained from the finite mixture specification. We tackle this issue by computing the probabilities that workers belong to the different worker types and associate them with their highest probability type. We then assign worker types to provinces by the addresses of the workers they include.

In the second stage of our empirical strategy, we regress our time varying sorting estimates on the exogenous variation of automation at the local level. This is captured through a shift-share instrument à la Acemoglu and Restrepo (2020), which imputes the sectoral changes in the IFR stock of robots over value added to a local economy based on its sectoral employment shares. The instruments is computed every two years to match the frequency at which sorting is estimated. The sectoral changes in the stock of robots over value added are constructed by averaging across the US, Japan and several European countries (other than Italy).

In the first stage, we find that the correlation between unobserved firm-class and worker-type characteristics is positive and accounts for a relevant part of the wage variance across matched firm-class and worker-type pairs. This part, though smaller than the part explained by unobserved worker-type characteristics, is larger than the part explained by unobserved firm-class characteristics. When we decompose the wage variance within and between firm classes, we also find evidence of a tendency of different worker types to appear in different firm classes. Hence, we observe not only 'sorting' as superior (inferior) firm classes tend to match with superior (inferior) worker types, but also 'segregation' as different worker types tend to cluster in different firm classes. As different firm classes are characterized by different occupations (but not necessarily in terms of complexity), segregation implies horizontal specialization. To shed light on the underlying mechanism, we correlate the firm classes with observable firm characteristics and find that superior firm classes are associated with higher value added per worker ('labor productivity') and are located in the most developed local economies. We also correlate the worker types with one-digit ISCO occupational categories ordered from the least to the most intensive in routine tasks, with higher routine intensity signalling lower task complexity. We find that superior worker types are associated with more complex tasks. We interpret this finding as evidence of vertical task specialization across firms. However, the correlation between worker types and occupational categories also implies that the segregation of worker types in different firm classes entails the parallel segregation of occupations in those classes. This is consistent with assortativity between workers' specialized skills and firms' specific tasks, which we interpret as evidence of horizontal task specialization across firms and confirm by assessing the relative importance of segregation for the correlation between firm classes and worker types.

Overall, the results of the first stage of our analysis support the conclusion that wage inequality is driven by both vertical and horizontal task specialization across firms. In the second stage, where we test whether specialization is caused by robot adoption, we find that this is indeed the case as our shift-share instruments foster both sorting and segregation.

Finally, we round off the paper by showing that our empirical findings can be rationalized within a theoretical random-search framework with two-sided heterogeneity, on-the-job search, job poaching, and a bargaining process à la Rubinstein (1982) in the wake of Cahuc et al. (2006) and Bagger and Lentz (2019) together with endogenous workers' search intensity as in Postel-Vinay and Robin (2004). In this framework, we model the impact of automation as strengthening production complementarities between workers' specialized skills and firms' specific tasks and disproportionately increasing the endogenous search intensity of higher worker types.[4] We then show that

[4]See also Elsby et al. (2022) for a model in which firm task specificity affects firms surplus and in

stronger complementarities and the implied higher search intensity for higher worker types increase wage dispersion with a relevant role played by both between and within firm class dispersion.

Literature Review. The relations of our paper with the existing literature are manifold. First, it relates to the expanding empirical literature that studies task-biased technological change and its role for inequality (see Autor et al. (2003), Autor et al. (2006) or, more recently, Cortes, Lerche, Schönberg, Tschopp 2020. Jaimovich et al. (2021) examines the link between automation and the fall in routine-intensive occupations.). Second, it relates to the literature examining more broadly within-versus between-firm inequality (see Card et al. (2013), Song et al. (2018), Barth et al. (2016), Helpman et al. (2017)) or studying the rise in market concentration and 'superstar' firms (see Autor et al. (2020), Dorn et al. (2017), Azar et al. (2020b) or Azar et al. (2020a)). While this literature emphasizes the role of productivity, we highlight the role of sorting in disproportionately increasing the surplus of the superstar firms, which are more likely to poach, hire and retain highly specialized workers for their tasks. Third, our focus on task-biased technological change links our paper to Autor et al. (2003), Autor et al. (2020) and Cortes et al. (2020), while the link between automation and wage inequality connects it to Acemoglu and Restrepo (2020), Hémous and Olsen (2022) and related studies. With respect to all these works, what distinguishes our paper is its focus on the effects of automation on wage inequality as mediated by horizontal task specialization on top of vertical task specialization. For this reason we devote particular attention to the robust identification of sorting in the data.

Fourth, in terms of econometric methodology, our work builds on a recent literature (see Andrews et al. (2012) and Bonhomme et al. (2019)) that introduces complementarities in the classical two-way fixed effect estimation of earnings distributions and worker composition on matched panel data with two-sided worker-firm unobserved heterogeneity (see Abowd et al. (1999) or Card et al. (2013)). With respect to this literature, we augment the finite-mixture model of Bonhomme et al. (2019) with a procedure to locally assign workers in specific occupations to the estimated wage bins. The methodology successfully deals also with the identification bias highlighted by Eeckhout and Kircher (2011) in the classical two-way fixed effect estimation.[5]

turn on-the-job search and Elsby and Gottfries (2022) for a model with firm dynamic and on-the-job search.

[5]Models à la Postel-Vinay and Robin (2002) and Cahuc et al. (2006) generate a dual wage hierarchy as firms can be ranked both in terms of the wages required to poach their employees and the wage

Finally, our paper contributes to the theoretical literature on two-sided heterogeneity, rent sharing, and firm-to-firm worker mobility due to on-the-job search and employeed poaching as modeled by Postel-Vinay and Robin (2002) and Cahuc et al. (2006) through a bargaining game à la Rubinstein (1982). With respect to this literature, we emphasize the role technological change plays for wage inequality in fostering the assortativity between workers' specialized skills and firms' specific tasks.

The rest of the paper is organized as follows. Section 2 introduces the dataset and documents the evolution of wage inequality. Section 3 presents the empirical strategy. Section 4 reports the econometric results. Section 5 presents the theoretical model. Section 6 concludes.

2. Data Description and Stylized Facts

Our investigation of the effects of robot adoption on wage inequality through firm-worker assortativity leverages a unique confidential dataset provided by the Italian National Social Security Agency (INPS) on contracts between Italian firms and workers from 1983 to 2020, enriched with information on robot acquisition made available by the International Federation of Robotics (IFR) from 1995 to 2017.

Contracts. The INPS dataset reports administrative social security data on a matched employer-employee panel covering the universe of Italian firms from 1983 to 2018. Contract level information for workers is linked to several firm-level variables.

Workers. On the worker side, the dataset reports information on all contracts in a given year. In each year we select the contract with the longest duration and we break ties by keeping the highest paying contract as standard in the literature. To control for the gender pay gap, we focus on male workers, which is also quite common in the literature. We correct for the intensive margin of labor supply by using weekly wages and accommodate for part-time workers by using 'full-time equivalent weeks' to compute the weekly rate.[6] Wages are deflated by the CPI index from the OECD with base year 2015. The dataset allows us to locate firms and workers at the NUTS 3 level ('province') from 1983 to 2019.

premia they offer to new hires. Exploiting a linked panel dataset similar to ours (i.e. the INPS-INVID dataset, which covers 5% of the full linked data), Di Addario et al. (2022) assess the relative importance of a worker's current employer and the employer from which the worker is hired for the determination of the hiring wage. Enriching the fixed effects approach in Abowd et al. (1999) with worker fixed effect, a fixed effect for the destination firm hiring the worker, and a separate fixed effect for the origin of the hire, they conclude that the latter is more important.

[6]'Full-time equivalent weeks' are reported in the dataset.

Firms. On the firm side, contracts are linked to identifiers from CERVED.[7] Specifically, the CERVED-INPS panel contains balance sheet data for the universe of private firms in Italy from 1996 to 2018. We use this information to estimate production functions and to compute sectoral employment shares.

Technology Adoption. The IFR collects information on automated robot purchases for all countries in the world. As we will describe in Section 3.2, we exploit a panel of 25 countries from 1995 to 2017 to build exogenous measures of robot adoption.

Wages. We employ two different measures of wage earnings. The first is salaries as reported in workers' contracts. The second is a Mincer residual from the following regression:

$$\log(Y_{ijt}) = \alpha + \sum_{r=2}^{3} \beta_r(\text{age}_{it} - 40)^r + \gamma\,\text{tenure}_{ijt} + \sum_m \sum_t (\eta_m \times \zeta_t) + u_{ijt}, \qquad (1)$$

where Y_{ijt} is salary of worker i employed by firm j in year t, age $_{it}$ is the worker's wage, tenure$_{ijt}$ is the time the worker has been employed by firm j, η_m is a dummy for two-digit Ateco sectors (the Ateco classification is the one employed in Italy and has a general correspondence with NAICS codes), ζ_t a year dummy and u_{ijt} is an error term. As in Boeri et al. (2021), we also control for sectoral fixed effects since the sectoral composition of the workforce at the local level may conflate the identified fixed effects.[8]

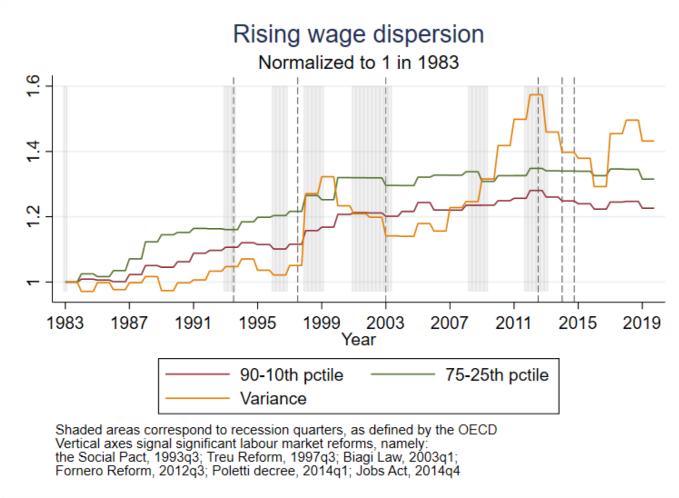
Time Intervals. An important aspect of the methodology we use to identify unobserved firm and worker characteristics concerns the choice of the time window for the estimation. On the one hand, larger windows increase the number of movers, which are crucial for identifying unobserved worker characteristics. On the other hand, smaller windows increase the number of point estimates in the time dimension. We choose a four-year window, and to increase the number of point estimates, we adopt the R-AKM technique formalized by Lachowska et al. (2020) and employ a rolling four-year window for the estimates of the worker-firm fixed effects (see Online Appendix A for additional details). This results in nine estimation intervals for which automated robot information is available.

Focusing on the salaries reported in workers' contracts for descriptive purposes, our data provide a vivid picture of the evolution of Italian wage inequality and its components

[7]Cerved is a leading Information Provider in Italy and one of the major rating agencies in Europe.
[8]Following Card et al. (2018), we use a third degree polynomial of age excluding the linear term.

in the period of observation. Three main facts stand out. First, as shown in Figure 1, this is a period of sizeable growth in wage inequality. The figure considers three inequality measures: the 90-10 percentile ratio, the 75-25 percentile ratio and the variance. They all go up, by roughly 10%, 20% and 30% respectively.

Figure 1: Wage dispersion Over Time



Note: Earnings dispersion, 1983-2019. We report the difference between the 75th and 25th percentile of the log weekly wage distribution, as well as the difference between the 90th and the 10th, and the variance of our earnings measure. Salaries are collected for male employees, aged 20-60, and corrected for part-time working arrangements.

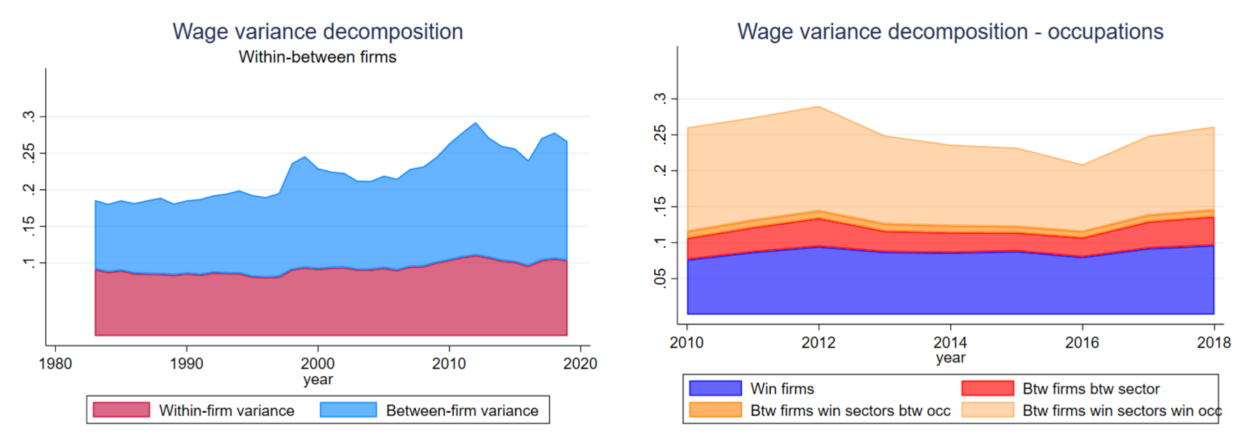
Second, the left panel of Figure 2 shows that the growth in wage inequality is mainly driven by rising wage dispersion between firms. Third, the between firm component has increased mostly in the form of more wage dispersion between tasks and sectors. These additional facts, displayed in the right panel of Figure 2, are revealed by the following wage variance decomposition, which enriches the canonical one (see Song et al. (2018)) by cutting across occupations ('tasks') and sectors:

$$\mathrm{var}(\log(Y_i)) = \underbrace{n^{-1}\sum_j \sum_{(i\in j)} (\log(Y_i) - \overline{\log(Y_j)})^2}_{\text{within firm}} + \underbrace{n^{-1}\sum_s n_s(\overline{\log(Y_j)} - \overline{\log(Y_s)})^2}_{\text{between firm, within sector}} +$$

$$+\underbrace{n^{-1}\sum_{so} n_{so}(\overline{\log(Y_s)} - \overline{\log(Y_{so})})^2}_{\text{between firm, between sector, within task}} + \underbrace{n^{-1}\sum_{so} n_{so}(\overline{\log(Y_{so})})^2}_{\text{between firm, between sector, between task}}$$

where Y_i is individual wage, the upper bar denotes means while j, o and s index firms, (3-digit) occupations and (3-digit) sectors respectively. The within-task component captures deviations of individual wages from the mean wage in their task, whereas the between-task component captures deviations of the mean task wage from the aggregate mean wage. The right panel in figure 2 shows that the between-firm, between-task and between-sector component accounts for the largest part of the variance.

Figure 2: Variance decompositions



(a) Between-Within Firms **(b) Tasks, Firms, Industries**

Note: The wage dispersion is here characterized as the sum of two components, the within-firm component and the between-firm component. The former is defined as the variance of the individual deviations from their employers average wage, the latter as the variance of the firm average wages.

Clearly, these facts do not necessarily imply an increase in the assortativity between workers' specialized skills and firms' specific tasks as between-firm wage inequality may be driven by the variance of unobserved firm characteristics, the variance of worker characteristics or the correlation between unobserved firm and worker characteristics (i.e. 'sorting'). Moreover, the increase in assortativity during the period of observation may not be necessarily due to technological change. In the next section we present how we approach the solution of these identification problems.

3. Empirical Strategy

Our approach to testing the impact of robot adoption on wage inequality through sorting consists of two main stages. In the first stage we employ an econometric methodology à la Bonhomme et al. (2019) to identify and estimate unobserved worker and firm characteristics and then their correlation. We will argue that this methodology has several advantages in terms of properly accounting for non-linearities and alleviating identification as well as incidental parameter biases. The first stage is itself divided into a clustering pre-stage, which helps to improve identification, and a sub-stage, in which we estimate a finite-mixture random effect specification for wages. The outcome of the first stage is a robust and precise estimation of the correlation between unobserved worker and firm characteristics. In the second stage, we regress the estimated correlation on an exogenous measure of robot adoption à la Acemoglu and Restrepo (2020), while controlling for other factors that may affect sorting.

3.1. First Stage: Estimating Sorting between Firms and Workers

Classical econometric specifications (see e.g. Abowd et al. (1999), Card et al. (2018)) devoted to separately identify the roles of firm characteristics, worker characteristics and their combination typically feature the separable additive structure:

$$y_{ijt} = \theta_i + \psi_{j(i,t)} + \bar{X}_{it}\bar{\beta} + \varepsilon_{ijt} \tag{2}$$

where i, j and t index workers, firms and time respectively (with assignment function $j \leftarrow j(i,t)$). According to (2), the log wage y_{ijt} of worker i, employed by firm j at time t depends on the employee's individual ability to command a wage premium (θ_i), the wage-setting policy or the characteristics of the employer, ($\psi_{j(i,t)}$) and a set of individual or match specific characteristics ($\bar{X}_{it}$). This specification implies that, if one abstracts from other covariates, the wage variance $\text{var}(y_{ijt})$ can be decomposed in the variance of worker characteristics, the variance of firm characteristics and their covariance:

$$\text{var}(y_{ijt}) = \text{var}(\theta_i) + \text{var}(\psi_{j(i,t)}) + 2\ \times \text{cov}(\theta_i,\ \psi_{j(i,t)}) + \text{var}(\varepsilon_{ijt}) \tag{3}$$

where the covariance of worker and firm characteristics is equal to the product of their correlation multiplied by their standard deviations: $\text{cov}(\theta_i,\ \psi_{j(i,t)}) = \text{corr}(\theta_i, \psi_{j(i,t)})\text{sd}(\theta_i)\text{sd}(\psi_{j(i,t)})$. As this correlation measures sorting between firms and workers, decomposition (3) highlights the role of sorting for wage dispersion. A richer decomposition of the wage variance (see e.g. Song et al. (2018)) shows that sorting affects wage inequality through its between-firm component:

$$\begin{aligned} \text{var}(y_{ijt}) = {} & \underbrace{\text{var}(\theta_i - \bar{\theta}^j) + \text{var}(\varepsilon_{ijt})}_{\text{within-firm inequality}} + \\ & + \underbrace{\text{var}(\psi_{j(i,t)}) + 2 \times \text{cov}(\bar{\theta}^j,\ \psi_{j(i,t)}) + \text{var}(\bar{\theta}^j)}_{\text{between-firm inequality}} \end{aligned} \tag{4}$$

where $\bar{\theta}^j$ denotes average worker characteristics in firm j. In particular, according to (4), with positive assortativity (i.e.$\text{corr}(\theta_i, \psi_{j(i,t)}) > 0$), more sorting (here expressed as higher $\text{corr}(\bar{\theta}^j, \psi_{j(i,t)})$) fosters wage inequality by raising its between-firm component through higher $\text{cov}(\bar{\theta}^j,\ \psi_{j(i,t)})$.

These decompositions show that a correct understanding of wage inequality hinges on a precise and robust estimation of firm and worker characteristics and their combination. Although very versatile, traditional specifications such as (2) face two main challenges. They neglect complementarity-induced non-linearities that plausibly exist between worker and firm characteristics, and are often plagued by an identification bias as well as by an incidental parameter bias. The latter arises whenever many parameters are estimated with relatively few observations (Andrews et al. (2012)). This is particularly relevant in our case. The estimation of firm fixed effects requires the same firm to employ different workers, which is typically what we see in the data. Analogously, the estimation of worker fixed effects requires the same worker to work for different firms, which however happens only to workers who change employer ('movers'). As the number of observations is given by the sum of the numbers of firms and workers (which are themselves equal to the numbers of firm and worker fixed effects) plus the number of movers (who are typically relatively few), the incidental parameter bias is often referred to as the 'low mover bias'. This is associated with lack of power in the estimation of worker fixed effects and may also lead to misleading identification.

Both challenges could be properly addressed using interacted random effects specifications, which are however computationally infeasible. Highly refined two-way fixed effects specifications could capture non-linearities, but the large number of parameters, against the low number of observations reduces estimates precision even further.[9] An alternative way has been recently put forth by Bonhomme et al. (2019) who propose to estimate the following finite-mixture specification:

$$y_{ijt} = \theta_j + \psi_{j(i,t)}\Theta_i + \bar{X}_{it}\bar{\beta} + \varepsilon_{ijt} \tag{5}$$

[9]Random effects imply fewer parameters to estimate and correlated random effects à la Chamberlain (1980) allow for non-linearities. The corresponding maximum-likelihood algorithm is, however, very demanding from a computational point of view.

3.1.1. Firm Classes

Firm clustering is obtained by applying a weighted k-means algorithm to the firm wage distribution that solves the following optimization problem:

$$\min_{k(1),\ldots,k(J),H_1,\ldots,H_{10}} \sum_{j=1}^{J} n_j \int \left(\hat{F}_j(y) - H_{k(j)}(y) \right)^2 d\mu(y) \tag{6}$$

where J is the number of firms in the dataset (indexed $j = 1, ..., J$), $k(1), \ldots, k(J)$ is a partition of them in K clusters ('classes'), $\hat{F}_j$ is the empirical cumulative density function of wages in firm j. The number of firm clusters K is set to 10, but we check robustness to changing their number (see Figure B.2 in the Online Appendix). $H_1, \ldots, H_{10}$ are cumulative density functions of the firm clusters. $\mu(y)$ are discrete moments of the firms' wage distributions - in particular, for each firm they are 19 percentiles (from the 5th to the 95th) of its wage distribution. n_j is the number of workers in firm j and its presence implies that in (6) each firm is weighted by the number of its employees.

3.1.2. Worker Types

Worker types are obtained by approximating the distribution of worker wages through a finite support population density. Following Bonhomme et al. (2019), we set the number of worker types L to 6, denoting them by α_1, α_2, α_3, α_4, α_5, α_6. The approximated distribution is then estimated through maximum likelihood using a finite mixture model. The estimation relies on the following joint probability distribution:

$$\begin{aligned} p(\bar{y}_{i|1}, \ldots, \bar{y}_{i|L} \mid m = 1) = \sum_{k=1}^{K} \sum_{k'=1}^{K} \mathbb{I}(\hat{k}_{i2} = k)\mathbb{I}(\hat{k}_{i3} = k') \ldots \\ \ldots \prod_{\alpha=1}^{L} p_{kk'}(\alpha, \theta_p) f_{y_{12},k,\alpha}(y_{i1}, \theta_f) f_{k,k',\alpha}(y_{i2}, y_{i3}, \theta_m) f_{y_{i3},k',\alpha}(y_{i4}, \theta_b) \end{aligned} \tag{7}$$

where $\bar{y}_{i\,|\,\alpha} = \{y_{i1}, y_{i2}, y_{i3}, y_{i4} \,|\, \alpha\}$ refers to the vector of observed yearly wages in the four-year estimation window for mover i ($m = 1$) of type α. The objects of interests are the probability $p_{kk'}(\alpha)$ that a worker of type α moves from firm class k to firm class k', and the wage distributions f's parametrized by θ_p, θ_f, θ_m and θ_b. Each wage distribution depends on both the firm class and the worker type ($f_{k,\alpha}$), and the wage at any time t depends on future and past earnings. The corresponding log-likelihood function then is:

$$\sum_{i=1}^{n} \log p(\bar{y}_{i|1}, \ldots, \bar{y}_{i|L} \mid m = 1) \tag{8}$$

where n is the total number of workers. The finite mixture log-likelihood function is maximized using the Expectation-Maximization (EM) algorithm. The log-likelihood function allows us to estimate the distribution parameters θ_p, θ_f, θ_m and θ_b. While (7) applies to movers ($m = 1$), the joint probability distribution for stayers ($m = 0$) can be obtained by replacing $p_{kk'}(\alpha)$ with the share of stayers of type α in firm class k so as to retrieve the overall density of workers of type α in firm class k_i. To address the possibility of local maxima, we iterate the EM algorithm multiple times, select iterations with the highest likelihood and, among these, eventually pick the one that maximizes network connectivity as defined by Jochmans and Weidner (2019).

3.1.3. Time-Varying Local Sorting

The correlation of estimated firm-class characteristics and estimated worker-type characteristics gives an estimate of sorting. For our purposes, however, this estimate has to vary both across time and across local economies. These features require to innovate with respect to Bonhomme et al. (2019) in two ways.

In a dynamic setting, the finite-mixture method estimation by Bonhomme et al. (2019) is applied using a single time window of four years. Since our data sample spans the period from 1983 to 2020, we can run our estimation on several windows, repeating our k-means classification at the start of each window. However, the possibility of obtaining a long enough time series of sorting estimates is crucial for our analysis given that its second stage consists in regressing the sorting estimates on measures of technology adoption. The statistical power of the second stage clearly depends on the number of our sorting estimates. To maximize this number while sticking to four-year windows, we re-estimate firm-class and worker-type characteristics every second year over partially overlapping 4-year intervals.

The second extension we make to Bonhomme et al. (2019) is motivated by the fact that, in order to estimate sorting in local economies, unobserved firm and worker characteristics have to be themselves estimated at the local level. Assigning firm classes to local economies by the addresses of the firms they include is relatively straighforward as the k-means algorithm provides an exact partition of firms into such classes. This is not the case, however, for worker types as the same does not hold for the probabilistic types obtained from the finite mixture specification. We tackle this issue by computing the probabilities that individual workers belong to the different worker types and associate them with their highest probability type. We then assign worker types to provinces by the addresses of the workers they include. Specifically, we assign worker i to the type that maximizes the posterior probability of the worker being of that type as estimated through Bayesian updating:

$$\arg\max_{\alpha^*} p(\alpha = \alpha^* | y_i, k_i) = \arg\max_{\alpha^*} \frac{f_{k_i,\alpha^*}(y_i)\ q_{k_i}(\alpha^*)}{\sum_j^J f_{k_i,\alpha_j}(y_i)\ q_{k_i}(\alpha_j)} \tag{9}$$

where y_i is worker i's wage, k_i is the worker's employer class, $f_{k_i,\alpha}(y_i)$ is the density of workers of type α in firm class i with wage y_i, and $q_{k_i}(\alpha)$ is the overall density of workers of type α in firm class k_i.

3.2. Second Stage: Estimating the Impact of Robots on Sorting

The key output of the first stage is the estimated correlation between firm-class and worker-type characteristics across local economies, which gives us a measure of sorting at biannual frequency. In the second stage this measure is regressed on exogenous robot adoption. We adopt a classical Bartik instrument based on shift shares along the lines of Acemoglu and Restrepo (2020). See Borusyak et al. (2022) who provide a recent discussion of shift-share instrumental variable (SSIV) regressions. Specifically the shift instrument reads as follows:

$$\text{automation}_p^{IV} = \sum_{s \in S} L_s^p \left[\frac{1}{N} \sum_{c=1}^{N} \left(\frac{dM_s^c}{L_s^c} - \frac{dY_s^c}{Y_s^c} \frac{M_s^c}{L_s^c} \right) \right] \tag{10}$$

where M_s^c and Y_s^c are the stocks of automated robot and value added in sector s respectively, dM_s^c and dY_s^c are the changes in the two stocks over a two-year interval matching the biennal frequency of the sorting measure, and L_s^c is the employment share of sector s. The superscript c refers to a country in a set of $N = 24$ countries including the US, Japan and Europe excluding Italy.[10]

Accordingly, the term between brackets measures the average exposure to robot adoption in sector s across that set of countries. As these countries are similar to Italy, their exposure should be similar to the Italian one, but independent of Italy-specific

[10] The detailed list of countries is: AT, BE, BG, CZ, DE, DK, EE, ES, FI, FR, HU, IE, JP, LT, LV, MT, NL, PL, PT, RO, SE, SK, UK, US.

developments. While data on M_s^c come from IFR, those on Y_s^c and L_s^c are taken from EUKLEMS. Different sectoral exposure is apportioned across Italian local economies ('provinces') through the importance of each sector for local employment L_s^p. Hence, automation_p^{IV} measures the exposure of Italian local economy p to robot adoption that is independent from local and national shocks in Italy. Thus, we use such measure to instrument for the endogenous robot adoption, measured as follows:

$$\text{automation}_p = \sum_{s \in S} L_s^p \left(\frac{dM_s}{L_s} - \frac{dY_s}{Y_s} \frac{M_s}{L_s} \right) \tag{11}$$

Equipped with an arguably exogenous measure of robot adoption, we test its impact on sorting through the following specification:

$$\begin{aligned} \text{sorting}_{p\tau} = \alpha + \beta \, \text{automation}_{p\tau} + \sum_{i \in \{25, 50, 75\}} \gamma_i \, \text{HHI}_{i\tau} + \\ + \delta_1 \, \text{Share Manufacturing}_{p\tau} + \delta_2 \, \text{Share Construction}_{p\tau} + \zeta_\tau \times \lambda + \varepsilon_{p\tau} \end{aligned} \tag{12}$$

where p is the local economy index, τ is a two-year time period and $\text{automation}_{p\tau}$ is instrumented robot adoption. The other variables control for factors that may affect sorting independently of robot adoption: $\text{HHI}_{i\tau}$ is the i-th percentile of the Herfindahl–Hirschman index measuring employment concentration across two-digit sectors, Share $\text{Manufacturing}_{p\tau}$ and Share $\text{Construction}_{p\tau}$ are the employment shares of the corresponding sectors, and ζ_τ is a bi-annual dummy. The interacted λ is a dummy for the geographical macro areas in which Italy is statistically partitioned in decreasing order of economic development: North-West, North-East, Centre, South, and Islands. The Herfindahl–Hirschman index serves the purpose of controlling for the possible effects of market concentration on sorting due to firm monopsonistic power.

4. Empirical Results

We now present the results of our two-stage analysis. As for the first stage, we start with a comparison of our estimated correlation between firm-class characteristics and worker-class characteristics with the correlation obtained from traditional approaches. This is followed by a discussion of the implications of our estimated correlation for wage variance decompositions. As the firm classes and the worker types derived from the finite-mixture approach have only implicit connections to the underlying firm and worker characteristics, to shed further light on the drivers of wage inequality, we then make those connections more explicit by relating the firm classes and the worker types to observable firm and worker characteristics. Lastly, we turn to the second stage discussing the results on whether and how robot adoption affects wage inequality through the recombination of firm classes and worker types.

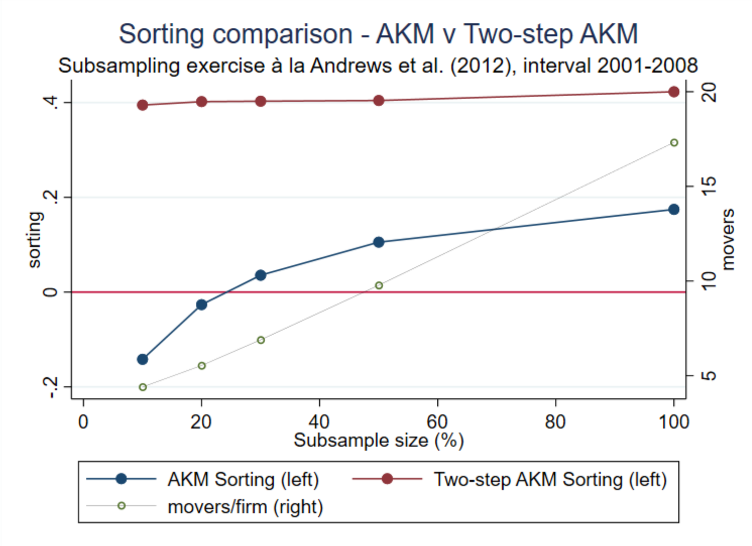
4.1. Sorting Estimation Revisited

Sorting is our key outcome variable. We have discussed the threats to the robustness and precision of its estimation in the traditional linearly additive fixed effect specification (2). To show that those threats are indeed consequential, we compare the sorting estimates obtained from the finite-mixture specification (5), which we call 'BLM', with those obtained from two alternative approaches. The first implements (2) with additive individual firm and worker fixed effects. The second gets closer to the two-step logic of (5) by implementing (2) with firm-class fixed effects after clustering firms using the k-means algorithm. We call the former 'AKM' and the latter '2s-AKM'. In all cases sorting is estimated in the first two years of partially overlapping four-year intervals. Compared with the approach based on (5), the second alternative approach dispenses with the maximum likelihood estimation of worker types.

First, we replicate the subsampling exercise of Andrews et al. (2012) to show the sensitivity of the sorting estimates based on (2) to sample size. In particular, we compare AKM and 2s-AKM estimates from the same subsamples. The exercise runs as follows. We take a 10% random sample of workers, and define p as the proportion of workers sampled ($p = 0.1$). We record the identities of all firms that employ the sampled workers and we keep the largest connected set of firms and workers. Keeping the sampled firms fixed, we increase p to 0.2, 0.3, 0.5 and 1. We generate ten k-means firm classes for each value of p and estimate the corresponding sorting by AKM and 2s-AKM. The results are summarized in Figure 3, which clearly shows how clustering leads to more stable estimates.

Second, we compare the sorting estimates from AKM, 2s-AKM and BLM. Here 2s-AKM is estimated using the same sample as AKM, for direct comparability. BLM, instead, retains only movers who switch job only between the second and third year. We present the three sorting estimates in Figure 4 (their cumulative increases in Figure C.1 is reported in Appendix C.2): the left panel show the level estimates and the right panel their period-on-period increases. The reason for showing those additional

Figure 3: Subsampling exercise



Note: Here we replicate the subsampling exercise from Andrews et al. (2012), showing how the sorting estimates of AKM are over-reliant on the number of movers per firm. we compare the AKM estimates with the ones from 2s-AKM. The latter are significantly more stable and less dependent on the number of movers.

transformations is twofold. First, cumulative and period-on-period increases track the dynamics of sorting. Second, it is often believed that the increment in sorting, which is a variable of interest in most analyses of worker and firm contributions to wage inequality, tends to remain stable independently of the degree of network connectivity. Hence, it should be unaffected by k-means clustering. Our results show that this is not the case.

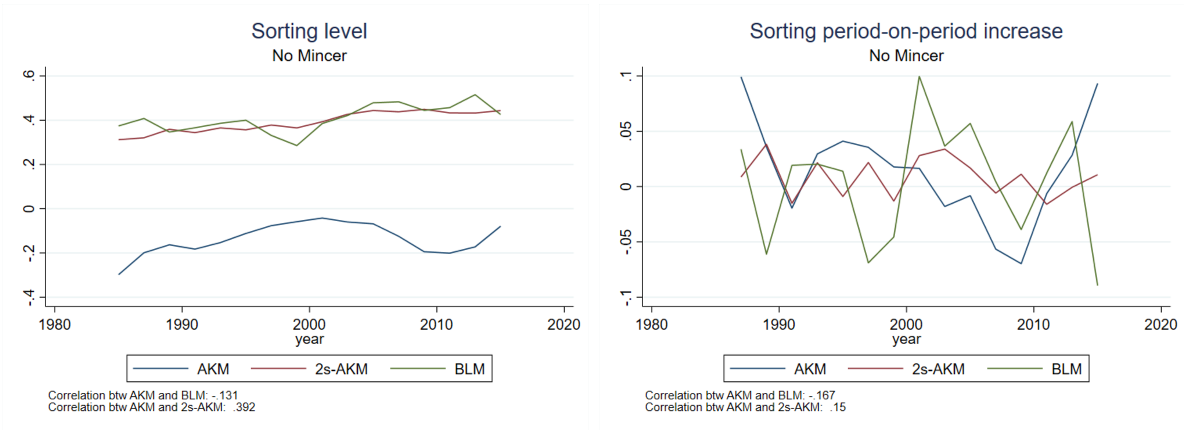
The left panel in Figure 4 shows that there is a large difference between the AKM estimates on the one side and those from 2s-AKM and BLM on the other. Moreover, the right panel in Figure 4 highlights that a large difference between estimates with and without clustering also exists in the change of sorting over time. The fact that the two-step approaches lead to very similar results suggests that firm clustering significantly mitigates the incidental parameter bias. However, the fact that the period-on-period and cumulative estimated increase is not perfectly aligned between 2s-AKM and BLM suggests that allowing for non-linearities may be also important.

4.2. Variance Decomposition Revisited

Focusing now on our implementation of BLM, Figure 5 reports the results for the variance decomposition (3) in the case of Mincer wages respectively.[11] The equivalent for raw wages is report in figure C.2 of Online Appendix C.2.

[11] The Mincer residual from OLS regression 1 explains 20% of total wage variance.

Figure 4: Sorting levels (left panel) and sorting increase (right panel).



(a) Sorting Levels **(b) Sorting Levels**

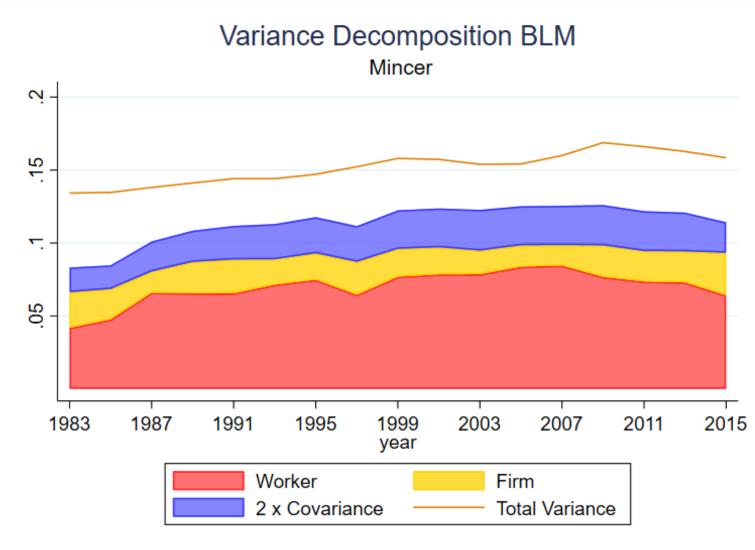
Note: Sorting levels (left panel) and period-on-period increase (right panel) estimated through Abowd et al. (1999), two-step additive fixed effects, and finite-mixture model. The first two algorithms run on the same sample of workers, hence they are directly comparable. The finite-mixture requires restrictions in the pattern of job movements allowed in the sample. Sorting is estimated on 4-year, partially overlapping intervals. The sorting is calculated in the first two year of each interval.

First, more than half of the explained variance is due to the worker heterogeneity. The role of firms is rather small, while the covariance component is significant and sizeable, a feature seldom captured by baseline Abowd et al. (1999) estimates. It is remarkable that, despite discretizing of two-sided heterogeneity, worker types and firm classes are able to explain about 80% of the total wage variance in all periods. Most importantly, positive covariance reveals the presence of positive assortativity.

The predominance of the worker fixed effect may seem in contradiction with our stylized fact in Figure 2 showing (in the right panel) that the between-firm, between-task and between-sector, component accounts for the largest part of the variance. This is not the case. The presence of positive assortativity sheds light on the underlying mechanism. Superior firms hire workers with specialized knowledge (‘core competencies’) that allow them to better perform the main tasks in their industries. They thereby generate larger surplus and pay higher wages by passing higher rents to those workers. This selective composition of the workforce explains why decomposition (3) assigns a larger role to worker characteristics.

To elaborate further on the role of sorting, we compute and plot the alternative decomposition (4), which separates the within-firm and between-firm components. The corresponding results are shown in Figure 6 for the Mincer and in figure C.3 from Online Appendix C.2 for the raw wages. Worker heterogeneity within firm class

Figure 5: Baseline wage variance decomposition - Mincer



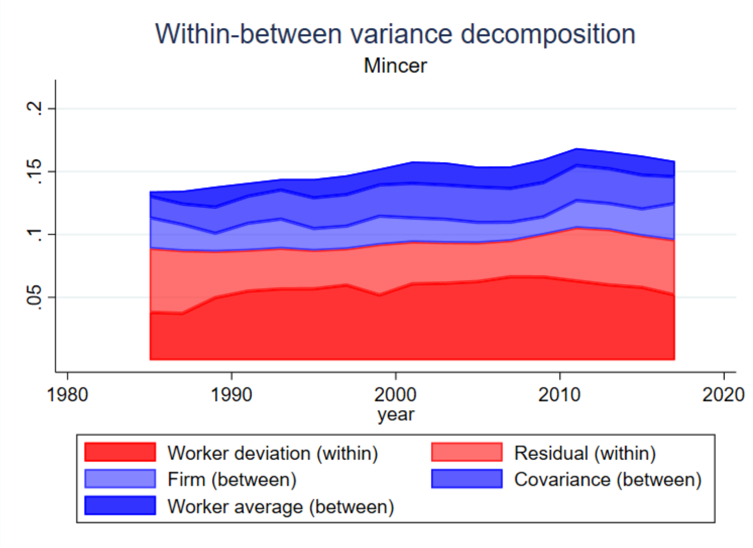
Note: The variance decomposition is obtained using estimates of 2 in which we employ the firm and worker effects, obtained respectively from the k-means clustering and the finite-mixture estimation. The measure of wage in this case is obtained by the estimated Mincer residual from equation 1.

(measured by $\text{var}(\theta_i - \bar{\theta}^j)$) accounts for the largest part of wage inequality. However, the contribution of worker heterogeneity between firm classes is also quite relevant. In particular, beyond firm heterogeneity (measured by $\text{var}(\psi_{j(i,t)})$), the figures highlights the importance of sorting (included in $\text{cov}(\bar{\theta}^j,\ \psi_{j(i,t)})$) and of the dispersion of average worker characteristics across firm classes (measured by $\text{var}(\bar{\theta}^j)$). The latter reveals the concentration of similar worker types in similar firm classes, a pattern that we may call 'worker segregation'. The fact that the estimated between-firm and within-firm components are remarkably stable throughout the period of observation testifies to the precision and robustness of the estimates.

4.3. Firm Classes and Worker Types

The firm classes and worker types derived from the BLM finite-mixture approach have only implicit connections to the underlying firm and worker characteristics. To shed light on the drivers of wage inequality, it is useful to try to make those connections more explicit by relating the firm classes and worker types to observable firm and worker characteristics. This provides further insight on the worker characteristics that explain within-firm inequality, the firm characteristics that explain between-firm inequality, and their possible association.

Figure 6: Between-within wage variance decomposition (Fixed effects) - Mincer



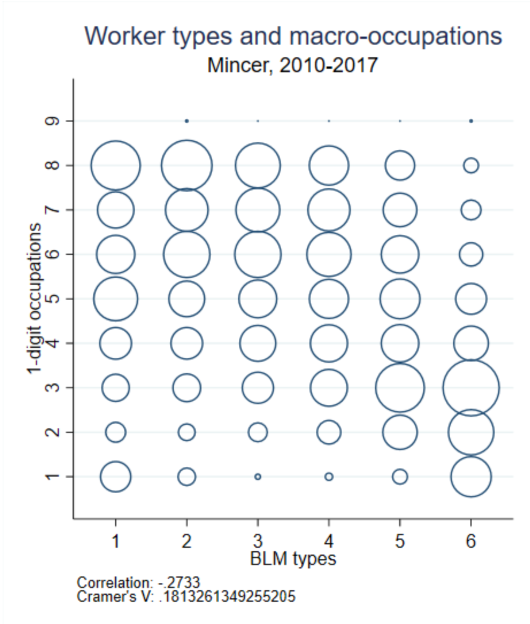
Note: The plot shows the wage variance decomposition following Equation 3 and using the estimated Mincer residual of the log weekly wage. "Worker dev" is worker heterogeneity within the same firm cluster and is measured as. var$(\theta_i - \bar{\theta}^j)$). "Worker avg" is aggregating workers of similar quality among the same employer, formally it is given by var$(\bar{\theta}^j)$, and we refer to it as worker segregation. The between and within components are estimated using the subsample built under Bonhomme et al. (2019) restrictions.

4.3.1. Characterizing Worker Types

We start by examining the relations between worker types and task complexity as measured by an occupation's relative intensity in non-routine versus routine intensive tasks. This requires matching the workers types with occupational categories. The latter are available in our dataset only from 2010 to 2017, hence we limit our matching exercise to that time period. The level of disaggregation is one-digit ISCO occupations, which entails ten occupational categories that we order from the most intensive in non-routine tasks to the most intensive in routine tasks.[12]. Figure 7 shows the share of worker types associated with each occupational category. The figure is based on Mincer wages: the equivalent with raw wages is presented in C.4 of Online Appendix C.2. The plot reveals that worker types commanding higher wages are associated with occupational categories characterized by lower routine task intensity. This applies no matter whether we consider raw wages or the Mincer residuals. We interpret these findings as suggestive evidence of vertical task specialization across firms, whereby

[12] One-digit ISCO occupation 1 corresponds to managers, 2 to professionals, 3 to technicians and associate professionals, 4 to clerical support workers, 5 to service and sales workers, 6 to skilled agricultural, forestry and fishery workers, 7 to craft and related trades workers, 8 to plant and machine operators and assemblers, 9 to elementary occupations, 10 to armed forces. With the notable exception of the last category (which we exclude), higher categories in the ordinal ISCO classification broadly correspond to occupations that are less complex in terms of their relative routine intensity.

Figure 7: Worker types - Occupational categories association- Mincer residual



Note: The plot shows the share of worker-year observation in type x associated to occupational category y, for $x \in \{1, \dots, 6\}$ and $y \in \{1, \dots, 9\}$. Worker types are estimated through finite-mixture method recursively every two years and using the Mincer residual. Occupations are the one-digit ISCO occupations, ten categories which are ordered from the most intensive in non-routine tasks to the most intensive in routine tasks.

firms requiring more complex tasks are more likely to hire workers possessing higher skills, and hence commanding higher wages. This provides another angle from which we can square the large contribution of the between-firm component with the large contribution of estimated worker types to the wage variance shown in Figures C.2 and 5.

We then look at other observable worker characteristics that are available in our dataset. In particular, we regress worker types on mover status, job description (six dummies for trainee, blue collar, white collar, manager or executive), age, weeks worked, and geographical macro area of employment. As there are potentially several concurring variables determining the worker type, we employ multinomial logit specifications. The estimates for each worker type are compared to type 3, which we therefore take as benchmark. Estimation is performed on the 2005-2008 four-year window. Table 1 shows the results for the Mincer residual. As usual the results for the equivalent under raw earnings are shown in table C.4 in Online Appendix C.2. While the link to some variables appear inconclusive, other relations are clearer. For instance, higher worker types (5 and 6) unequivocally have a higher probability of being employed in top position within the firm's organization (executives). Age as well as weeks worked are good predictors for the workers being assigned to higher types (5 or 6).[13] The finding on age can be explained in terms of human capital accumulation: as workers become more experienced, they are more likely to be hired by firms paying higher wages. The geographical distribution of workers is as expected given the economic development of the macro areas: workers of higher wage types are located mostly in the North.

4.3.2. Characterizing Firm Classes

In the case of firms, the dependent variables for the multinomial logit regressions are the firm classes, taking class 5 as benchmark. Estimates are again performed on the 2005-2008 four-year window. The regressors are employment, value added, local sales shares, poaching score and geographical macro area dummies. Following Bagger and Lentz (2019) the poaching score is computed as the four-year average of the ratio of new hires from other firms over total new hires.

Results are shown in Table 2 when firm clusters are obtained using Mincer residuals. As usual the equivalent one using raw wages (Table C.5) is shown in Online Appendix C.2. Results are intuitive. Firms in higher wage classes exhibit higher value added per worker ('labor productivity') and are less likely to be located in the less developed South. The fact that they are smaller in terms of employment is in line with growing evidence in the literature (see e.g. Autor et al. (2020)) that better performing firms employ fewer workers as they are more technologically advanced and pay higher wages due to larger sales shares. We do not find evidence that firms paying higher wages have also larger sales shares. This maybe due to the fact that we measure sales shares at the local level while better performing firms may be disproportionately active in the national and international markets.

Finally, the poaching score provides a measure of firms' labour market power or rent extraction. If firms compete more aggressively for workers, they are also willing to transfer larger rents to them. The relation between the poaching score and the firm classes is actually blurred: with raw earnings intermediate classes are more likely to poach workers; with the Mincer residual high firm classes poach more. We think that the latter is more reliable as the Mincer regression purges the residual from other worker covariates that may alter the role of firm labor market power. In this respect, the results in Table 2 point to the existence of some degree of monopsony in high

[13] The number of weeks worked tends to be positively associated with worker wage classes, except for the highest class.

Table 1: Multinomial logit - Worker Types on Observables

VARIABLES	(1) 1	(2) 2	(3) 3	(4) 4	(5) 5	(6) 6
mover	0.397***	0.0774***		0.0649***	0.405***	0.881***
	(0.0172)	(0.00913)		(0.0123)	(0.0164)	(0.0287)
trainee	1.091***	0.312***		-0.954***	-2.376***	-4.959***
	(0.0277)	(0.0132)		(0.0215)	(0.0556)	(0.294)
blue collar	0.462***	0.398***		-1.218***	-2.664***	-4.993***
	(0.0267)	(0.0106)		(0.0126)	(0.0166)	(0.0368)
white collar	-0.848***	-0.716***		0.170***	-0.115***	-1.605***
	(0.0264)	(0.0105)		(0.0124)	(0.0152)	(0.0240)
"quadro"	-0.741***	-2.961***		2.315***	3.532***	2.874***
	(0.0564)	(0.0452)		(0.0181)	(0.0194)	(0.0262)
executive	0.0705	-1.858***		2.164***	5.421***	7.920***
	(0.164)	(0.135)		(0.0828)	(0.0782)	(0.0800)
age	0.0400***	0.0141***		0.0151***	0.0402***	0.0600***
	(0.000423)	(0.000170)		(0.000231)	(0.000358)	(0.000803)
weeks	-0.0898***	-0.0273***		-0.00495***	-0.0266***	-0.0709***
	(0.000443)	(0.000281)		(0.000412)	(0.000600)	(0.00106)
Centre	0.0847**	-0.174***		0.555***	0.948***	1.344***
	(0.0339)	(0.0170)		(0.0191)	(0.0240)	(0.0403)
North-East	-0.275***	-0.446***		0.708***	1.216***	1.651***
	(0.0340)	(0.0170)		(0.0191)	(0.0240)	(0.0404)
North-West	-0.215***	-0.462***		0.728***	1.155***	1.526***
	(0.0337)	(0.0169)		(0.0190)	(0.0238)	(0.0399)
South	0.264***	0.101***		0.264***	0.644***	0.859***
	(0.0338)	(0.0170)		(0.0193)	(0.0247)	(0.0431)
Islands	0.223***	0.0909***		0.380***	0.705***	0.999***
	(0.0352)	(0.0175)		(0.0204)	(0.0268)	(0.0490)
Observations	86,704	1,279,224	877,061	451,237	199,316	64,852

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Multinomial logit for the period 2005-2008 where the worker types, estimated through finite-mixture models, are regressed on a set of observables, namely the status of mover, the qualification (six dummies for being employed as a trainee, blue collar, white collar, "quadro" or executive), age, weeks worked, and macroarea of employment. The reference category is the third worker type. Estimation algorithm for worker fixed effects employs Mincer residuals. Reference cluster is number 3.

firm types. While better performing firms pay higher wages, they still pass lower rent shares to their employees.

4.4. Decomposing Sorting in Horizontal and Vertical Specialization

Having established the presence of both vertical and horizontal specialization, it is interesting to assess their relative relevance for sorting. This can be achieved by further elaborating on the variance decomposition (3) and (4). By definition we have

$$\mathrm{var}(\theta_i) = \mathrm{var}(\theta_i - \bar{\theta}^j) + \mathrm{var}(\bar{\theta}^j) + 2\ \times \mathrm{cov}(\bar{\theta}^j,\, \theta_i - \bar{\theta}^j),$$

Table 2: Multinomial logit - Firm Clusters on Observables

VARIABLES	(1) 1	(2) 2	(3) 3	(4) 4	(5) 5	(6) 6	(7) 7	(8) 8	(9) 9	(10) 10
log_l	1.402***	0.727***	0.199***	0.356***		-0.349***	-0.473***	-1.180***	-1.243***	-2.176***
	(0.0491)	(0.0325)	(0.0344)	(0.0273)		(0.0324)	(0.0302)	(0.0353)	(0.0366)	(0.0457)
log_va	-1.921***	-1.182***	0.0779**	-0.635***		0.844***	0.682***	1.720***	1.717***	2.731***
	(0.0423)	(0.0288)	(0.0316)	(0.0243)		(0.0300)	(0.0276)	(0.0326)	(0.0338)	(0.0415)
(mean) poach_score	-1.444***	-0.743***	-0.324***	-0.239***		0.0630	0.179***	0.0950	-0.0183	-0.323***
	(0.0939)	(0.0571)	(0.0644)	(0.0475)		(0.0636)	(0.0558)	(0.0759)	(0.0778)	(0.111)
(mean) share_sales	0.267	0.342	-0.117	0.158		-0.851***	-0.303	-1.796***	-2.094***	-4.739***
	(0.327)	(0.215)	(0.238)	(0.195)		(0.248)	(0.238)	(0.294)	(0.323)	(0.457)
(p50) (p50) macro_area==Centro	0.0562	1.441***	0.645**	0.335		0.119	-0.437**	-0.232	-0.715***	-0.432*
	(0.457)	(0.449)	(0.253)	(0.241)		(0.207)	(0.203)	(0.211)	(0.213)	(0.254)
(p50) (p50) macro_area==Isole	1.174**	2.441***	1.257***	0.835***		0.0347	-0.703***	-0.529**	-1.328***	-0.823***
	(0.460)	(0.451)	(0.258)	(0.244)		(0.219)	(0.213)	(0.235)	(0.246)	(0.296)
(p50) (p50) macro_area==Nord-est	-0.719	0.515	0.391	-0.158		0.286	0.0560	0.0609	-0.102	-0.417*
	(0.458)	(0.449)	(0.252)	(0.241)		(0.205)	(0.202)	(0.209)	(0.210)	(0.252)
(p50) (p50) macro_area==Nord-ovest	-0.714	0.536	0.524**	-0.158		0.396*	0.0508	0.332	0.0937	0.371
	(0.458)	(0.449)	(0.252)	(0.241)		(0.205)	(0.202)	(0.209)	(0.209)	(0.249)
(p50) (p50) macro_area==Sud	1.573***	2.703***	1.320***	0.897***		0.00998	-0.922***	-0.786***	-1.292***	-1.345***
	(0.457)	(0.449)	(0.254)	(0.242)		(0.210)	(0.207)	(0.221)	(0.224)	(0.281)
Observations	5,236	20,651	8,999	28,833	25,999	8,993	13,350	5,797	5,600	3,060
Workforce	84,941	292,452	365,631	421,435	552,617	395,438	461,264	269,919	263,801	110,753

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Multinomial logit for the period 2005-2008 where the kmeans firm cluster is regressed on a set of observables, namely (log) size of workforce, (log) value added, share of sales on markets defined as province x 3-digit Ateco sector (the Italian classification for sectors), poaching score and macroarea dummies. The reference category is the fifth cluster. Firm clusters are estimated using the Mincer wages.

$$\text{cov}(\theta_i,\, \psi_{j(i,t)}) = \text{sd}(\theta_i)\text{sd}(\psi_{j(i,t)})\text{corr}(\theta_i,\, \psi_{j(i,t)})$$

and

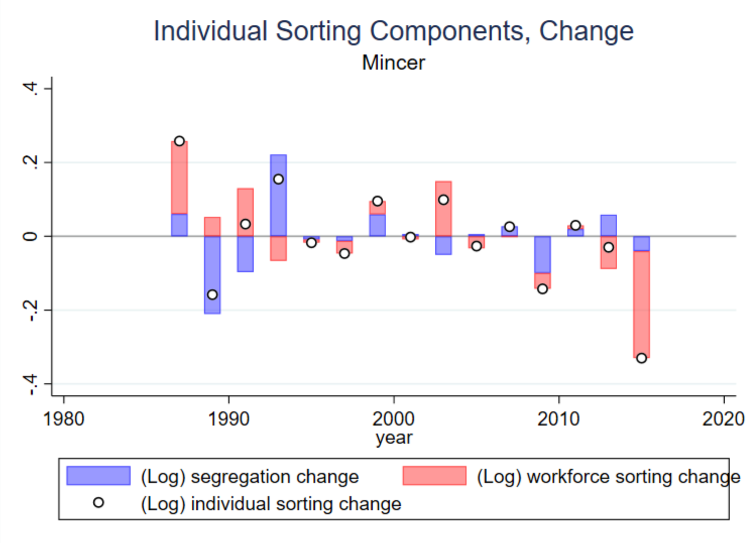
$$\text{cov}(\bar{\theta}^j,\, \psi_{j(i,t)}) = \text{sd}(\bar{\theta}^j)\text{sd}(\psi_{j(i,t)})\text{corr}(\bar{\theta}^j,\, \psi_{j(i,t)}).$$

Then, equating the right hand sides of those two variance decompositions gives

$$\text{corr}(\theta_i,\, \psi_{j(i,t)}) = \frac{\text{sd}(\bar{\theta}^j)}{\text{sd}(\theta_i)}\text{corr}(\bar{\theta}^j,\, \psi_{j(i,t)}) \tag{13}$$

where θ_i and $\bar{\theta}^j$ denote worker type i's individual characteristics and firm class j's average worker type characteristics. According to (13), (individual) sorting ($\text{corr}(\theta_i,\, \psi_{j(i,t)})$) can be decomposed in two components, segregation ($\text{sd}(\bar{\theta}^j)/\text{sd}(\theta_i)$) and workforce sorting ($\text{corr}(\bar{\theta}^j,\, \psi_{j(i,t)})$), which capture horizontal and vertical specialization respectively. With sorting superior firm classes tend to match with superior worker types. With segregation different worker types tend to cluster in different firm classes. As superior firm classes are associated with higher labor productivity and superior worker types are associated with more complex tasks, sorting implies vertical task specialization across

Figure 8: Segregation and workforce sorting log changes



Note: BLM estimates, Mincer residuals. Switching to Mincer residuals the role of the two components seems more balanced.

firms. However, the correlation between worker types and occupational categories also implies that the segregation of worker types in different firm classes entails the parallel segregation of occupations in those classes. This is consistent with assortativity between workers' specialized skills and firms' specific tasks, which can be seen as evidence of horizontal task specialization across firms. In particular, (13) features the following segregation index:[14]:

$$\text{segregation score} = \frac{\text{sd}(\bar{\theta}^j)}{\text{sd}(\theta_i)} \tag{14}$$

which measures the share of wage dispersion due to similar worker types working together in the same firm class.

Figure 8 depicts the evolution of the two (log) components of (individual) sorting focusing on Mincer wages for parsimony. It shows that both vertical specialization and horizontal specialization contribute to (individual) sorting with varying relative importance through time.

4.5. Robot Adoption, Sorting and Segregation

Having obtained robust and precise measures of sorting ($\text{corr}(\theta_i, \psi_{j(i,t)})$) and segregation ($\text{sd}(\bar{\theta}^j)/\text{sd}(\theta_i)$), we can now test whether and how robot adoption affects wage inequality through the recombination of firm classes and worker types. The results for sorting

[14] See also Song et al. (2018).

are obtained by estimating regression (12) and are reported in Table 3, whereby the top part shows results when including all industries and the bottom part shows them when excluding the automobile industry. As this industry accounts for a large portion of Italian manufacturing, its exclusion allows us to purge the overall patterns from specific characteristics of this sector that may possibly affect sorting beyond robot adoption. OLS and IV estimates are reported in Table 3 with and without the automobile industry and for the sorting estimates based on raw wages. Various robustness checks are reported in Online Appendix C.3. In the majority of cases, more so for the IV estimates, the coefficient is positive, significant and has a meaningful economic magnitude (as confirmed by our counterfactual reported further below). This is true also for the Mincer regression reported in Appendix C.3.

Table 3: Impact of Automated Robots on Sorting - With and Without Automobile Industry

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV2	(10) IV2	(11) IV2	(12) IV2
New Robots	0.0234***	0.00189	0.00355	0.00309	0.0182***	0.00620	0.00942**	0.0107***	0.00269	0.0107	0.0115	0.0167*
	(0.00341)	(0.00267)	(0.00276)	(0.00280)	(0.00434)	(0.00387)	(0.00396)	(0.00413)	(0.0110)	(0.00951)	(0.0101)	(0.0101)
Observations	937	937	937	937	937	937	937	937	937	937	937	937
R^2	0.281	0.570	0.596	0.634	0.279	0.570	0.595	0.632	0.256	0.567	0.594	0.627
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	No	No	No	No	No	No	No	No	No	No	No	No
A-R p value					2.52e-05	0.108	0.0160	0.0103	0.811	0.261	0.260	0.0862
M-P F stat					546.9	344.9	333.4	327.7	13.36	13.54	12.36	13.19
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Anderson-Rubin tests for joint null of orthogonality and non-significance of endogenous regressors.

F-stats above the critical value reject null of weak identification.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV2	(10) IV2	(11) IV2	(12) IV2
New Robots	0.0760***	0.114***	0.102***	0.101***	0.0772***	0.0859***	0.0802***	0.0854***	0.0679**	0.182***	0.141***	0.164***
	(0.0107)	(0.0204)	(0.0177)	(0.0176)	(0.0115)	(0.0212)	(0.0191)	(0.0182)	(0.0323)	(0.0559)	(0.0511)	(0.0449)
Observations	937	937	937	937	937	937	937	937	937	937	937	937
R^2	0.285	0.589	0.611	0.648	0.285	0.588	0.610	0.647	0.285	0.582	0.608	0.642
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	No	No	No	No	No	No	No	No	No	No	No	No
A-R p value					9.44e-11	9.44e-05	4.93e-05	1.01e-05	0.0520	0.00303	0.0114	0.00135
M-P F stat					4362	1890	1899	1747	219.1	168.1	154.8	162.7
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Anderson-Rubin tests for joint null of orthogonality and non-significance of endogenous regressors.

F-stats above the critical value reject null of weak identification.

Note: OLS and IV estimates of the impact of automated robots installations on sorting, computed at the province level and estimated through the finite-mixture model. Controls include three quartiles of two-digit sector HHI for employment shares - HHI_{25}, HHI_{50} and HHI_{75}. Covariates include the share of workforce employed in the manufacturing and construction macro-sectors. Finally, macro-area and two-year period fixed effects are included either separately or interacted. For the IV estimates we employ two shift-share instruments - in columns (5) to (8) we exploit robot information for Europe, US and Japan; in columns (9) to (12) we exclude Europe from the instrumental variable construction. The sample in the top table includes the automobile industry, the sample in the bottom table excludes the automobile industry

To support our argument that automation has affected both vertical and horizontal specialization, we now repeat the estimation by employing as dependent variable either the covariance term of decomposition (3) or our segregation measure, namely $\text{cov}(\bar{\theta}^j,\ \psi_{j(i,t)})$. On average across Italian provinces, the segregation score is about 18% (see Table C.3 in the Online Appendix), a value in line with the existing estimates for the US ranging from 16% to 25% across the years. Estimates for both specifications are shown, focusing on the more conservative sample that excludes the automobile industry, in Table 4. Online Appendix C.3 reports additional robustness checks. The top part in Table 4 shows results when the covariance is the dependent variable, the bottom part shows results for the care in which segregation is the dependent variable. The coefficient of interest is positive and significant. To appreciate the economic magnitude we run a counterfactual. First, for the case in which covariance is the dependent variable and under our preferred specification (column 8), if we were to bring a province from the 25th to the 75th percentiles of robot installations per worker, the covariance term would increase by one-third of its standard deviation. Next, for the case in which segregation is our dependent variable and under our preferred specification (column 8), if we were to bring a province from the 25th to the 75th percentiles of robot installations per worker, the segregation score would increase by 3.5 percentage points (half its interquartile range).

Overall, our findings support the conclusion that robot adoption increases wage inequality by fostering both horizontal and vertical task specialization across firms. On the one hand, in local economies where robot penetration has been more pronounced, workers performing similar tasks have disproportionately clustered in the same firms ('segregation'). On the other hand, such clustering has been characterized by the concentration of higher earners performing more complex tasks in firms paying higher wages ('sorting'). These firms are more productive and poach more aggressively.[15]

[15] One may be concerned that our estimated sorting captures not only production complementarities but also other potential drivers of geographical assortativity, such as local amenities or agglomeration externalities. In Appendix C.4 we discuss why this is unlikely to be the case.

Table 4: Impact of Automated Robots on Covariance and Segregation Excluding Automobile Industry

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV2	(10) IV2	(11) IV2	(12) IV2
New Robots, without automobile sector = L,	0.000866	0.0130***	0.0124***	0.0121***	0.00142	0.0104***	0.0101***	0.00986***	0.00549	0.0201***	0.0183***	0.0218***
	(0.00123)	(0.00285)	(0.00272)	(0.00288)	(0.00122)	(0.00296)	(0.00290)	(0.00294)	(0.00354)	(0.00726)	(0.00696)	(0.00648)
Observations	937	937	937	937	937	937	937	937	937	937	937	937
R^2	0.278	0.376	0.381	0.415	0.278	0.375	0.380	0.414	0.263	0.368	0.376	0.401
Period FEs	No	Yes	Yes	No	No	No	No	No	No	No	No	No
Macroarea FEs	No	No	Yes	No	No	No	No	No	No	No	No	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	No	No	No	No	No
Mincer	No	No	No	No	No	No	No	No	No	No	No	No
A-R p value					0.250	0.000746	0.000824	0.00154	0.135	0.00947	0.0143	0.00245
M-P F stat					4362	1890	1899	1747	219.1	168.1	154.8	162.7
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV2	(10) IV2	(11) IV2	(12) IV2
New Robots	-0.0180*	0.109***	0.100***	0.0928***	-0.0396***	0.0899***	0.0992***	0.0823***	-0.00125	0.241***	0.183***	0.153***
	(0.00982)	(0.0184)	(0.0148)	(0.0152)	(0.0104)	(0.0196)	(0.0171)	(0.0158)	(0.0274)	(0.0443)	(0.0353)	(0.0353)
Observations	937	937	937	937	937	937	937	937	937	937	937	937
R^2	0.354	0.466	0.584	0.650	0.350	0.465	0.584	0.650	0.352	0.432	0.571	0.644
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	No	No	No	No	No	No	No	No	No	No	No	No
First-stage F					4362	1890	1899	1747	219.1	168.1	154.8	162.7

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: OLS and IV estimates of the impact of automated robots installations on covariance (top table) and segregation (bottom table), cov($\bar{\theta}^j$, $\psi_{j(i,t)}$), computed at the province level. Controls include three quartiles of two-digit sector HHI for employment shares - HHI_{25}, HHI_{50} and HHI_{75}. Covariates include the share of workforce employed in the manufacturing and construction macro-sectors. Finally, macro-area and two-year period fixed effects are included either separately or interacted. For the IV estimates we employ two shift-share instruments - in columns (5) to (8) we exploit robot information for Europe, US and Japan; in columns (9) to (12) we exclude Europe from the instrumental variable construction.

5. A Search Model with Worker-Firm Complementarities

Our empirical findings can be rationalized through the lens of a dynamic search model in which within an occupation robot adoption strengthens the complementarities between firm classes and worker types. For the model to be consistent with the finite-mixture specification (5), its predicted cross-sectional distribution of wages should be expressible in terms of a joint distribution of worker characteristics, firm characteristics, and their correlation across firm-worker matches. Moreover, to be consistent with the identification of those characteristics through movers, the model has to predict firm-to-firm worker mobility. Finally, to generate within-firm-class wage dispersion for any given worker type, the model has to feature some randomness in the outcome of wage negotiation.

A canonical model with all these properties has been proposed by Cahuc et al. (2006), which extends previous work by Postel-Vinay and Robin (2002) to allow for a bargaining parameter that regulates the impact of the randomness on wage negotiation outcomes. The model generates this randomness through search frictions in the presence of two-sided heterogeneity, rent sharing, and firm-to-firm worker mobility due to on-the-job search and employee poaching. Unemployed workers negotiate with a single employer in a conventional way, but, when an employed worker receives an outside job offer, a three-player bargaining process starts between the worker, his original employer and the employer who has made the outside offer. The bargaining process follows an infinite-horizon alternating-offers bargaining game à la Rubinstein (1982), which links the share of the match surplus a worker obtains from negotiation to other search friction parameters. A firms offer a worker a wage that depends on his types. It can counter the offer the worker receives from another firm. If it does so, the firm makes take-it-or-leave-it counteroffers. If the worker's negotiation fails with both firms, he continues in his job at the preexisting terms. Wage contracts are long-term. They can be renegotiated only with mutual consent, which rules out wage cuts. There is no endogenous firing motive as an existent contract must be profitable and nothing can happen that may turn a profitable contract into an unprofitable one. Workers and firms have complete information. In particular, while matching is random, they know each other's types and classes. These are, therefore, unobserved only to the econometrician.

For our purposes, a shortcoming of the setup in Postel-Vinay and Robin (2002) and Cahuc et al. (2006) is that it predicts that the within-firm-class distribution of worker types in a given occupation is the same for all firm classes. In other words, there is neither sorting or segregation of worker types across firm classes that may affect the wage distribution. Hence, any observed variation in the within-firm-class distribution of worker types across firm classes has to be explained in terms of different occupational composition. This speaks to Section 4.3, where we discussed the relations of firm classes and worker types to observable firm and worker characteristics.

Different occupational composition is not the only possible explanation of segregation and sorting. As discussed by Postel-Vinay and Robin (2002) and Cahuc et al. (2006), their setup could be modified to deliver variation in the within-firm-class distribution of worker types across firm classes even abstracting from occupational composition. Appropriate modifications that could lead to segregation and sorting include removing the assumptions of constant returns to worker ability, scalar heterogeneity, and undirected search. We now show that the simplest modifications that would allow the canonical model to rationalize our empirical findings concern the marginal productivity of a firm-worker match and the arrival rate of outside offers to employed workers. Specifically, we let the former be log-supermodular with robot adoption increasing the degree of log-supermodularity.[16] Moreover, we endogenize the arrival rate of outside offers to employed workers by letting them choose their costly search intensity as in Postel-Vinay and Robin (2004). With these modifications, the canonical model predicts that the within-firm-class distribution of worker types in a given occupation exhibits higher density of higher x-types workers in higher y-class firms, and that robot adoption enhances such a positive assortative matching.

5.1. Environment

Consider the labor market for an occupation, in which an exogenous measure M of atomistic workers face a continuum of competitive firms, with mass normalized to 1, producing a single multipurpose good, with price also normalized to 1. Time is continuous. The market is populated by infinitely lived risk neutral workers and firms. Workers are heterogenous in their characteristics x, which are distributed according to a time-invariant distribution $g(x)$ with support $[\underline{x}, \overline{x}]$. Firms are heterogenous in their own characteristics y, which are distributed according to a time-invariant distribution $f(y)$ with support $[\underline{y}, \overline{y}]$. Hence, x and y correspond to the 'latent' worker types and firm classes of the empirical finite mixture model. For later use we also define the cumulative function $F(y)$ and its complement $\overline{F}(y) = 1 - F(y)$. Workers maximize utility and firms minimize costs intertemporally. Workers and firms make neither saving nor investment decisions, and spend all their earnings in every period. Both workers and firms discount the future at a rate ρ.

Due to search frictions, a worker can be matched with a firm ('employed') or unmatched ('unemployed'). In both cases the worker searches for new job offers. The Poisson

[16] Given that in the empirical part we have focused on the exogenous component of robot adoption, here we take robot adoption by all firms as exogenously given. In particular, we assume that technology adoption hits all firms at the same time as an unanticipated exogenous shock, to which firms and workers then react.

arrival rates of job offers are λ_o and $\lambda_1(x)$ for unemployed and employed workers respectively. They are strictly smaller than one due to search frictions, which gives labour market power to firms. While the arrival rate for the unemployed does not depend on worker type, the arrival rate for the employed is an increasing function of worker type: $\lambda_1'(x) \geq 0$. Higher x-type workers receive (weakly) more offers in any given interval of time. The limit case $\lambda_1'(x) = 0$ is the one assumed by Postel-Vinay and Robin (2002) and Cahuc et al. (2006).

We will show how $\lambda_1'(x) \geq 0$ can be derived from employed workers' endogenous choices of search intensity à la Postel-Vinay and Robin (2004) by assuming that the arrival rate of offers to on-the-job searchers is $\lambda_1(x) = \lambda_1 + \mu(x)$, where $\lambda_1 > 0$ is the arrival rate to passive workers and $\mu(x)$ is the endogenous intensity of an x-type worker's search activity faced with a search cost increasing and convex in the search effort: $c(\mu) > 0$, $c'(\mu) > 0$ and $c''(\mu) > 0$.

Firm-worker matches are separated exogenously at a Poisson rate δ, which replenishes the unemployed pool, or endogenously by poaching, which generates firm-to-firm worker flows.

A match produces a (flow) surplus every period determined by the match's marginal productivity in supplying the multipurpose good. As the price of this good is normalized to 1, a match's surplus and its marginal productivity coincide. We use $s(x, y)$ to denote the surplus of a match between a x-type worker and a y-class firm with $s(x, y) > 0$, $s_x(x, y) > 0$ and $s_y(x, y) > 0$. We capture complementarities between worker types and firm classes by assuming (weak) log-supermodularity: $s_{xy}(x, y) \geq 0$. A common functional form in the literature has constant elasticity of substitution:

$$s(x, y) = \left(x^{\xi} + y^{\xi}\right)^{\frac{2}{\xi}} \tag{15}$$

which is strictly log-supermodular (log-submodular) for $\xi < 0$ ($\xi > 0$), with its degree of modularity increasing with the absolute value of ξ. In the limit case of ξ going to 0, (15) converges to the Cobb-Douglas match marginal productivity in Postel-Vinay and Robin (2002) and Cahuc et al. (2006), $s(x, y) = xy$, which is neither log-supermodular nor log-submodular.[17] While we will use the functional form (15) as an example that facilitates the comparison with Postel-Vinay and Robin (2002) and Cahuc et al. (2006),

[17]The CES function (15) is assumed to be homogeneous of degree 2 to obtain exactly the Cobb-Douglas function used by Postel-Vinay and Robin (2002) and Cahuc et al. (2006) as ξ goes to 0.

the results that follow will be derived for any $s(x, y)$ satisfying the aforementioned properties. In particular, we will argue that the positive impact of robot adoption on sorting and segregation we have found in the data is consistent with robot adoption strengthening production complementarities between worker types and firm classes (which in (15) is captured by larger ξ).

5.2. Sharing Rules and Value Functions

Let $U(z) = z$ denote the instantaneous utility of a worker earning a flow of income z, and let $V(x, w, y)$ denote the lifetime utility of a x-type worker when employed at y-class firm at wage w with match surplus $s(x, y)$. Subsequent notation and derivations can be simplified by modelling unemployment as employment in a 'virtual firm' in 'virtual class' $b > 0$, paying an unemployed worker a benefit equal to the entire match surplus $s(x, b)$. The lifetime utility of an unemployed x-type worker x is then given by $V_0(x) \equiv V(x, s(x, b), b)$.

A y-class firm is able to hire an unemployed x-type worker only if the match is productive enough to at least compensate him for foregone unemployment income, which by (15) requires $y \geq b$. Therefore, for the worker to prefer employment in any firm class to unemployment, the lower bound of the support of the distribution of firm classes has to be no less than b, i.e. $\underline{y} \geq b$. Under this condition, any firm in any class is willing to hire any x-type unemployed worker when it meets him on the search market.

This implies that, whenever an unemployed worker of any type and a firm in any class meet, they sign a contract. The contract is based on an agreement on the wage reached through bargaining. As shown by Cahuc et al. (2006), the bargaining process delivers the generalized Nash-bargaining solution, in which the worker receives a constant share $\eta \in (0, 1)$ of the match rent. To at least compensate the worker for foregone unemployment income, the firms offers him the wage $\phi_0(x, y) \equiv \phi(x, b, y)$ that solves

$$V(x, \phi_0(x, y), y) = V_0(x) + \eta \left[V(x, s(x, y), y) - V_0(x)\right]$$

where $V(x, \phi_0(x, y), y)$ is the unemployed worker lifetime utility once hired, $V_0(x)$ is his threat point given by his lifetime utility as unemployed, and $V(x, s(x, y), y)$ is the maximum lifetime utility the worker would achieve by extracting the entire match surplus from the firm. Accordingly, the value function of an unemployed x-type worker can be stated as:

$$\rho V_0(x) = s(x,b) + \lambda_o \int_{y_{\inf}}^{\overline{y}} \eta \left[V(x, s(x,y'), y') - V_0(x) \right] dF(y')]$$

given that he receives offers at Poisson arrival rate λ_0 and the offerer's firm class y' cannot fall short of the value $y_{\inf}$ such that $V(x, s(x, y_{\inf}), y_{\inf}) = V_0(x)$. If there is free entry and exit of firms in the search market, then $y_{\inf} = \underline{y}$ holds, which we assume henceforth.

Turn now to employed workers and consider a x-type worker currently employed by a y-class firm. When the worker receives an offer from another firm in class y', he starts bargaining with both firms. There are three outcomes. A y'-class firm is able to poach the worker only if their match is productive enough to at least compensate him for foregone income at the current employer. If the poacher belongs to a higher class than the employer's ($y' > y$), competition between them can be seen as an auction where the bidder with the higher valuation wins and pays the second price. The winner is, therefore, the poacher which is forced by the auction to match the current employer's highest feasible bid $s(x, y)$. This is valued by the worker at $V(x, s(x, y), y)$, which becomes the fallback position for the negotiation game that the worker and the poacher subsequently play. As the firm has to at least compensate the worker for the foregone income implied by the employer's bid, the poacher offers him the wage $\phi(x, y, y')$ that solves:

$$V(x, \phi(x,y,y'), y') = V(x, s(x,y), y) + \eta \left[V(x, s(x,y'), y') - V(x, s(x,y), y) \right] \quad (16)$$

where $V(x, \phi(x, y, y'), y')$ is the worker lifetime utility once hired by the y'-class poacher, $V(x, s(x, y), y)$ is his threat point given by his lifetime utility at the current employer, and $V(x, s(x, y'), y')$ is the maximum lifetime utility the worker would achieve by extracting the entire match surplus from the poacher. Hence, when the worker meets a y'-class poacher, the first outcome is that, for $y < y' \leq \overline{y}$, the x-type worker moves to y' and is paid $\phi(x, y, y')$.

If the poacher does not belong to a higher class than the current employer's ($y' \leq y$), the winner of the auction is the current employer, which may or may not be forced by the auction to match the poacher's highest feasible bid $s(x, y')$. This is valued by the worker at $V(x, s(x, y'), y')$, which becomes the fallback position for the negotiation game that the worker and the current employer play. As the current employer has to at least compensate the worker for the foregone income implied by the poacher's bid, it offers him the wage $\phi(x, y', y)$ that solves

$$V(x, \phi(x, y', y), y') = V(x, s(x, y'), y') + \eta\left[V(x, s(x, y), y) - V(x, s(x, y'), y')\right] \quad (17)$$

where $V(x, \phi(x, y', y), y')$ is the worker lifetime utility if he chooses the current y-class employer to the y'-class poacher, $V(x, s(x, y'), y')$ is his threat point given by his lifetime utility at the poacher, and $V(x, s(x, y), y)$ is the maximum lifetime utility the worker would achieve by extracting the entire match surplus from the employer. The employer has to renegotiate with the worker only if the current wage w falls short of the poacher's offer, that is, only if $w < \phi(x, y', y)$ holds. As in Cahuc et al. (2006), the $\phi(x, y', y)$ that solves (17), is increasing in y' and thus there exists a unique threshold level for y' solving $w = \phi(x, y', y)$ such that $w < \phi(x, y', y)$ holds for $y' > q(x, w, y)$ and $w > \phi(x, y', y)$ holds for $y' < q(x, w, y)$. We use $q(x, w, y)$ to denote this threshold as the solution to $w = \phi(x, q, y)$. Hence, for $y < y' \leq \overline{y}$, there are two additional outcomes when the worker meets a y'-class poacher. For $q(x, w, y) < y' \leq y$, the worker stays with the current employer and his wage rises to $\phi(x, y', y)$, whereas, for $\underline{y} \leq y' \leq q(x, w, y)$, the worker stays with the current employer and his wage is unchanged.

Given the three possible outcomes, the value function of a x-type worker employed by a y-type firm at wage w can be stated as follows:

$$\begin{aligned}
\rho V(x, w, y) = {} & w - c(\mu) + \delta[V_0(x) - V(x, w, y)] \\
& + \lambda_1(x) \left[\int_{q(x,w,y)}^{y} \left[\eta V(x, s(x, y), y) + (1-\eta)\, V(x, s(x, y'), y')\right] dF(y') \right. \\
& + \int_{y}^{\overline{y}} \left[\eta V(x, s(x, y'), y') + (1-\eta)\, V(x, s(x, y), y)\right] dF(y') \\
& \left. - \int_{q(x,w,y)}^{\overline{y}} V(x, w, y) dF(y') \right]
\end{aligned} \quad (18)$$

as he receives an offer from a poacher at Poisson arrival rate $\lambda_1(x)$ incurring search cost $c(\mu)$, accepts the offer for $y < y' \leq \overline{y}$ with bargaining outcome (16), rejects the offer and stays with the current employer at higher wage for $q(x, w, y) < y' \leq y$ with renegotiation outcome (17), and stays with the current employer at the pre-existing wage for $\underline{y} \leq y' \leq q(x, w, y)$. The search intensity μ is endogenous and determined by the worker's optimal choice, which we now turn to.

5.3. Optimal Choice of Search Intensity

The optimal search intensity maximizes the worker's value function in equation (18). In Online Appendix D.1 we show how that equation can be rewritten as:

$$(\rho+\delta)V(x, s(x,y), y) = s(x,y) - c(\mu) + \delta V_0(x) + \lambda_1(x)\eta \int_y^{\overline{y}} \frac{s_y(x,p)\overline{F}(p)}{\rho+\delta+\lambda_1(x)\eta\overline{F}(y)} dp. \tag{19}$$

Recalling $\lambda_1(x) = \lambda_1 + \mu(x)$, the value function of the left hand side is maximized with respect to the search intensity for the value of $\mu(x)$ that satisfies:

$$c'(\mu(x)) \left[\rho+\delta+(\lambda_1+\mu(x))\,\eta\overline{F}(y)\right]^2 = (\rho+\delta)\,\eta \int_y^{\overline{y}} s_y(x,p)\overline{F}(p)dp. \tag{20}$$

The left hand side of (20) is an increasing function of μ as $c''(\mu) > 0$ holds due to the convexity of the search intensity cost. Moreover, its right hand side is an increasing function of x as $s_{xy}(x,y) > 0$ holds for the surplus to be log-supermodular. Hence, the optimal search intensity $\mu(x)$ and thus the implied arrival rate of offers $\lambda_1(x)$ are both increasing functions of the worker's type: $\mu'(x) > 0$ and thus $\lambda_1'(x) > 0$ as assumed so far.[18]

As for the impact of robot adoption on the optimal search intensity, by increasing the log-supermodularity of $s(x,y)$ robot adoption makes $s_{xy}(x,y)$ larger for any given x and y. It therefore makes $\mu(x)$ a steeper function of x. This implies an increase in the search intensity of higher x-type workers relative to lower x-type workers, which translates in a relative increase of the former workers' offer arrival rate. By putting relatively more effort into on-the-job search than before, higher x-type workers raise their relative probability of receiving offers, which allows them to climb the firm-class ladder relatively faster than before, raising in turn the relative probability of finding them in higher class firms. We expand on the implications for sorting and segregation in the next section.

[18] By (20) a worker's optimal search intensity depends not only on his type x but also on the wage w paid by his current employer and the employer's class y. We leave this dependence implicit for ease of notation.

5.4. Wage Determination with Complementarities

Following Cahuc et al. (2006), equation (18) can be solved to obtain the wage of a x-type worker who, after bargaining with two alternative firms y and y' with $y' < y$, has decided to work for the former. His contract is such that his wage equals the share of match surplus $s(x, y)$ he would appropriate from the y-class firm in the absence of on-the-job search (with threat point $s(x, y')$ given by the highest possible offer of the y'-class firm) plus the option value of differential income gains from mobility in the y-class firm with respect to the y'-firm:[19]

$$\phi(x, y', y) = s(x, y) - (1-\eta)\left(\int_{y'}^{y} \frac{\rho + \delta + \lambda_1(x)\overline{F}(p)}{\rho + \delta + \lambda_1(x)\eta\overline{F}(p)} s_y\,(x, p)\,dp\right) \tag{21}$$

The option implies the possibility that a worker accepts a lower wage today against higher expected wage increases in the future. As the worker is paid a wage resulting from a bargain between the worker and two alternative employers, the cross-sectional distribution of wages predicted by the model is driven by three aspects: the worker's type, the employer's class, and, because of search frictions, a statistical summary of the last random wage mobility the worker enjoyed.

An analogous expression applies to unemployed workers, for whom the wage is determined by competition between a y-class firm and a 'virtual' b-class employer with $b \leq \underline{y}$:

$$\phi_0(x, y) \equiv \phi(x, \underline{y}, y) = s(x, \underline{y}) - (1-\eta)\left(\int_{\underline{y}}^{y} \frac{\rho + \delta + \lambda_0\overline{F}(p)}{\rho + \delta + \lambda_0\eta\overline{F}(p)} s_y\,(x, p)\,dp\right) \tag{22}$$

Expression (21) is the theoretical counterpart of the finite-mixture specification (5).

Under the assumptions of Cahuc et al. (2006), we have Cobb-Douglas surplus $s(x, y) = xy$ (and thus $s_y(x, y) = x$) and offer arrival rate $\lambda_1(x) = \lambda_1$ independent of worker type. The logarithm of (21) becomes:

$$\ln \phi(x, y', y) = \ln x + \ln \phi(1, y', y)$$

The model thus predicts a log-linear decomposition of wages that separates the effect of unobserved worker characteristics on one side ($\ln x$) and the effect of unobserved firm characteristics as well as of recent labor market history on the other ($\ln \phi(1, y', y)$).

[19] See Online Appendix D.1 for derivation.

The effect of history is independent of the worker type because (as we will show in the next section) employers do not sort workers by their characteristics. In contrast, history is not independent of firm class as higher class firms suffer less from labor market competition. Firms' bargaining power is key for this feature. If firms had no bargaining power ($\eta = 1$), firm and worker effects would be separable because labor market history does not matter when the worker appropriates the whole surplus up front.

If the offer arrival rate is a function of worker type, the log of expression (21) is:

$$\ln \phi(x, y', y) = \ln x + \ln \left(y - (1-\eta) \left(\int_{y'}^{y} \frac{\rho + \delta + \lambda_1(x)\overline{F}(p)}{\rho + \delta + \lambda_1(x)\eta\overline{F}(p)} dp \right) \right)$$

Labor market history now depends on both firm class and worker type in a non-separable way. In particular, as higher worker types receive offers more often ($\lambda_1'(x) > 0$), they climb the firm-class ladder faster, and thus there is a higher probability of finding them in higher class firms (with $\eta < 1$ the ratio inside the integral is an increasing function of x). Hence, allowing the offer arrival rate to be higher for higher type workers generates sorting and segregation, which affect wage variance through the positive correlation of firm and worker characteristics. Firms' bargaining power is again key for this result. If the firm had no bargaining power ($\eta = 1$), we would be back to the separable case.

For $\lambda_1'(x) > 0$ and $s(x, y) = xy$, the effect of worker type cannot be separated from the effects of firm class and labor market history. Another reason why they cannot be separated is the presence of production complementarities between firm and work characteristics. Consider $\lambda_1(x) = \lambda_1$, $s(x, y) = \left(x^\xi + y^\xi\right)^{\frac{2}{\xi}}$ and thus $s_y(x, y) = 2y^{\xi-1}\left(x^\xi + y^\xi\right)^{-(\xi-2)/\xi}$). Expression (21) evaluates to:

$$\phi(x, y', y) = \left(x^\xi + y^\xi\right)^{\frac{1}{\xi}} - (1-\eta) \left(\int_{y'}^{y} \frac{\rho + \delta + \lambda_1\overline{F}(p)}{\rho + \delta + \lambda_1\eta\overline{F}(p)} 2y^{\xi-1}\left(x^\xi + y^\xi\right)^{-(\xi-2)/\xi} dp \right)$$

In this case sorting and segregation come from two sources. First, even in the limit case in which the firm has no bargaining power ($\eta = 1$) and the worker extracts all surplus up front, the surplus function transmits its log-supermodularity to the wage. Second, as log-supermodularity $s_{xy}(x, y) \geq 0$ implies that higher type workers receive disproportionately higher offers, these workers climb the firm-class ladder faster, and thus they are more likely to be found in higher class firms (the term $s_y(x, y) =$

$2y^{\xi-1}\left(x^{\xi}+y^{\xi}\right)^{-(\xi-2)/\xi}$ inside the integral is an increasing function of x). Hence, the positive impact of robot adoption on sorting and segregation we have found in the data is consistent with robot adoption strengthening production complementarities between worker types and firm classes when higher type workers are more likely to receive job offers due to higher endogenous on-the-job search intensity. The stronger those complementarities are, the higher the relative importance of between versus within firm-class wage dispersion.

5.5. Wage Distribution with Complementarities

The wage distribution can be derived from the steady state equilibrium flow conditions. We start by defining a few variables that enter the definition of the flows. We use $G(w \mid x, y)$ to denote the share of x-type workers employed by y-class firms who are paid a wage no higher than w. We define u as the share of unemployed workers, $(1-u)$ as the share of employed workers searching on the job, and $\ell(x, y)$ as the density of x -type workers employed by y-class firms. Due to sorting, this density is not necessarily given by the product of the population densities of worker types $g(x)$ and firm classes $f(y)$.

In steady state the outflows and the inflows of workers in the different worker pools have to balance. These worker pools are the pool of unemployed workers and several pools of employed workers that differ in terms of employee type, employer class and wage, given that on-the-job search allows two identical workers hired by the same firm to be paid differently. In each period a share δ of firm-worker matches $(1-u)M$ is destroyed and a share λ_o of the unemployed uM finds a job. Hence, for the unemployed pool to be in steady state, we need $\lambda_o u = \delta(1-u)$. As for the employed pools, consider the pool of x-type workers employed by y-class firms at wage no higher than w. On the outflow side, a share $G(w \mid x, y)\ell(x, y)(1-u)$ of them exits either because they are laid off (at separation rate δ) or because they obtain a wage rise. The latter event happens to a share $\lambda_1(x)\overline{F}(q(x, w, y))$ of workers, as a share $\lambda_1(x)$ of them receives an offer and a share $\overline{F}(q(x, w, y))$ of those receiving an offer either leaves to another firm or negotiates a wage rise with the current employer. On the inflow side, a share $g(x)$ of unemployed workers are of type x. A share λ_0 of them receives an offer, which they accept, and a share $f(y)$ of the offers comes from y-class firms. In addition, some workers are poached from other firms. A share $f(y)$ of these firms is in class y, a share $\lambda_1(x)$ of their employees receives an offer. The share of x-type workers accepting offers from y-class firms is $\int_{\underline{y}}^{q(x,w,y)} \ell(x,p)dp$. Hence, for the outflow to equal the inflow the following condition has to hold:

$$\left(\delta + \lambda_1(x)\overline{F}(q(x,w,y))\right)G(w \mid x,y)\ell(x,y) = \left[\delta g(x) + \lambda_1(x)\left(\int_{\underline{y}}^{q(x,w,y)} \ell(x,p)dp\right)\right]f(y) \tag{23}$$

where we have imposed also balanced flows for the unemployed: $\lambda_o u = \delta(1-u)$. Condition (23) applies to all x-type workers matched with y-class firms at wage lower or equal to w. Consider now the subset (x,w,y) consisting of workers just at the margin of those who do not get a pay rise. A worker in this subset is an employee of a y-type firm who receives an offer from a poaching firm in class $y' = q$. As already discussed, competition between the current employer and the poacher implies that the current employer has to pay as wage the entire surplus that the employee would generate if hired by the poacher: $w = \phi(x,q,y) = \phi(x,y',y) = s(x,y')$. If the poacher is itself a y-firm, then $y' = y$ implies $w = s(x,y)$, and thus also $G(w = s(x,y) \mid x,y) = 1$ given that no worker can be paid more than the match surplus. In other words, Bertrand competition for a x-type worker between two y-class firms transfers the entire match surplus from the employer to the employee as wage payment. In this case, condition (23) simplifies to:

$$\left(\delta + \lambda_1(x)\overline{F}(y))\right)\ell(x,y) = \left[\delta g(x) + \lambda_1(x)\left(\int_{\underline{y}}^{y} \ell(x,p)dp\right)\right]f(y)$$

which, after integration by parts, can be solved for the density of x-type workers employed by y-class firms:[20]

$$\ell(x,y) = \frac{1 + \frac{\lambda_1(x)}{\delta}}{\left(1 + \frac{\lambda_1(x)}{\delta}\overline{F}(y)\right)^2} g(x)f(y) \tag{24}$$

Integrating (24) gives the corresponding cumulative density:

$$L(x,y) = \int_{\underline{y}}^{y} \ell(x,p)dp = \frac{F(y)}{1 + \frac{\lambda_1(x)}{\delta}\overline{F}(y)} \tag{25}$$

[20]See Online Appendix D.2 for further details.

Using (24) and (25) allows us to solve (23) for the share of x-type workers employed by y-class firms who are paid a wage no higher than w:

$$G(w \mid x, y) = \left(\frac{1 + \frac{\lambda_1(x)}{\delta} \overline{F}(y)}{1 + \frac{\lambda_1(x)}{\delta} \overline{F}(q(x, w, y))} \right)^2 \tag{26}$$

with $w \in [\phi_0(x, y), s(x, y)]$. The threshold $q(x, w, y)$ is an increasing function of w.[21] This implies that, as w rises, $G(w \mid x, y)$ also rises, reaching 1 for $w = s(x, y)$ as firms cannot pay workers more than match surplus. Again, $q(x, w, y) = y$ holds for $w = s(x, y)$. The following remarks are in order. On the one hand, with $\lambda_1(x) = \lambda_1$, (24) becomes:

$$\ell(x, y) = g(x)\ell(y)$$

where: $\ell(y) = \frac{1+\frac{\lambda_1}{\delta}}{\left(1+\frac{\lambda_1}{\delta}\overline{F}(y)\right)^2} f(y)$ is the density of workers in y-type firms. In this case, the density $\ell(x, y)$ of x-type workers employed by y-class firms can be decomposed into two multiplicative components, one of which is a function of x only while the other is a function of y only. Accordingly, the composition of employment in terms of worker types is the same for all firm classes. In other words, within an occupation there is no sorting of worker types across firm classes. This result holds for any functional form of match surplus $s(x, y)$. Therefore, it is unaffected by production complementarities. This does not mean, however, that also $G(w \mid x, y)$ is unaffected. Stronger complementarities increase $q(x, w, y)$ as well as both $\phi_0(x, y)$ and $s(x, y)$, but the latter more than the former, fostering within-firm-class wage dispersion despite unchanged employment composition.[22]

On the other hand, when $\lambda_1(x)$ depends on worker type, (24) is not separable into worker-type and firm-class components. In this case, $\lambda_1'(x) > 0$ generates sorting and segregation: in higher firm classes the composition of employment is skewed towards higher worker types. The more so, the larger $\lambda_1'(x)$ is and the stronger the complementaries. The more so, the stronger the complementarities as these also make $\lambda_1'(x)$ larger. In this case, stronger complementarities together with the implied increased search intensity of higher worker types promote both within- and between-

[21] This comes from the definition of $q(x, w, y)$ as the solution to $\phi(, q, p) = w$ and the fact that the $\phi(x, y', y)$ solution is increasing in x and y' (but not necessarily in y). See Cahuc et al. (2006) for details.

[22] In the case of $\eta = 0$, even though the three components of the steady-state earnings distribution are not independent, Postel-Vinay and Robin (2002) decompose the variance of wages into three separate components. This is achieved by allocating the total within-firm variance of wages not explained by worker type to market frictions and all the between-firm variance to the firm class.

firm-class wage dispersion also through the change of employment composition across firm classes.

6. Conclusion

Leveraging the geographic dimension of a large administrative panel on employer-employee contracts, we have studied the impact of robot adoption on wage inequality through changes in worker-firm assortativity. Using recently developed methods to correctly and robustly identify firm unobserved characteristics, firm unobserved characteristics and their combination, we have linked worker-firm sorting and worker segregation across firms to robot adoption across local economies.

We have found that robot adoption increases wage inequality by fostering both horizontal and vertical task specialization across firms. In local economies where robot penetration has been more pronounced, workers performing similar tasks have disproportionately clustered in the same firms ('segregation'). Moreover, such clustering has been characterized by the concentration of higher earners performing more complex tasks in firms paying higher wages ('sorting'). These firms are more productive and poach more aggressively.

We have rationalized these findings through a simple extension of a well-established class of models with two-sided heterogeneity, on the job search and rent sharing through a generalized Nash bargaining process under Bertrand competition in employee poaching and workers' endogenous search intensity.

References

[1] Abowd, John M, Francis Kramarz, and David N Margolis, "High wage workers and high wage firms," *Econometrica*, 1999, *67* (2), 251–333.

[2] Acemoglu, Daron and Pascual Restrepo, "Robots and jobs: Evidence from US labor markets," *Journal of Political Economy*, 2020, *128* (6), 2188–2244.

[3] Addario, Sabrina Di, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten, "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages," *Journal of Econometrics*, 2022.

[4] Andrews, Martyn J, Leonard Gill, Thorsten Schank, and Richard Upward, "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Economics Letters*, 2012, *117* (3), 824–827.

[5] Autor, D. H., Lawrence F Katz, and Melissa S Kearney, "The polarization of the US labor market," *American economic review*, 2006, *96* (2), 189–194.

[6] Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

[7] Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics*, 2003, *118* (4), 1279–1333.

[8] Azar, José, Ioana Marinescu, and Marshall Steinbaum, "Labor market concentration," *Journal of Human Resources*, 2020, pp. 1218–9914R1.

[9] ___ , ___ , ___ , and Bledi Taska, "Concentration in US labor markets: Evidence from online vacancy data," *Labour Economics*, 2020, *66*, 101886.

[10] Bagger, Jesper and Rasmus Lentz, "An empirical model of wage dispersion with sorting," *The Review of Economic Studies*, 2019, *86* (1), 153–190.

[11] Barth, Erling, Alex Bryson, James C Davis, and Richard Freeman, "It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States," *Journal of Labor Economics*, 2016, *34* (S2), S67–S97.

[12] Boeri, Tito, Andrea Ichino, Enrico Moretti, and Johanna Posch, "Wage equalization and regional misallocation: evidence from Italian and German provinces," *Journal of the European Economic Association*, 2021, *19* (6), 3249–3292.

[13] Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "A distributional framework for matched employer employee data," *Econometrica*, 2019, *87* (3), 699–739.

[14] Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-Experimental Shift-Share Research Designs," *Review of Economic Studies*, 2022, *89* (1), 181–213.

[15] 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.

[16] Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

[17] ___ , Jörg Heining, and Patrick Kline, "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 2013, *128* (3), 967–1015.

[18] Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 1980, *47* (1), 225–238.

[19] Cortes, Guido Matias, Adrian Lerche, Uta Schönberg, and Jeanne Tschopp, "Technological Change, Firm Heterogeneity and Wage Inequality," Technical Report, Mimeo (https://www. dropbox. com/s/jcpm28binsgqhbp/Draft_v1_0. pdf 2020.

[20] Dorn, David, Lawrence F Katz, Christina Patterson, John Van Reenen et al., "Concentrating on the Fall of the Labor Share," *American Economic Review*, 2017, *107* (5), 180–85.

[21] Eeckhout, Jan and Philipp Kircher, "Identifying sorting—in theory," *The Review of Economic Studies*, 2011, *78* (3), 872–906.

[22] Elsby, Michael, Axel Gottfries, Ryan Michaels, and David Ratner, "Vacancy Chains," 2022.

[23] Elsby, Michael WL and Axel Gottfries, "Firm dynamics, on-the-job search, and labor market fluctuations," *The Review of Economic Studies*, 2022, *89* (3), 1370–1419.

[24] Faia, Ester, Sébastien Laffitte, Max Mayer, and Gianmarco IP Ottaviano, "Automation, globalization and vanishing jobs: a labor market sorting view," 2020.

[25] Helpman, Elhanan, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J Redding, "Trade and inequality: From theory to estimation," *The Review of Economic Studies*, 2017, *84* (1), 357–405.

[26] Hémous, David and Morten Olsen, "The rise of the machines: Automation, horizontal innovation, and income inequality," *American Economic Journal: Macroeconomics*, 2022, *14* (1), 179–223.

[27] Jaimovich, Nir, Itay Saporta-Eksten, Henry Siu, and Yaniv Yedid-Levi, "The macroeconomics of automation: Data, theory, and policy analysis," *Journal of Monetary Economics*, 2021, *122*, 1–16.

[28] Jochmans, Koen and Martin Weidner, "Fixed-Effect Regressions on Network Data," *Econometrica*, 2019, *87* (5), 1543–1560.

[29] Katz, Lawrence F and Kevin M Murphy, "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 1992, *107* (1), 35–78.

[30] Lachowska, Marta, Alexandre Mas, Raffaele D Saggio, and Stephen A Woodbury, "Do firm effects drift? Evidence from Washington administrative data," Technical Report, National Bureau of Economic Research 2020.

[31] Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium wage dispersion with worker and employer heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

[32] ___ and ___ , "To match or not to match? Optimal wage policy with endogenous worker search intensity," *Review of Economic Dynamics*, 2004, *7* (2), 297–330.

[33] Rubinstein, Ariel, "Perfect equilibrium in a bargaining model," *Econometrica: Journal of the Econometric Society*, 1982, pp. 97–109.

[34] Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality*," *The Quarterly Journal of Economics*, 2018, *134* (1), 1–50.

Robot Adoption, Worker-Firm Sorting and Wage Inequality: Evidence from Administrative Panel Data

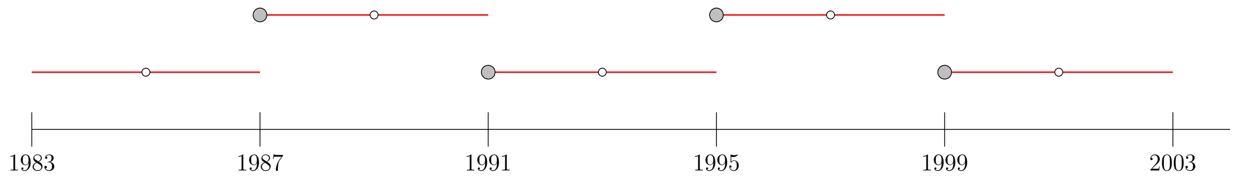
Ester Faia, Goethe University Frankfurt and CEPR
Gianmarco Ottaviano, Bocconi University and CEPR
Saverio Spinella, New York University

ONLINE APPENDIX

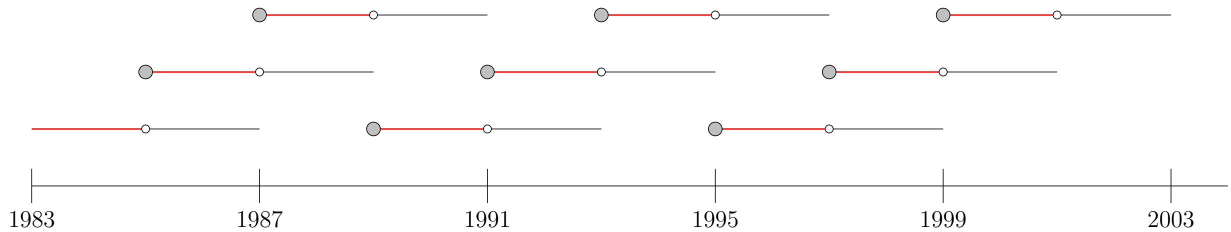
A. Overlapping intervals - Rolling Windows

To increase the number of estimated sorting parameters we employ rolling windows. Figure A.1 gives a graphic representation of our estimation intervals and compares them to the ones envisaged in Bonhomme et al. (2019). In each interval we keep only workers who have a contract each year and firms that survive as well for the whole period. This restricts considerably the sample size. Furthermore, as in Bonhomme et al. (2019) movers are allowed to change employer only between the second and third year (signalled by white dots). Every time we define a new interval and consider the possibility of new entrants in the labor market. Hence, we reclassify workers and firms again, so as to consider those who survive in the following four years. This operation encompasses a larger quantity of job movements, indicated by a gray dot. If we were to use non-overlapping intervals, sorting would be computed on the whole 4-year interval (the red portion of each segment). Our rolling windows, as exemplified in Panel (b), accommodate larger job movements every two years. Sorting is then computed only for the first two of each interval (once again, the red portion of each segment). This allows us to have a clear measure of sorting which is frequently updated, and informed by sizeable job movements and recent k-means clusterization (which we implement at the start of each period).

Figure A.1: Estimation Intervals



(a) Intervals following Bonhomme et al. (2019)



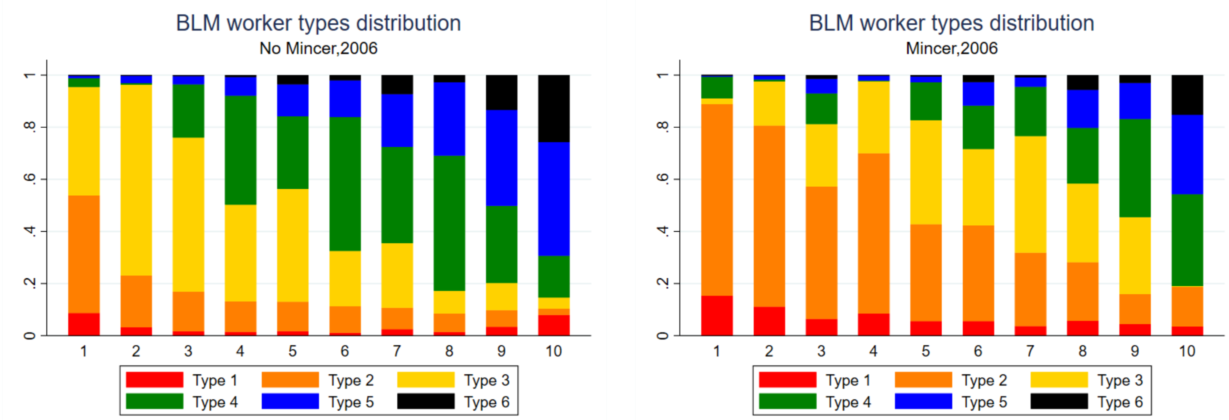
(b) Intervals using Rolling-Windows

Note: Representation of the rolling estimation windows. Movers are allowed to change employer only between the second and third year (signalled by white dots). Every time a new interval starts. Workers and firms are then reclassified to consider those who survive in the following four years.

B. Algorithm estimates - Other results

Workforce composition. The finite-mixture estimation algorithm delivers, prior to the Bayesian assignment, also the composition of the workforce for each firm cluster. This represents the prior that inserted in the Bayes' rule 9. Figure B.1 plots worker compositions for the interval 2005-2008. The plot reveals sorting of workers in high wage types to firms in high wage classes.

Figure B.1: Workforce Composition, Mincer



(a) No Mincer

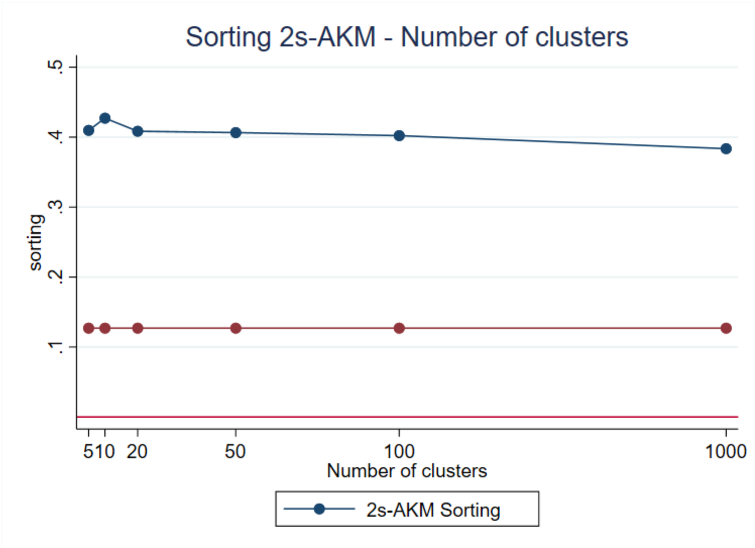
(b) Mincer

Note: The figure plots for each firm cluster the relative representation of each worker type in the workforce. Types and clusters are estimated using raw earnings (panel a) and Mincer wages (panel b)

B.1. Firm clusters

The exercise in 3 is a powerful display of the strength of the clustering approach. Here we show that, using the full sample, the choice of the number of clusters does not matter for the gap between sorting computed by baseline AKM and our two-step alternative. Indeed, it remains fairly stable even upon using one thousand clusters, confirming that the approach - by addressing the issue of low network connectedness - is able to uncover the true measure of sorting.
Clearly, one can also sense from Figure B.2 that, as the number of clusters approaches the number of individual firms populating the network, sorting measures will clearly converge - what is interesting, it is the slow pace of such convergence.Having established that the estimates are not over-reliant on the choice of K, one can qualify each of the ten baseline firm clusters looking at the unconditional-mean table in B.1.

Figure B.2: Testing different numbers of kmeans clusters



Note: Sorting measures for two-step AKM using K =5,10,20,50,100,1000 clusters are plotted against the baseline AKM estimate.

Table B.1: Firm clusters - No Mincer

	clus_recap clus1	clus2	clus3	clus4	clus5	clus6	clus7	clus8	clus9	clus10
mean_wage	5.223	5.8	6.027	6.161	6.317	6.44	6.475	6.644	6.84	7.178
y_25	4.625	5.697	5.928	6.029	6.17	6.176	6.322	6.422	6.553	6.853
median_wage	5.452	5.816	6.03	6.151	6.284	6.353	6.448	6.595	6.801	7.162
y_75	5.908	5.926	6.129	6.277	6.422	6.627	6.591	6.808	7.07	7.493
poach_score	.38	.452	.477	.51	.515	.552	.6	.62	.588	.598
lab_intensity	.408	.496	.443	.403	.308	.207	.298	.226	.216	.243
va_worker	64.389	70.282	110.255	130.947	119.967	164.893	149.194	208.672	349.684	433.171
va_sales	.489	.562	.51	.47	.385	.287	.384	.37	.324	.314
share_intangible	.249	.23	.254	.262	.169	.237	.151	.255	.279	.345
size	12	13.5	14.9	21.2	28	32.6	39.3	63.3	52.5	35.9
trainee	.026	.043	.036	.034	.023	.029	.013	.017	.022	.009
bluecollar	.799	.847	.801	.73	.585	.491	.667	.391	.205	.065
whitecollar	.168	.1	.153	.213	.353	.393	.274	.473	.555	.48
quadro	.002	.001	.002	.007	.023	.049	.03	.077	.147	.295
executive	.001	0	.001	.002	.005	.022	.008	.026	.06	.142
manuf	.161	.247	.267	.338	.474	.592	.616	.663	.59	.43
constr	.045	.067	.074	.055	.044	.023	.033	.019	.028	.023
serv	.794	.686	.658	.606	.482	.386	.351	.317	.382	.547
age	36.4	38.8	39.1	39.4	41.8	41	42.9	42.9	42.5	43.2
northwest	.29	.21	.23	.31	.35	.46	.4	.44	.46	.55
northeast	.14	.14	.19	.25	.29	.31	.3	.3	.24	.11
center	.21	.21	.22	.21	.19	.14	.16	.16	.17	.26
south	.28	.33	.26	.17	.12	.07	.11	.07	.08	.05
islands	.08	.1	.11	.07	.05	.02	.03	.04	.04	.04
share_empl	.006	.068	.138	.163	.172	.105	.117	.146	.068	.016
n_firms	1562	15449	28416	23395	18806	9800	9113	7021	3975	1389
n_workers	18699	208046	422146	496825	526604	319683	358326	444303	208681	49814

Note: Summarizing several variables characterizing the ten k-means firm clusters estimated in the 2005-2008 interval.

B.2. Worker types

Table B.2: Worker types - No Mincer

	type_recap					
	type1	type2	type3	type4	type5	type6
wage	5.84175	5.982744	5.972135	6.262899	6.643395	7.194921
trainee	.113	.084	.075	.049	.014	.002
bluecollar	.75	.711	.592	.578	.27	.063
whitecollar	.124	.184	.101	.344	.627	.401
quadro	0	0	.03	.001	.068	.373
executive	0	0	.195	0	.001	.151
occ1	.001	.002	.161	.004	.022	.157
occ2	.019	.029	.053	.066	.185	.276
occ3	.066	.088	.074	.154	.311	.331
occ4	.118	.136	.094	.144	.146	.119
occ5	.132	.173	.186	.122	.096	.055
occ6	.242	.205	.13	.19	.121	.036
occ7	.131	.172	.089	.208	.081	.018
occ8	.292	.196	.213	.112	.038	.008
manuf	.331	.33	.301	.465	.498	.507
constr	.062	.066	.058	.058	.048	.033
serv	.606	.603	.641	.477	.454	.461
age	35.735	36.312	38.995	37.364	40.026	43.475
northwest	.269	.26	.35	.356	.434	.489
northeast	.197	.205	.175	.277	.282	.258
center	.198	.208	.191	.18	.166	.167
south	.254	.237	.223	.131	.082	.063
islands	.083	.091	.061	.056	.035	.023
n	190397	391889	22275	395429	122180	45090
share	.1631145	.3357341	.0190832	.3387668	.1046725	.0386289

Note: Summarizing several variables characterizing the six worker grouped random effect estimated in the 2015-2018 interval.

C. Last stage - Additional Results

C.1. Descriptive statistics

Table C.1: Shift-share Robot Installations measure

	n obs	mean	var	p10	p25	p50	p75	p90
All sectors	1044	.6091146	.6139619	.0268438	.1690594	.4167585	.754465	1.295757
Without auto	1044	.2787364	.0664967	.0200644	.0466083	.2485207	.4219752	.584039
Only auto	984	.3505233	.566917	-.0415398	.0094493	.0864324	.3798713	.9880844

Note: Summarizing our alternative measures of automatization. More than half of all robot installations come from the automobile industry; the relative measure also exhibits substantial variance, due to the widely different employment shares in such sectors across provinces.

Table C.2: Covariance measure

	n obs	mean	var	p10	p25	p50	p75	p90
No Mincer	1056	.0254638	.0001016	.0152654	.0192701	.0239352	.0303524	.0376067
Mincer	1056	.0184392	.0000493	.011137	.0141712	.0174062	.0215385	.0274435

Note: Summarizing our alternative measures of covariance, computed as twice the correlation between the firm and worker fixed effects times the respective standard deviations.

Table C.3: Segregation score measure

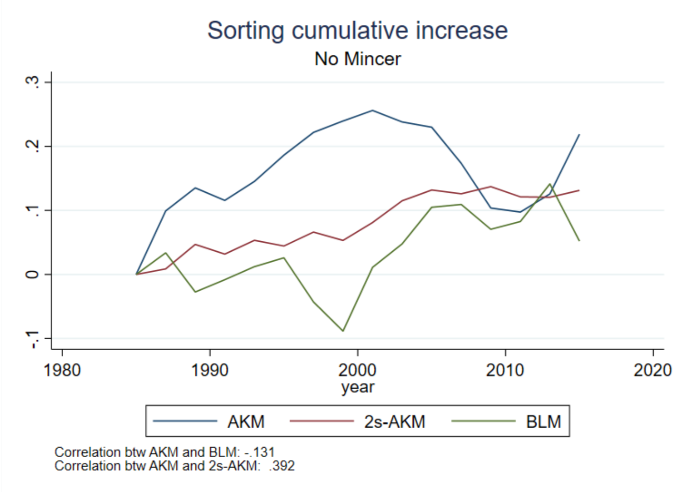
	n obs	mean	var	p10	p25	p50	p75	p90
No Mincer	1056	.4775343	.0083294	.3763741	.4147443	.4647163	.5293838	.5930381
Mincer	1056	.4163355	.0070687	.3325594	.3617232	.4035098	.4512173	.5118459

Note: Summarizing our alternative measures of segregation score, computed as the ratio between the variance of the firm's average worker fixed effect and the total variance of the worker fixed effect.

C.2. Other Figures and Tables

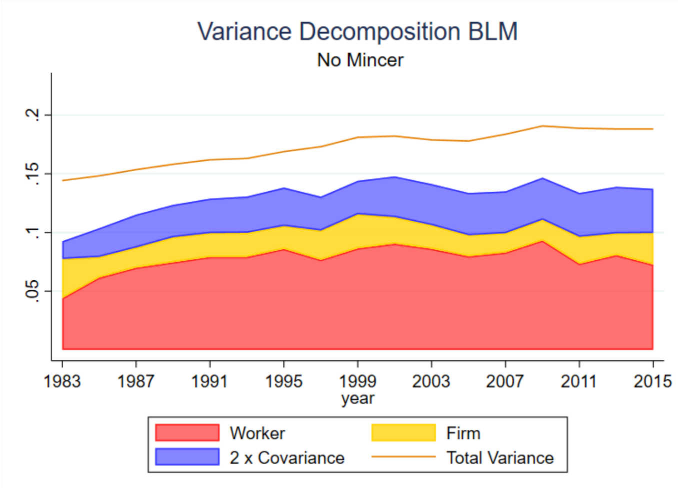
Figure C.1 shows the differences in the cumulative increase in sorting across the different estimation methods, namely Abowd et al. (1999), two-step AKM, and Bonhomme et al. (2019).
Figure C.2 shows the variance decomposition between the firm and worker fixed effects in the Bonhomme et al. (2019) and using raw wages instead of Mincer.
Figure C.3 plots the between-within wage variance decomposition based on Equation 3 and using raw wages.
Figure C.4 shows the share of worker-year observation in type x associated with occupational category y and using the raw wages.

Figure C.1: Cumulative sorting increase



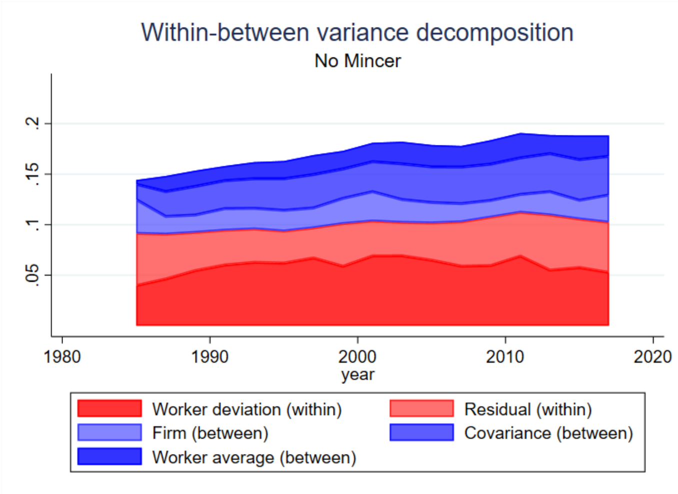
Note: Cumulative sorting increase estimated through Abowd et al. (1999), two-step AKM, and Bonhomme et al. (2019). The first two algorithms run on the same sample of workers, and are thus directly comparable. The latter algorithm imposes some restrictions in the pattern of job movements allowed in the sample. We estimate sorting on 4-year, partially overlapping intervals. The sorting is calculated in the first two year of each interval.

Figure C.2: Baseline wage variance decomposition - No Mincer



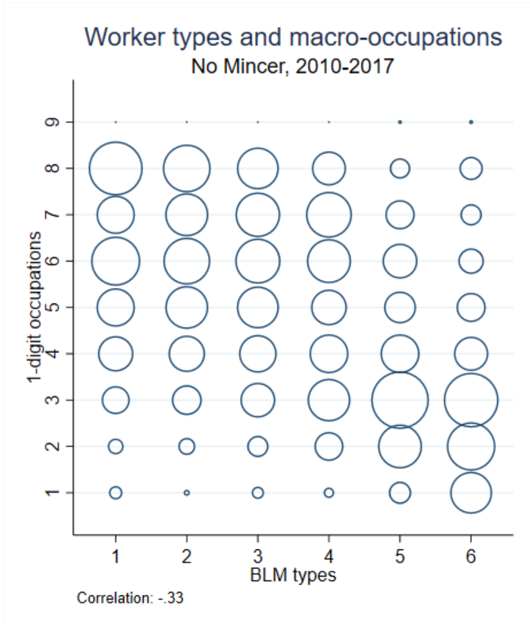
Note: The variance decomposition is obtained using estimates of 2 in which we employ the firm and worker effects, obtained respectively from the k-means clustering and the finite-mixture estimation. The measure of wage in this case is raw earnings.

Figure C.3: Between-within wage variance decomposition (Fixed effects) - No Mincer



Note: The plot shows the log weekly wage variance decomposition based on Equation 3. "Worker dev" is worker heterogeneity within the same firm cluster and is measured as. $\mathrm{var}(\theta_i - \bar{\theta}^j)$). "Worker avg" is aggregating workers of similar quality among the same employer, formally it is given by $\mathrm{var}(\bar{\theta}^j)$, and we refer to it as worker segregation. The between and within components are estimated using the subsample built under Bonhomme et al. (2019) restrictions.

Figure C.4: Worker types - Occupational categories association - Raw Earnings



Note: The plot shows the share of worker-year observation in type x associated to occupational category y, for $x \in \{1, \dots, 6\}$ and $y \in \{1, \dots, 9\}$. Worker types are estimated through finite-mixture method recursively every two years and using the observed log weekly wage. Occupations are the one-digit ISCO occupations, ten categories which are ordered from the most intensive in non-routine tasks to the most intensive in routine tasks.

Table C.4 reports the estimated coefficients for the Multinominal Logit on worker types. Note that the estimates with raw earnings exhibit few puzzling patters such as the association of the lower worker type (1) with some characteristics that are typical of high income earners, such as being an executive. There are good statistical and

Table C.4: Multinomial logit - Worker Types on Observables

VARIABLES	(1) 1	(2) 2	(3) 3	(4) 4	(5) 5	(6) 6
mover	0.307***	-0.344***		0.0364***	0.0365**	-0.234***
	(0.0292)	(0.0142)		(0.00942)	(0.0146)	(0.0263)
trainee	1.632***	1.774***		-1.476***	-3.035***	-3.901***
	(0.0427)	(0.0166)		(0.0245)	(0.0885)	(0.277)
blue collar	-2.268***	-0.281***		-0.579***	-2.399***	-4.765***
	(0.0369)	(0.0182)		(0.0114)	(0.0149)	(0.0268)
white collar	-1.313***	-0.588***		0.732***	0.837***	-0.376***
	(0.0367)	(0.0186)		(0.0112)	(0.0147)	(0.0197)
"quadro"	4.447***	0.480***		3.145***	6.178***	6.693***
	(0.0820)	(0.169)		(0.0723)	(0.0716)	(0.0728)
executive	8.551***	-0.371		0.696***	3.894***	7.463***
	(0.135)	(0.360)		(0.148)	(0.133)	(0.133)
age	0.0679***	-0.0115***		0.0442***	0.0911***	0.106***
	(0.000895)	(0.000337)		(0.000177)	(0.000278)	(0.000588)
weeks	-0.0748***	-0.0377***		0.0647***	0.0569***	-0.000258
	(0.000731)	(0.000328)		(0.000332)	(0.000537)	(0.000935)
Centre	0.734***	0.0359		0.305***	0.850***	1.240***
	(0.0478)	(0.0287)		(0.0172)	(0.0220)	(0.0343)
Nord-East	0.896***	0.137***		0.728***	1.366***	1.858***
	(0.0480)	(0.0286)		(0.0172)	(0.0220)	(0.0345)
Nord-West	1.039***	0.157***		0.766***	1.383***	1.744***
	(0.0474)	(0.0285)		(0.0171)	(0.0219)	(0.0341)
South	0.612***	-0.0653**		-0.0816***	0.218***	0.483***
	(0.0482)	(0.0285)		(0.0172)	(0.0223)	(0.0358)
Islands	0.454***	-0.140***		-0.179***	0.324***	0.649***
	(0.0526)	(0.0295)		(0.0178)	(0.0234)	(0.0392)
Observations	28,365	161,766	1,163,935	1,074,202	428,508	101,618

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Multinomial logit for the period 2005-2008 where the assigned worker types, estimated through finite-mixture models, are regressed on a set of observables, namely the status of mover, the qualification (six dummies for being employed in one of the following categories: trainee, blue collar, white collar, "quadro" or executive), age, weeks worked, and macroarea of employment. The reference category is the third worker type. Estimation algorithm for worker effects employs raw earnings. Reference cluster is number 3.

economic reasons for that. Worker type 1 is very small, counting only 28 thousand workers in a sample of almost 3 million workers. This makes it very responsive to outliers. The puzzling pattern indeed disappears when we employ as dependent variable the worker fixed effects estimated using the Mincer residual (shown in Table 1 in the main text). Moreover, worker type 1 may capture individual contracts of non-executive board members, whose salaries albeit high may not parallel those of executives, or of other professions, such as civil servants, that command lower wages than their equivalent in the private sector.

Finally, Table C.5 shows coefficient estimates for the Multinomial logit on firm clusters. Results are very intuitive and in line with the equivalent using Mincer residual and shown in the main draft.

Table C.5: Multinomial logit - Firm Clusters on Observables

VARIABLES	(1) 1	(2) 2	(3) 3	(4) 4	(5) 5	(6) 6	(7) 7	(8) 8	(9) 9	(10) 10
log_l	1.270***	0.528***	0.494***	0.387***		0.0519	-0.801***	-0.685***	-1.419***	-2.046***
	(0.0494)	(0.0363)	(0.0312)	(0.0311)		(0.0330)	(0.0343)	(0.0415)	(0.0411)	(0.0552)
log_va	-2.418***	-1.794***	-1.206***	-0.786***		-0.103***	0.993***	0.761***	1.708***	2.401***
	(0.0435)	(0.0328)	(0.0289)	(0.0289)		(0.0307)	(0.0321)	(0.0385)	(0.0382)	(0.0500)
(mean) poach_score	-0.712***	-0.134**	-0.0815	0.0167		-0.0792	0.0640	-0.382***	-0.332***	-0.971***
	(0.0947)	(0.0656)	(0.0583)	(0.0590)		(0.0646)	(0.0732)	(0.0879)	(0.0960)	(0.143)
(mean) share_sales	-1.229***	1.292***	0.784***	0.128		-0.460*	-1.627***	-1.484***	-3.302***	-6.621***
	(0.406)	(0.244)	(0.220)	(0.225)		(0.248)	(0.270)	(0.340)	(0.358)	(0.598)
(p50) (p50) macro_area==Centro	-0.789*	-0.618**	-0.150	-0.0350		-0.115	-0.309	-0.842***	-0.865***	-0.866***
	(0.404)	(0.305)	(0.233)	(0.225)		(0.221)	(0.200)	(0.233)	(0.215)	(0.287)
(p50) (p50) macro_area==Isole	0.0465	0.380	0.422*	-0.212		-0.654***	-0.839***	-1.392***	-1.509***	-1.284***
	(0.409)	(0.310)	(0.239)	(0.233)		(0.236)	(0.227)	(0.269)	(0.265)	(0.353)
(p50) (p50) macro_area==Nord-est	-1.394***	-1.666***	-0.759***	-0.0984		0.326	0.0501	-0.241	-0.466**	-0.908***
	(0.403)	(0.305)	(0.233)	(0.224)		(0.219)	(0.198)	(0.230)	(0.212)	(0.285)
(p50) (p50) macro_area==Nord-ovest	-1.506***	-1.834***	-0.894***	-0.220		0.235	0.185	-0.251	-0.0937	0.115
	(0.403)	(0.305)	(0.232)	(0.224)		(0.219)	(0.198)	(0.230)	(0.210)	(0.278)
(p50) (p50) macro_area==Sud	0.741*	0.634**	0.474**	-0.104		-0.772***	-0.953***	-1.280***	-1.649***	-2.044***
	(0.403)	(0.306)	(0.234)	(0.227)		(0.225)	(0.209)	(0.243)	(0.236)	(0.341)
Observations	6,132	24,264	31,330	21,990	11,335	13,096	7,945	4,590	4,029	1,807
Workforce	100,107	278,892	504,090	452,543	436,514	467,347	340,983	350,710	211,951	75,114

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: Multinomial logit for the period 2005-2008 where the kmeans firm cluster is regressed on a set of observables, namely (log) size of workforce, (log) value added, share of sales on markets defined as province x 3-digit Ateco sector (the Italian classification for sectors), poaching score and macroarea dummies. The reference category is the fifth cluster. Firm clusters are estimated on raw earnings.

C.3. Alternative Specifications for Sorting

In this Section we report estimates for alternative specifications of Equation 12. We report results for the case in which sorting is estimated based on the Mincer-residual and the results for the case in which automated robots are exclusively installed in the automobile industry.

C.3.1. Automation Estimates

Table C.6: Impact of Automated Robots on Sorting - Mincer wages- With and Without Automobile.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	OLS	OLS	OLS	OLS	IV	IV	IV	IV	IV2	IV2	IV2	IV2
New Robots	0.0108***	0.00690***	0.00691***	0.00506**	-0.0105**	0.00998***	0.0109***	0.0113***	0.0353***	0.00388	0.00179	-0.00148
	(0.00232)	(0.00228)	(0.00231)	(0.00209)	(0.00422)	(0.00339)	(0.00351)	(0.00325)	(0.0121)	(0.00792)	(0.00826)	(0.00761)
Observations	937	937	937	937	937	937	937	937	937	937	937	937
R^2	0.140	0.553	0.574	0.641	0.100	0.553	0.573	0.638	0.087	0.553	0.572	0.638
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
A-R p value					0.00882	0.00288	0.00152	0.000335	0.000163	0.631	0.831	0.849
M-P F stat					546.9	344.9	333.4	327.7	13.36	13.54	12.36	13.19
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Anderson-Rubin tests for joint null of orthogonality and non-significance of endogenous regressors.

F-stats above the critical value reject null of weak identification.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	OLS	OLS	OLS	OLS	IV	IV	IV	IV	IV2	IV2	IV2	IV2
New Robots	-0.0532***	0.0663***	0.0632***	0.0640***	-0.0602***	0.0230	0.0262	0.0359*	-0.102***	0.104**	0.0798*	0.113***
	(0.0107)	(0.0199)	(0.0182)	(0.0188)	(0.0106)	(0.0202)	(0.0190)	(0.0194)	(0.0320)	(0.0522)	(0.0483)	(0.0429)
Observations	937	937	937	937	937	937	937	937	937	937	937	937
R^2	0.156	0.560	0.579	0.647	0.155	0.556	0.576	0.646	0.134	0.557	0.579	0.643
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
A-R p value					2.71e-08	0.263	0.177	0.0767	0.000754	0.0620	0.118	0.0158
M-P F stat					4362	1890	1899	1747	219.1	168.1	154.8	162.7
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Anderson-Rubin tests for joint null of orthogonality and non-significance of endogenous regressors.

F-stats above the critical value reject null of weak identification.

Note: OLS and IV estimates of the impact of automated robots installations on sorting, corr(θ, $\psi_{j(i,t)}$), computed at the province level. Controls include three quartiles of two-digit sector HHI for employment shares - HHI_{25}, HHI_{50} and HHI_{75}. Covariates include the share of workforce employed in the manufacturing and construction macro-sectors. Finally, macro-area and two-year period fixed effects are included either separately or interacted. For the IV estimates we employ two shift-share instruments - in columns (5) to (8) we exploit robot information for Europe, US and Japan; in columns (9) to (12) we exclude Europe from the instrumental variable construction. Estimated sorting is based on the Mincer residual. Top table includes automobile, bottom table excludes automobile

C.3.2. Automobile Robots - Automation Estimates

Table C.7: Impact of Automated Robots on Sorting - Automobile Industry only. Top table is raw wages. Bottom table is Mincer.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV2	(10) IV2	(11) IV2	(12) IV2
New Robots	0.0176***	-0.00145	5.81e-05	-0.000239	0.00808**	0.00157	0.00406	0.00492	-0.00870	0.000393	0.00246	0.00582
	(0.00295)	(0.00269)	(0.00268)	(0.00276)	(0.00389)	(0.00379)	(0.00375)	(0.00389)	(0.0119)	(0.00870)	(0.00898)	(0.00840)
Observations	883	883	883	883	883	883	883	883	883	883	883	883
R^2	0.263	0.574	0.605	0.641	0.258	0.573	0.604	0.640	0.222	0.574	0.605	0.639
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	No	No	No	No	No	No	No	No	No	No	No	No
A-R p value					0.0406	0.681	0.280	0.216	0.417	0.964	0.788	0.502
M-P F stat					308	339.5	329.4	331	11.23	14.19	12.93	13.63
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Anderson-Rubin tests for joint null of orthogonality and non-significance of endogenous regressors.

F-stats above the critical value reject null of weak identification.

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) IV	(8) IV	(9) IV2	(10) IV2	(11) IV2	(12) IV2
New Robots	0.0178***	0.00413*	0.00444**	0.00226	-0.00240	0.00667**	0.00752**	0.00707**	0.0552***	-0.00146	-0.00178	-0.00767
	(0.00282)	(0.00220)	(0.00214)	(0.00201)	(0.00433)	(0.00335)	(0.00335)	(0.00310)	(0.0178)	(0.00742)	(0.00753)	(0.00719)
Observations	883	883	883	883	883	883	883	883	883	883	883	883
R^2	0.162	0.552	0.582	0.649	0.128	0.552	0.581	0.648	0.046	0.550	0.580	0.644
Period FEs	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No
Macroarea FEs	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Macroarea x Period FEs	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Mincer	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
A-R p value					0.574	0.0454	0.0234	0.0220	7.32e-08	0.844	0.814	0.266
M-P F stat					308	339.5	329.4	331	11.23	14.19	12.93	13.63
M-P .05 critical value					37.42	37.42	37.42	37.42	37.42	37.42	37.42	37.42

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Anderson-Rubin tests for joint null of orthogonality and non-significance of endogenous regressors.

F-stats above the critical value reject null of weak identification.

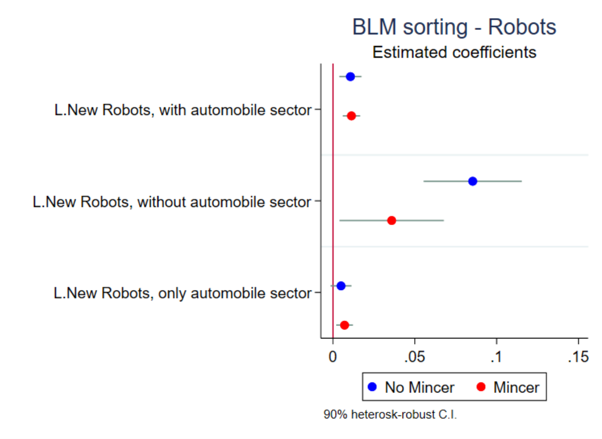
Note: OLS and IV estimates of the impact of automated robots installations on sorting, corr(θ, $\psi_{j(i,t)}$), computed at the province level and focusing solely on automobile industry. Controls include three quartiles of two-digit sector HHI for employment shares - HHI_{25}, HHI_{50} and HHI_{75}. Covariates include the share of workforce employed in the manufacturing and construction macro-sectors. Finally, macroarea and two-year period fixed effects are included either separately or interacted. For the IV estimates we employ two shift-share instruments - in columns (5) to (8) we exploit robot information for Europe, US and Japan; in columns (9) to (12) we exclude Europe from the instrumental variable construction. Top table is raw wages. Bottom table is Mincer.

C.3.3. Summary and Comparison of Estimates Across Specifications

We entertain in total six different specifications, by combining those with different automated robot installations considered (whole economy, excluding automobile industry, automobile industry only) and those with different measures of sorting (Mincer and non Mincer). For each of these specifications we estimate positive coefficients. The presence of the automobile industry drags down the point estimates. Its exclusion rises the economic impact on sorting, but makes estimates less precise.[23] The following graph plots our coefficient of interest across specifications.

[23]This is to be expected given the relatively lower variance that our shift-share measure for automation exhibits without the automobile sector, as shown in Section C.

Figure C.5: Impact of Automation on Sorting, different specifications



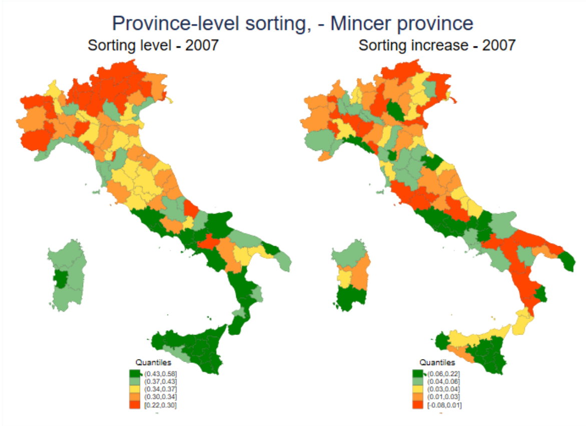
Note: Coefficient of lagged automated robot installations with IV estimates, including interacted period and area FEs - across different automation definitions and sorting computations.

C.4. Production Complementarities vs. Geographical Assortativity

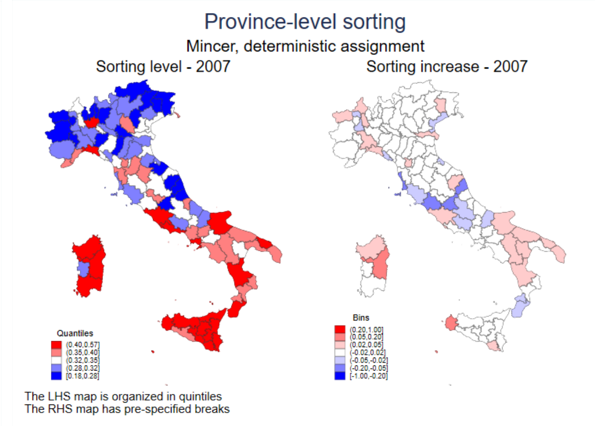
Our strategy exploits cross-sectional variation across provinces by relying on a Bayesian assignment of worker types to provinces. A concern one may have is that the estimated sorting captures not only production complementarities but also other possible drivers of geographical assortativity, such as local amenities or agglomeration externalities that induce workers to sort into certain provinces. This is unlikely to be the case. On the one hand, the dominant part of Italian manufacturing activity takes place in northern provinces that do not differ much in terms of local amenities and agglomeration economies. On the other hand, the largest differences in the Italian industrial landscape are observed between macro areas (North-West, North-East, Centre, South, and Islands) rather than within them, which we take account of by including macro area controls in the last stage of our econometric analysis.

That said, to further support our claim that the sorting we estimate is likely determined by genuine firm-worker production complementarities rather than other drivers of geographical assortativity, we report here results obtained by including province controls in the Mincer regression, which should allow us to purge the wage variation from its geographical component. For parsimony we rely on 2s-AKM as this is easier to implement than BLM and, as shown in Section 4.1, delivers sorting estimates that are quite close to those obtained through BLM. The results are depicted through maps in panel (a) of Figure C.6, where the estimated sorting and its increase are plotted across Italian provinces using orange, yellow and green to color code high, intermediate and low sorting (left map) or sorting increase (right map). For comparison, panel (b) reports the corresponding maps when sorting and its increase are estimated through BLM without province controls in the Mincer regression. To better distinguish the two panels visually, in the right one we adopt a different palette using blue, white and red color to color code high, intermediate and low estimated sorting (left map) or sorting increase (right map). The comparison between the two panels reveals that sorting and its increase are very similarly distributed across provinces with and without the province controls. We interpret this similarity as confirming that our estimates are indeed not driven by geographical assortativity.

Figure C.6: Regional sorting across specifications: the maps below plot the estimated sorting by province in the case in which the Mincer includes province controls (left panel) and in the case in which it does not (right panel).

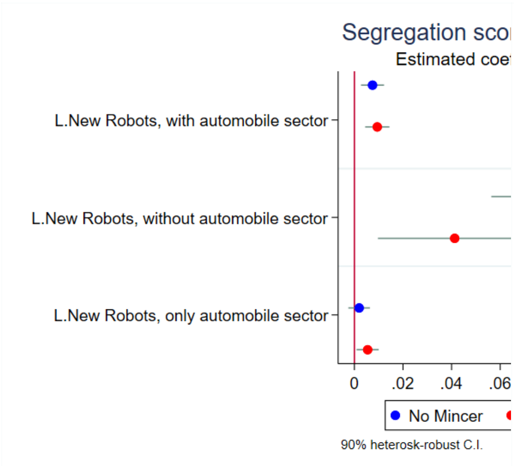


(a) **With province controls**

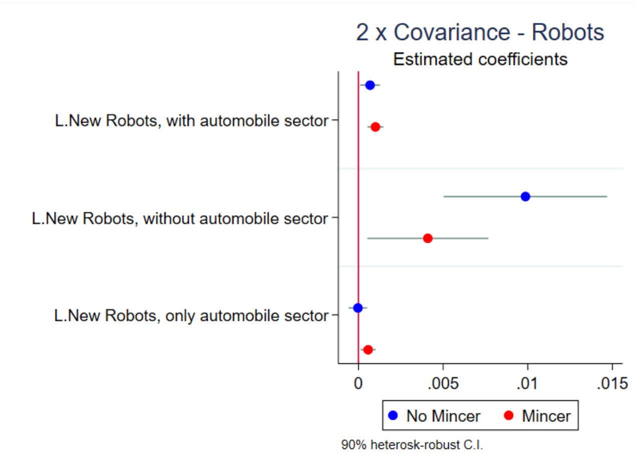


(b) **Without province controls**

Figure C.7: Impact of Automation on Segregation, different specifications



(a) **Segregation**



(b) **Covariance**

Note: Coefficient of segregation (left panel) and covariance (right panel) on lagged automated robot installations under a specification based on an IV estimate, which includes interacted period and area FEs - across different automation definitions and segregation score computations.

C.5. Alternative dependent variables

The very same pattern concerning the automobile sector in terms of sorting can be found when using covariance of worker and firm effects or segregation as dependent variables of our last stage.

D. Model Derivations

D.1. Wage Equation

The value function of a x-type worker employed by a y-type firm at wage w is:

$$\rho V(x,w,y) = w - c(\mu(x,w,y)) + \delta[V_0(x) - V(x,w,y)] + \lambda_1(x)\left[\int_{q(x,w,y)}^{y} (\eta V(x,s(x,y),y)\right. \tag{27}$$

$$+ (1-\eta)V(x,s(x,p),p))dF(p) + \int_{y}^{\overline{y}} (\eta V(x,s(x,p),p)+$$

$$\left. +(1-\eta)V(x,s(x,y),y))dF(p) - \int_{q(x,w,y)}^{\overline{y}} V(x,w,y)dF(p)\right]$$

Once integrated by parts, the term between square brackets evaluates to:

$$\begin{aligned} &-\overline{F}(y)V(x,s(x,y),y) + \overline{F}(q)[\eta V(x,s(x,y),y) + (1-\eta)V(x,s(x,q),q)]+ \\ &+ (1-\eta)\int_{q}^{y} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp+ \\ &+ \overline{F}(y)V(x,s(x,y),y) + \eta\int_{y}^{\overline{y}} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp - \overline{F}(q)V(x,w,y) \end{aligned} \tag{28}$$

where $\overline{F}(p) = 1 - F(p)$. Given that $q(x,w,y)$ is defined as the value of q that solves $w = \phi(x,q,y)$, the sharing rule implies

$$V(x,w,y) = \eta V(x,s(x,y),y) + (1-\eta)V(x,s(x,q),q) \tag{29}$$

so that (28) simplifies to:

$$(1-\eta)\int_{q}^{y} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp + \eta\int_{y}^{\overline{y}} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp \tag{30}$$

Substituting (30) into (27) entails:

$$(\rho+\delta)V(x,w,y) = w - c(\mu(x,w,y)) + \delta V_0(x) + \lambda_1(x)\left[(1-\eta)\int_{q(x,w,y)}^{y} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp\right. \tag{31}$$

$$\left. +\eta\int_{y}^{\overline{y}} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp\right]$$

Imposing $w = s(x,y)$ and thus $q(x,s(x,y),y) = y$, equation (31) becomes:

$$(\rho+\delta)V(x,s(x,y),y) = s(x,y) - c(\mu(x,w,y)) + \delta V_0(x) + \lambda_1(x)\eta\int_{y}^{\overline{y}} \overline{F}(p)\frac{d}{dp}V(x,s(x,p),p)dp \tag{32}$$

Differentiating the above with respect to y delivers:

$$\frac{d}{dy}V(x,s(x,y),y) = \frac{s_y(x,y)}{\rho + \delta + \lambda_1(x)\eta\overline{F}(y)} \tag{33}$$

Substituting (33) into (32) yields:

$$(\rho+\delta)V(x,s(x,y),y)=s(x,y)-c(\mu(x,w,y))+\delta V_0(x)+\lambda_1(x)\eta\int_y^{\overline{y}}\frac{s_y(x,p)\overline{F}(p)}{\rho+\delta+\lambda_1(x)\eta\overline{F}(y)}dp \tag{34}$$

Furthermore, substituting (33) into (31) delivers:

$$(\rho+\delta)V(x,w,y)=w-c(\mu(x,w,y))+\delta V_0(x)+\lambda_1(x)\left[(1-\eta)\int_{q(x,w,y)}^{y}\frac{s_y(x,p)\overline{F}(p)}{\rho+\delta+\lambda_1(x)\eta\overline{F}(y)}dp+\right. \tag{35}$$

$$\left.+\eta\int_y^{\overline{y}}\frac{s_y(x,p)\overline{F}(p)}{\rho+\delta+\lambda_1(x)\eta\overline{F}(y)}dp\right]$$

For $w=\phi(x,y',y)$ with $y'\leq y$, (34) and (35) imply:

$$\phi(x,y',y)=\eta s(x,y)+(1-\eta)s(x,y')-\lambda_1(x)(1-\eta)^2\int_{y'}^{y}\frac{s_y(x,p)\overline{F}(p)}{[\rho+\delta+\lambda_1(x)\eta\overline{F}(p)]} \tag{36}$$

$$\phi(x,y',y)=\eta s(x,y)+(1-\eta)s(x,y')-(1-\eta)^2\lambda_1(x)\int_{y'}^{y}\frac{s_y\left(x,p\right)\overline{F}(p)}{\rho+\delta+\lambda_1(x)\eta\overline{F}(p)}dp \tag{37}$$

Using $s(x,y')=s(x,y)-\int_{y'}^{y}s_y\left(x,p\right)dp$, (37) can be restated in more compact firm as:

$$\phi(x,y',y)=s(x,y)-(1-\eta)\left(\int_{y'}^{y}\frac{\rho+\delta+\lambda_1(x)\overline{F}(p)}{\rho+\delta+\lambda_1(x)\eta\overline{F}(p)}s_y\left(x,p\right)dp\right)$$

which is expression (21) in the main text.

D.2. Wage Distribution

Consider the balanced flows condition (23) for $y'=q$ and $y'=y$. In this case, we have $w=s(x,y)$ and $G(w=s(x,y)\mid x,y)=1$. The condition simplifies to:

$$\left(\delta+\overline{F}(y)\lambda_1\left(x\right)\right)\ell(x,y)=\left(\delta g(x)+\lambda_1\left(x\right)\int_{\underline{y}}^{y}\ell(x,p)dp\right)f(y) \tag{38}$$

To intergrate by parts, define: $u=\left(\delta+\overline{F}(y)\lambda_1\left(x\right)\right)=\left[\delta+\left(1-F(y)\right)\lambda_1\left(x\right)\right]$, $u'=-\lambda_1\left(x\right)f(y)$,$v=\int_{\underline{y}}^{y}\ell(x,p)dp$, $v'=\ell(x,y)$. Then, (38) can be restated as:$-\frac{\delta}{\lambda_1(x)}g(x)u'=u'v+uv'$, which can be intregrated to yield:$-\frac{\delta}{\lambda_1(x)}g(x)\int_{\underline{y}}^{y}u'=\int_{\underline{y}}^{y}\left(u'v+uv'\right)=uv$. Hence, recalling the definition above for u, U', v and v', we have:

$$\delta g(x)\left(1-\overline{F}(y)\right)=\left(\delta+\overline{F}(y)\lambda_1\left(x\right)\right)\int_{\underline{y}}^{y}\ell(x,p)dp$$

which can be rewritten as: $\int_{\underline{y}}^{y} \ell(x,p)dp = \frac{\delta g(x)\left(1-\overline{F}(y)\right)}{\left(\delta+\lambda_1(x)\overline{F}(y)\right)}$. Plugging this last expression into (38) and simplifying gives:

$$\left(\delta + \overline{F}(y)\lambda_1\left(x\right)\right)\ell(x,y) = \left(\frac{\delta + \lambda_1\left(x\right)}{\delta + \lambda_1\left(x\right)\overline{F}(y)}\right)g(x)f(y)$$

which can be solved for:

$$\ell(x,y) = \frac{1 + \frac{\lambda_1(x)}{\delta}}{\left(1 + \overline{F}(y)\frac{\lambda_1(x)}{\delta}\right)^2} g(x)f(y)$$

which is expression (24) in the main text.