Sorting Robots: How Automation Shapes the Allocation of Workers Across Firms*

Salvatore Lattanzio

University of Cambridge

July 26, 2020

Preliminary and incomplete. Please do not cite.

Abstract

This paper provides evidence on the effect of automation on the allocation of workers across firms in the labor market. Combining automation data at the sector level and matched employer-employee data over the period 1993-2016 for Italy, the paper shows that automation contributes to positive assortative matching, i.e. the tendency of high-wage workers to be employed by high-wage firms. In particular, at the regional level one additional robot per 1000 workers increases sorting by 9.3% relative to the mean. At the firm level, only firms in the top quartile of the firm effect distribution match with better workers in response to higher automation exposure. These results contribute to the understanding of the effects of automation on labor demand and shed light on one mechanism that determine labor market sorting, a key driver of rising earnings inequality.

Keywords: Automation, Sorting, Matching, Matched Employer-Employee Data

***JEL* codes:** J23, J30, O33

1 Introduction

The rise in earnings inequality has been widely documented in many advanced economies, including the US, the UK, Germany and Italy (Autor et al., 2008; Belfield et al., 2017; Dustmann et al., 2009; Franzini and Raitano, 2019). Traditional explanations put forward in the literature highlight the role of changing returns to skills and occupations as drivers of earnings inequality (Katz and Autor, 1999). More recent evidence discusses the role of firm-specific determinants and studies the contribution of between-firm pay differences in explaining earnings differentials between otherwise observationally equivalent workers. In

*I would like to thank Massimo Anelli, Alessandra Casarico, Italo Colantone, Kai Liu, Hamish Low and audiences at the University of Cambridge and Collegio Carlo Alberto for helpful comments and discussion. I gratefully acknowledge financial support from the Keynes Fund.

particular, this literature highlights the importance of the sorting of high-wage workers into high-wage firms as a key driver of earnings inequality (Barth et al., 2016; Card et al., 2013; Song et al., 2018).

Despite being acknowledged as one of the key drivers of earnings inequality, little is known about the determinants of rising assortative matching in the labour market. A recent strand of the literature investigates the impact of outsourcing (Goldschmidt and Schmieder, 2017) and trade (Bombardini et al., 2019; Colantone et al., 2019; Smith, 2018) on sorting and the labour demand of high-skill workers. Very little is known, however, on the impact of automation on labour market sorting. And yet automation adoption may have an important effect on the allocation of workers across firms. In a simple theoretical framework, the decision to automate by a firm can be seen as the choice between producing a given task with human labour or with machines (Acemoglu and Restrepo, 2019). Assuming that routine tasks, carried out by low-skilled workers in the absence of automation, can be performed by machines, the decision to automate will affect labour demand by displacing low-skilled workers and by increasing the value of high-skilled workers who are complementary to automation, as they can perform non-routine and non-automatable tasks. Rising assortative matching is therefore an expected consequence of increased automation, given that firms which automate become more productive and increase demand for high-skill workers.

So far the literature has found mixed results across Europe and the US as to the effect of automation on the labour market and the labour demand (Acemoglu and Restrepo, 2020; Bessen et al., 2019; Dauth et al., 2018; Graetz and Michaels, 2018), but there is still a scant literature (probably also because of data limitations) on the relation between automation and sorting, one key mechanism through which the impact of automation on labour demand can materialise.

This paper aims at filling this gap in the literature by providing evidence on the causal impact of automation on sorting, both at the geographical level and at the firm level. To do so, I exploit rich administrative matched employer-employee data for Italy over the period 1993-2016, linked to information on robot adoption at the industry level through the International Federation of Robotics. The availability of worker and firm identifiers allows to estimate a two-way fixed effects wage equation (Abowd et al., 1999, AKM henceforth). With the estimated fixed effects, I conduct two analyses. First, at the region level, I compute sorting as the correlation between worker and firm effects in each year and region. I then regress this measure on robot adoption at the region level, computed as a shift-share variable, exploiting the number of industrial robots by sector and historical industry shares in each region, in the spirit of Acemoglu and Restrepo (2020). Second, I conduct a firm-level analysis where I regress the average and the dispersion of worker effects in the firm on robot adoption by sector. For both analyses I provide OLS and instrumental variable estimates, exploiting robot adoption in the same sectors in seven other European countries (Germany, France, Spain, Sweden, Denmark, Finland and the United Kingdom).

In the most conservative estimate, the results suggest that one additional robot per 1000 worker in a region increases sorting by 0.3 log points or 9.3% relative to the mean. When focusing only on manufacturing and construction sectors, I do not find a significant effect of automation on sorting, suggesting that the effects is mainly driven by a reallocation of workers out of manufacturing into services. When differentiating between different types of workers I find positive and significant effects for men, workers older than 35 and white-collar jobs.

At the firm level, I find that one additional robot per worker in the sector which the firms belongs to raises wages by 0.03 log points, but I do not find significant effects on the average worker effect or its dispersion. However, when splitting the sample by quartiles of the firm fixed effect distribution I find that both the average worker quality significantly increases in firms in the top quartile, highlighting again the increase in assortativeness between high-wage workers and high-wage firms.

This paper contributes to two strands of the literature. First, the one investigating the drivers of rising assortativeness in the labour market, which would also benefit the understanding of earnings inequality, given the evidence proving that sorting is one key driver behind its rise (see Card et al., 2018, for a review). For Italy, Devicienti et al. (2019) show that the growth in assortativeness between worker and firm types explains more of the rise in the variance of wages between 1982 and 2001. Complementary to this evidence, the present paper highlights that a part of this increase in assortativeness and inequality is due to increased automation adoption. Second, it contributes to the literature that investigates the mechanisms through which automation affects labour demand (Acemoglu and Restrepo, 2020; Dauth et al., 2018). One specific and under-explored channel is precisely the reallocation of workers of different skills across firms of different productivity.

The remainder of the paper is organized as follows. Section 2 describes the data. Section 3 provides details on the empirical strategy. Section 4 shows the results of the effect of automation on labor market sorting. Section 5 concludes.

2 Data and Descriptive Statistics

There are two main data sources. On the worker-firm side, I use LoSaI-Inps records, a matched employer-employee dataset that contains a random sample of the universe of workers in the Italian non-agricultural private sector. The data covers approximately 7% of the universe of employees over the period 1985-2016. The dataset is comprised of different archives. In the worker archive, I observe the entire working history of each individual and, specifically, I have information on annual gross earnings,[1] the number of weeks and days worked in a year, the type of contract (full-time or part-time and permanent or temporary) and broad occupation categories (apprentice, blue-collar, white-collar, middle manager, executive). In a separate archive, the dataset contains demographic information on each em-

[1]The measure of earnings is gross of labor incomes taxes and pension contributions on the side of the employee.

ployee, such as year of birth, gender and region of residence. Finally, the firm archive records total firms size in discrete brackets and firm's sector.

On the automation side, data on robot adoption are from the Industrial Federation of Robotics, which records the number of industrial robots per sector in a number of different countries over the period 1993-2016. For the purpose of this paper, I will focus on the number of industrial robots per sector in Italy plus seven European countries (Germany, France, Spain, Sweden, Denmark, Finland and the UK). The sectors included in the IFR data are detailed in Table B.1 in the appendix. Given that LoSaI data contain sector information at the 2-digit level I group sectors into 18 groups (as shown in the third column of Table B.1), corresponding, approximately, to 2-digit sectors in manufacturing and 1-digit sectors in non-manufacturing. Although data on robots are for non-service sectors only, I retain firms belonging also to the service economy in the employer-employee matched data, in order to capture possible reallocation of workers across sector.

Sample restrictions In the worker data, I restrict the sample to the main job, i.e. the job with the highest number of weeks or days worked in a given year. If two jobs have the same number of weeks/days worked, I keep the one with the highest earnings. Remaining duplicates are dropped randomly (less than 0.2% of total person-year observations). I only retain workers aged 25 to 60 and wages of part-time workers are normalized to take into account the reduced number of hours they work in a year.[2]

Finally, I focus on the largest connected sample, i.e. the sample of workers and firms connected by worker mobility. Specifically, a connected set contains all workers that have ever been employed by one of the firms in the sample and all the firms that have ever employed one of the workers in the set. The choice is motivated by the empirical model used to estimate worker and firm fixed effects, which will be discussed in more detail in Section 3. For the same reasons, I keep only firms with at least 10 person-year observations.

Table 1 reports descriptive statistics for the full sample and the largest connected set in columns 1 and 2, respectively. The total number of person-year observations in the connected sample is 19 million, comprising approximately 93% of observations in the full sample. The number of workers is 1.9 million (94% of the full sample), whereas the number of firms is 367 thousand (83% of the full sample). Mean age and tenure are, respectively, 40 and 6 years. The share of female workers is 36%. The share of workers with part-time contracts is 11%. Blue-collar workers comprise 56% of the sample, whereas white-collar workers are 40%, with little differences between the full sample and the connected sample. Real daily wages are slightly higher in the largest connected set, but the difference is very small in economic terms: 97.6 against 96.3 Euro.[3]

[2]The dataset contains a variable with the adjusted number of weeks worked in a year, which takes into account part-time contracts. I convert the variable into days and use the adjusted days to compute daily wages as the ratio of annual earnings to days worked.

[3]Wages are expressed in 2015 prices.

Table 1: Summary statistics

	(1) Full sample	(2) Connected sample
Age	40.18	40.19
Tenure	5.97	5.79
Female	0.36	0.36
Part-time	0.11	0.11
Blue-collar	0.56	0.56
White-collar	0.40	0.40
Real daily wage	96.25	97.60
Person-year obs.	20,381,212	19,009,639
Share of full sample		*93.3%*
Number of workers	1,996,554	1,880,906
Share of full sample		*94.2%*
Number of firms	445,495	367,476
Share of full sample		*82.5%*

3 Empirical Strategy

This section provides details on the measurement of sorting, automation adoption and the empirical model that links the two.

Sorting and Matching I first derive measures of worker and firm quality from an AKM two-way fixed effects wage regression (Abowd et al., 1999):

$$w_{it} = \alpha_i + \psi_{J(i,t)} + X_{it}'\beta + \varepsilon_{it} \tag{1}$$

where w_{it} are log daily wages of worker i at time t, α_i is a worker fixed effect, $\psi_{J(i,t)}$ is a firm fixed effect – with $J(i,t)$ indicating the firm that employs worker i at time t –, X_{it} contains observable worker characteristics (cubic polynomials in age and tenure, occupation dummies and the interaction of all these variables with a female dummy) and year fixed effects, ε_{it} is an error term.

I measure sorting by computing the correlation between the estimated worker and firm effects at the region level:[4]

$$y_{\ell t} = \text{Corr}_{\ell t}\left(\widehat{\alpha}_i, \widehat{\psi}_{J(i,t)}\right) \tag{2}$$

where ℓ indexes regions, whereas hats over letters indicate estimated quantities from equation 1.

[4]There are 20 regions in Italy, corresponding to NUTS-2 groups. One shortcoming of the data is that the only geographical information available is the last region of residence of the worker. Hence, I assume that individuals work in the same region where they live and they do not change region of residence throughout the whole period of analysis. This is of course a rather restrictive assumption, which could generate some measurement error in the outcome variable. However, this should affect only the precision of the estimates, provided that the measurement error is random.

At the firm-level, following Bombardini et al. (2019), I measure the quality of workers matched with the firm by computing the mean and the standard deviation of firm-level worker effects:

$$m_{jt} = \frac{1}{n_{jt}} \sum_{i \in J(i,t)} \widehat{\alpha}_i \tag{3}$$

$$s_{jt} = \frac{1}{n_{jt}} \sqrt{\sum_{i \in J(i,t)} \left(\widehat{\alpha}_i - m_{jt}\right)} \tag{4}$$

where n_{jt} is the number of workers employed in firm j in year t and $\widehat{\alpha}_i$ is the estimated worker effect from equation 1.

Automation adoption At the regional level, I compute a time-varying measure of automation adoption, by assigning sector level variation to regions using historical sector employment shares:

$$\text{Regional exposure}_{\ell t} = \sum_s \frac{L_{\ell s}^{t_0}}{L_{\ell}^{t_0}} \times \frac{R_s^{t-1}}{L_s^{t_0}} \tag{5}$$

where $L_{\ell s}^{t_0}/L_{\ell}^{t_0}$ indicate the historical employment shares[5] in sector s and region ℓ and $R_s^{t-1}/L_s^{t_0}$ indicates the lagged number of robots per worker at the sector-level.

At the firm level, I can exploit variation in the sector-level number of robots per worker:

$$\text{Firm exposure}_{S(j)t} = \frac{R_{S(j)}^{t-1}}{L_{S(j)}^{t_0}} \tag{6}$$

where $R_{S(j)}^{t-1}/L_{S(j)}^{t_0}$ is the lagged number of robots per worker in sector $S(j)$, i.e. the sector which firm j belongs to.

Estimating equations The effect of automation on sorting at the region- or firm-level can be estimated through the following OLS regressions:

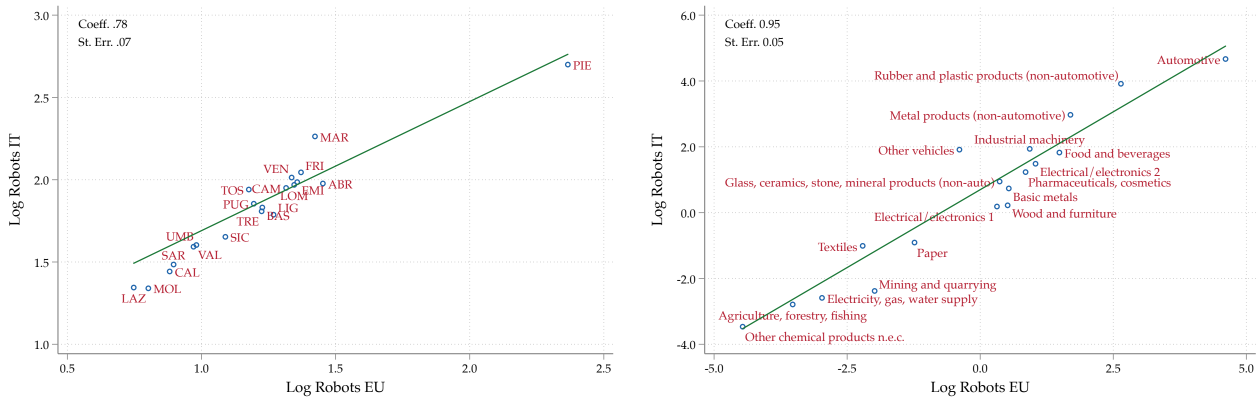
$$y_{\ell t} = \beta \text{Regional exposure}_{\ell t} + \zeta Z_{\ell t} + \lambda_{g(\ell)} + \delta_t + \varepsilon_{\ell t} \tag{7}$$

$$y_{jt} = \gamma \text{Firm exposure}_{S(j)t} + \theta X_{jt} + \eta_j + \delta_t + \epsilon_{jt} \tag{8}$$

where $y_{\ell t}$ is defined in (2) and $y_{jt} = \{m_{jt}, s_{jt}\}$. $Z_{\ell t}$ and X_{jt} are region- and firm-level controls, respectively;[6] $\lambda_{g(\ell)}$ are area fixed effects; η_j are firm fixed effects; δ_t are year effects; $\varepsilon_{\ell t}$ and

[5]The superscript t_0 indicates that employment shares are computed in years before the first automation wave, specifically using the 1991 wave of the census of firms in manufacturing and services provided by the Italian Statistical Institute (Istat). The use of past rather than contemporaneous employment shares addresses the potential endogeneity of employment to automation.

[6]Specifically, $Z_{\ell t}$ contains the 1991 region-level shares of blue collar, female, manufacturing and skilled workers, and the initial correlation (i.e. in 1993) between worker and firm effects. X_{jt} contains the firm-level shares of female, part-time and blue-collar workers, average age and tenure of the workforce, a quadratic polynomial in the number of sampled workers in the data, and 14 dummy variables for firm size.



(a) Regions

(b) Sectors

Figure 1: First stage

ϵ_{jt} are error terms.

OLS estimation of (7) and (8) would likely provide biased estimates if regions and firms with ex-ante higher levels of sorting are more exposed to automation or if omitted factors bias the OLS estimates. Therefore, following a standard approach in the literature (Acemoglu and Restrepo, 2020; Autor et al., 2013), I use the region- and firm-level robot exposure in countries different than Italy as instruments, i.e.

$$\text{IV Region exposure}_{\ell t} = \sum_s \frac{L_{\ell s}^{t_0}}{L_\ell^{t_0}} \times \frac{\widetilde{R}_s^{t-1}}{L_s^{t_0}} \tag{9}$$

$$\text{IV Firm exposure}_{S(j)t} = \frac{\widetilde{R}_{S(j)}^{t-1}}{L_{S(j)}^{t_0}} \tag{10}$$

where $\widetilde{R}_{S(j)}^{t-1}$ is the lagged average number of industrial robots in seven European countries (Germany, France, Spain, Sweden, Denmark, Finland and the UK). Two stage least squares estimates of (7) and (8) would provide unbiased estimates of the causal effect of automation on sorting provided that the instruments capture changes in robot exposure that are common across markets but uncorrelated with geographical or firm-level specific shocks.

Figure 1 plots the log average number of robots by region, against the instrument in panel (a). Panel (b) does the same using sectors. In other terms, the figures show the first stage relationship for both empirical strategies at the regional level and firm level. The association between the number of robots and the instrument is positive and strongly significant. The elasticities are, respectively, 0.78 and 0.95 in panels (a) and (b) and both are significant at 1% level.

4 The Impact of Automation on Sorting

4.1 AKM Estimation

As detailed in Section 3, I estimate the AKM two-way fixed effects model on the largest connected sample of workers and firms. I use the whole period 1985-2016 for the estimation, although only years 1993-2016 will be used in the final analysis. This choice is made in order to avoid a loss in efficiency due to the exclusion of job mobility episodes from the sample. Indeed, Abowd et al. (2002) show that identification of firm effects is guaranteed by job mobility within connected sets of workers and firms. Crucially, the identification rests on the assumption of conditional random mobility, i.e.

$$E\left(\varepsilon_{it}|\alpha_i, \psi_{J(i,s)}, X_{is}\right) \ \forall\ s, t.$$

Provided that this assumption is fulfilled in the data, OLS estimation of (1) gives consistent estimates of worker and firm fixed effects. Conditional random mobility implies that, conditional on workers and firms' fixed and time-varying characteristics, mobility is not determined by idiosyncratic shocks to workers or firms. I leave to Appendix A a more thorough discussion of the validity of such assumption. Overall, I conclude that conditional random mobility holds in the data.

One additional concern with the estimation of equation (1) is the limited mobility bias, as explained in Andrews et al. (2008). When the number of movers per firm is low, there is a negative bias in the estimation of covariance terms between firm and worker components. This is particularly important as one of the outcome variables is the correlation between worker and firm components. To partially address this issue, I restrict the sample to firms with at least 10 person-year observations over the period of analysis. Moreover, the measurement error in sorting on the left hand side of equation (7) would only affect the precision of the estimates, provided that the bias in the calculation of sorting is uncorrelated with automation adoption.

Details about the estimation of the AKM two-way model are reported in Table 2, panel A. The R-squared is 0.82 and the root mean squared error is 0.19. The standard deviation of worker effects is larger than the standard deviation of firm effects. The correlation between worker and firm effects is small but positive, indicating some degree of positive assortative matching. Panel B reports a decomposition of the variance of wages:

$$\begin{aligned} Var(w_{it}) =& Var(\alpha_i) + Var(\psi_{J(i,t)}) + Var(X'_{it}\beta) + Var(\varepsilon_{it}) \\ &+ 2 \times Cov(\alpha_i, \psi_{J(i,t)}) + 2 \times Cov(\alpha_i, X'_{it}\beta) + 2 \times Cov(\psi_{J(i,t)}, X'_{it}\beta) \end{aligned}$$

The variance of log daily wages is 0.2. The variance of worker effects explains 52% of it, whereas the variance of firm effects explains 27% of the variance of wages, indicating the importance of between-firm pay differences. The covariance between worker and firm effects explain very little of the variance of log wages, around 2%.

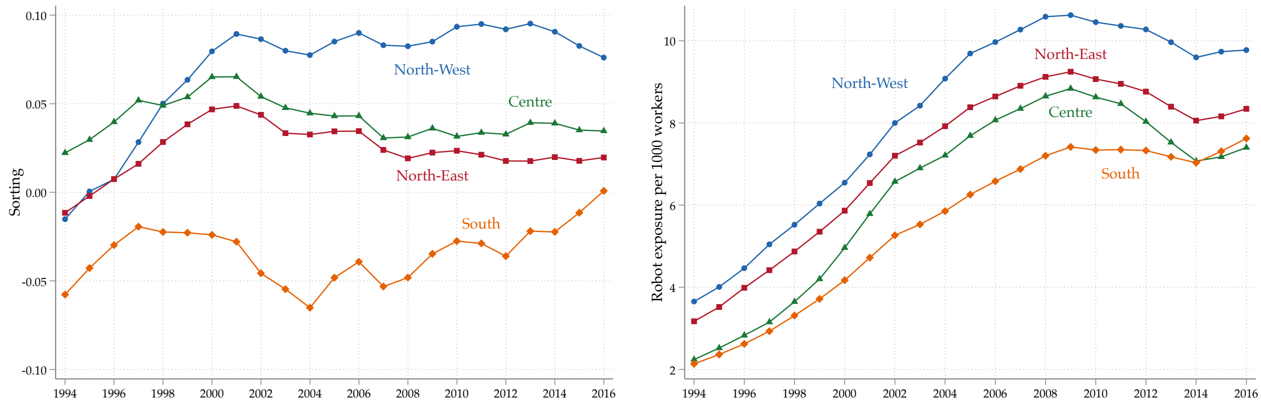
Table 2: AKM estimation and variance decomposition

Panel A. AKM estimation results		
R-squared	0.820	
Root MSE	0.191	
Standard deviation log daily wage	0.450	
Standard deviation α	0.324	
Standard deviation ψ	0.234	
Standard deviation $X\beta$	0.135	
Standard deviation residual	0.191	
Corr(α, ψ)	0.028	
Corr(α, $X\beta$)	-0.167	
Corr(ψ, $X\beta$)	-0.029	
Panel B. Variance decomposition		
Variance log daily wage	0.202	100%
Variance α	0.105	52%
Variance ψ	0.055	27%
Variance $X\beta$	0.018	9%
Variance residual	0.036	18%
$2 \times$ Cov(α, ψ)	0.004	2%
$2 \times$ Cov(α, $X\beta$)	-0.015	-7%
$2 \times$ Cov(ψ, $X\beta$)	-0.001	0%

4.2 Descriptive Evidence on the Evolution of Sorting and Automation Adoption Over Time

Although the correlation between worker and firm effects is low and little of the variance of wages is explained by their covariance, there are considerable differences in sorting between regions and over time. Figure 2a shows that sorting is positive and considerably larger in regions in the North-West (the more productive ones) than in the rest of the country. Moreover, there has been a steep increase in sorting between 1994 and 2001 and, after that, it remained constant. Regions in the Centre and North-East see a similar, but less steep, trajectory before 2001 and, after that time, sorting remains positive but closer to 0. In the South, instead, sorting is negative for most years in the sample, indicating the presence of negative assortative matching, but with an upward trend between 2004 and 2016.

How do differences in sorting between regions and over time compare to changes in automation adoption? The pattern depicted in Figure 2a shows an association with the one in Figure 2b, which depicts the evolution of automation adoption, computed as in (5), by macro-region over time. Automation adoption, in terms of the number of robots per 1000 workers, is higher in the North of the country than in the South and it flattens after 2007.



(a) Sorting (b) Automation adoption

Figure 2: Evolution of sorting and automation adoption over time and by macro-area

4.3 Causal Effect of Automation on Sorting

This section provides estimates of the causal impact of automation adoption on sorting, using the empirical strategy outlined in Section 3.

4.3.1 Regional level analysis

Table 3 shows the regression results. Columns 1-3 report OLS estimates of equation (7), whereas columns 4-6 report IV results. Panel A reports coefficients for all sectors in the economy. OLS estimates are not significant and small in columns 1-3, where column 1 includes no controls, column 2 controls for pre-determined area characteristics (blue collar share, female share, manufacturing share, skilled workers share, and the initial correlation between worker and firm effects) and macro-area dummies (North-West, North-East, Centre, South), column 3 includes region fixed effects. Instrumental variable estimates are reported in columns 4-6 (where column 4 has no controls, column 5 controls for region characteristics and macro-area fixed effects, column 6 includes region fixed effects) and are positive and statistically significant. First-stage F-statistics are reported in the bottom part of the table and show that the first stage effect is strong and significant. The most conservative estimates in columns 5 and 6 indicate that one additional robot per 1000 workers increases sorting by 0.33 log points or by approximately 9.3% relative to the mean of sorting across regions and time. Panel B shows results for industry only (i.e. mining, manufacturing, construction and electricity). Interestingly, the effects vanishes if one focuses on this restricted set of sectors. Hence, automation adoption entailed a reallocation of workers out of manufacturing into services, such that high-wage workers ended up working into high-wage firms in regions more exposed to the automation wave and over time. Moreover, the results are robust to exclusion of Piedmont and Marche, the two regions with the highest level of automation adoption, as shown in Figure B.1

Figure 3 shows heterogeneous effects by worker's characteristics. Specifically, it reports results by gender, age and occupation, where sorting is computed as the correlation between

Table 3: Impact of automation on sorting, regional-level analysis

	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
				Panel A: All sectors		
Regional exposure	0.0044	0.0019	0.0007	0.0061***	0.0033**	0.0033***
	(0.0032)	(0.0014)	(0.0033)	(0.0021)	(0.0013)	(0.0011)
Mean of dep. var.	0.035	0.035	0.035	0.035	0.035	0.035
R^2	0.16	0.83	0.43	0.15	0.83	0.42
				Panel B: Industry		
Regional exposure	0.0004	-0.0008	-0.0018	0.0029	0.0007	0.0008
	(0.0048)	(0.0026)	(0.0023)	(0.0030)	(0.0025)	(0.0008)
Mean of dep. var.	-0.030	-0.030	-0.030	-0.030	-0.030	-0.030
R^2	0.11	0.76	0.45	0.09	0.75	0.44
Kleibergen-Paap F-stat	-	-	-	424.63	153.59	174.89
# geo. fixed effects	0	4	20	0	4	20
Controls	No	Yes	No	No	Yes	No
Observations	460	460	460	460	460	460

Notes. The table reports OLS (columns 1-3) and IV (columns 4-6) estimates of equation (7). Columns 1 and 4 have no controls; columns 2 and 5 control for the region-level blue collar share, female share, manufacturing share, skilled workers share, the initial correlation between worker and firm effects, and macro-area dummies (North-West, North-East, Centre, South); columns 3 and 5 report control for region fixed effects. All regressions are weighted for total employment in the region in 1991. Cluster-robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

worker and firm effects conditional on belonging to a specific category. Looking at the results by gender, while the effect is negative for women – both in the sample with all sectors and in the one considering only industry –, the effect is positive for men. At the same time, for workers younger than 35 years old the effect is null in all sectors and negative in industry, whereas for older workers the effect is generally positive (although significant only in all sectors). Finally, when looking at differences by occupation, the effect is positive and significant only for white-collar workers.

4.3.2 Firm level analysis

Table 4 reports OLS and IV estimates of equation (8) in panels A and B, respectively. Additionally to estimating the effect on the mean and the standard deviation of worker effects, I also use the mean and standard deviation of wages as an alternative measure of worker's quality and therefore of matching. All regressions control for firm fixed effects and the following set of covariates: firm-level shares of female, part-time and blue-collar workers, average age and tenure of the workforce, a quadratic polynomial in the number of sampled workers in the data, and 14 dummy variables for firm size. The Table shows that there is only a modest and insignificant effect of automation adoption at the sector level on match-

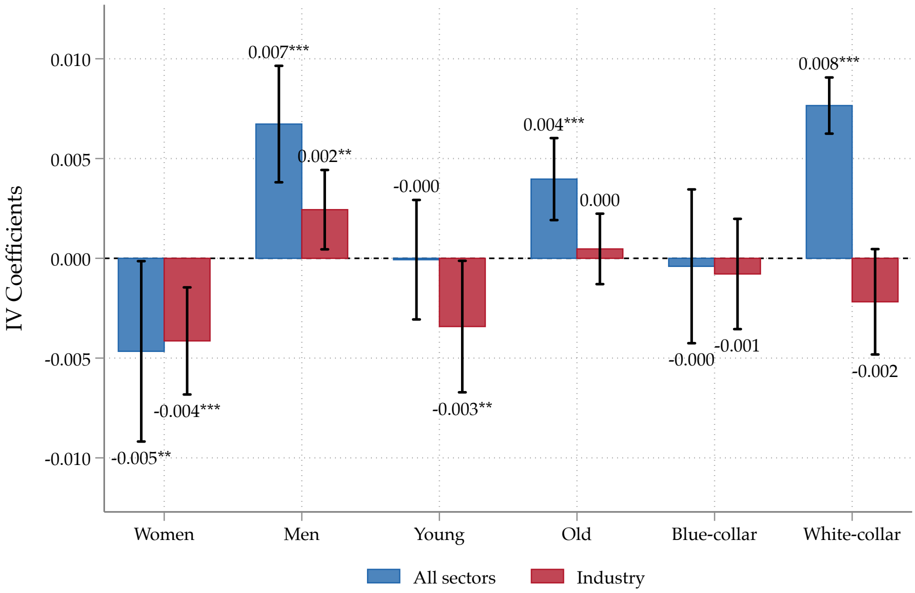


Figure 3: Heterogeneous effects

ing, as measured by the mean or dispersion of worker effects (columns 1 and 3). The IV estimates suggest that there is a significant effect on wages (column 2), which is however very tiny in economic terms: one additional robot per 1000 worker in a given sector raises firm-level wages by 0.03 log points.

The effect of automation may differ between low-wage and high-wage firms. For the latter group, indeed, the benefit of hiring a high-wage worker may be higher in response to increased automation. In fact, Figure 4 shows heterogeneous effects by quartiles of the firm fixed effects distribution. The figure highlights that only firms in the top quartile of the firm effects distribution see a positive and significant increase in the average quality of their workforce, as measured by both the average worker effect and average log daily wage. The dispersion of worker effects and wages increases, too, but the effect is not statistically significant. It may be that firms match with better and increasingly similar workers: in this case, average worker quality would increase, while dispersion would remain constant or even decrease. The raise in average worker quality in top quartile firms therefore confirms again that in response to automation the allocation of workers into firms changes in the direction of positive assortative matching.

5 Conclusion

This paper provides evidence on the effect of automation on labor market sorting and matching. Combining automation data at the sector level and matched employer-employee data over the period 1993-2016 for Italy, the paper shows that automation contributes to positive assortative matching, i.e. the tendency of high-wage workers to be employed by high-wage firms. To prove this point, I first set up an AKM two-way fixed effects model that recovers

Table 4: Effect of automation on matching, firm level regressions

	(1) Mean WFE	(2) Mean wage	(3) S.D. WFE	(4) S.D. wage
		Panel A: OLS		
Firm exposure	-0.0033	-0.0029	0.0044	0.0127**
	(0.0036)	(0.0061)	(0.0042)	(0.0063)
		Panel B: IV		
Firm exposure	0.0053	0.0267**	0.0080	0.0094
	(0.0070)	(0.0123)	(0.0074)	(0.0111)
Mean of dep. var.	-0.0278	4.5257	0.2243	0.2857
Kleibergen-Paap F-stat	5,311.8	5,311.8	5,311.8	5,311.8
Observations	158,389	158,389	158,389	158,389

Notes. The table reports OLS (panel A) and IV (panel B) estimates of equation (8). All regressions include firm and year fixed effects, and control for firm-level shares of female, part-time and blue-collar workers, average age and tenure of the workforce, a quadratic polynomial in the number of sampled workers in the data, and 14 dummy variables for firm size. Cluster-robust standard errors are reported in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

estimates of firm and worker effects. With these estimates at hand, I run a region-level analysis where I study the impact of automation exposure on the labor market sorting in each Italian region over time, where sorting is measured as the correlation between worker and firm effects at region-year level. I find that one additional robot per 1000 workers increases sorting by 9.3% relative to the mean. The effects are heterogeneous according to worker characteristics. They are larger in magnitude for male workers, individuals older than 35 years old and workers employed in white-collar occupations. I then conduct a firm level analysis, where I study the effect of automation exposure at the sector level on the quality of workers the firm matches with. On average, firms in sectors more exposed to automation do not change their worker quality, as measured by the average worker effect, but do increase wages at the firm level by a tiny margin. However, when splitting the sample according to the quartiles of the firm effect distribution, I find that only firms in the top quartile match with better workers in response to higher automation exposure, therefore confirming positive assortative matching in response to increased automation exposure. These results contribute to the understanding of the effects of automation on labor demand and to the analyses of the mechanisms that determine labor market sorting, which ultimately reflect in higher earnings inequality.

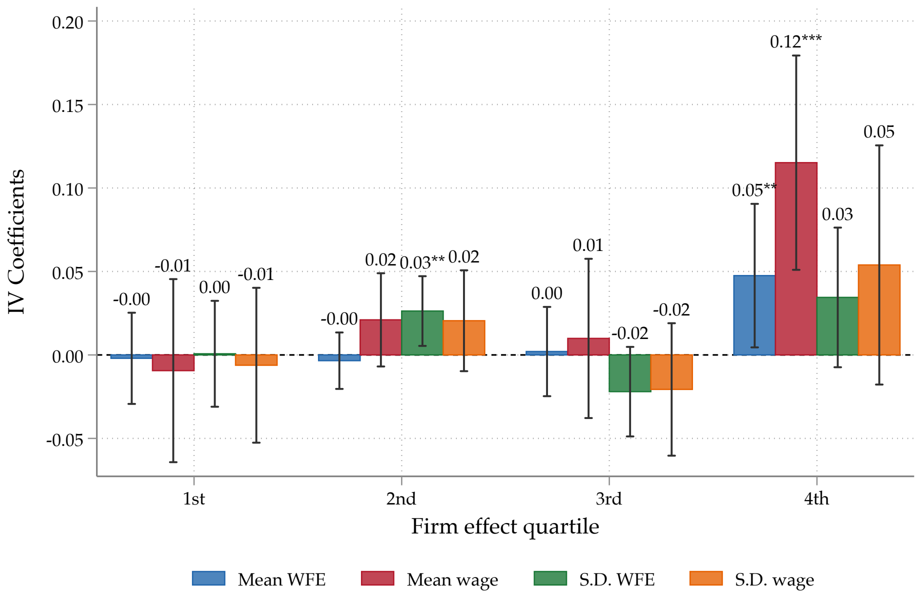


Figure 4: Heterogeneous effects by firm effect quartile

References

Abowd, J., Creecy, R. H., and Kramarz, F. (2002). Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal employer-household dynamics technical papers, Center for Economic Studies, U.S. Census Bureau.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. and Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2):3–30.

Acemoglu, D. and Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*, forthcoming.

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2008). High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3):673–697.

Autor, D. H., Dorn, D., and Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, 103(6):2121–68.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics*, 90(2):300–323. Data and Replication Files.

Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, 34(S2):S67–S97.

Belfield, C., Blundell, R., Cribb, J., Hood, A., and Joyce, R. (2017). Two Decades of Income Inequality in Britain: The Role of Wages, Household Earnings and Redistribution. *Economica*, 84(334):157–179.

Bessen, J., Goos, M., Salomons, A., and van den Berge, W. (2019). Automatic Reaction – What Happens to Workers at Firms that Automate? CPB Discussion Paper 390, CPB Netherlands Bureau for Economic Policy Analysis.

Bombardini, M., Orefice, G., and Tito, M. D. (2019). Does exporting improve matching? Evidence from French employer-employee data. *Journal of International Economics*, 117(C):229–241.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, 36(S1):S13–S70.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics*, 128(3):967–1015.

Colantone, I., Matano, A., and Naticchioni, P. (2019). New imported inputs, wages and worker mobility. *Industrial and Corporate Change*, 29(2):423–457.

Dauth, W., Findeisen, S., Suedekum, J., and Woessner, N. (2018). Adjusting to Robots: Worker-Level Evidence. Opportunity and Inclusive Growth Institute Working Papers 13, Federal Reserve Bank of Minneapolis.

Devicienti, F., Fanfani, B., and Maida, A. (2019). Collective bargaining and the evolution of wage inequality in italy. *British Journal of Industrial Relations*, 57(2):377–407.

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisiting the German Wage Structure. *The Quarterly Journal of Economics*, 124(2):843–881.

Franzini, M. and Raitano, M. (2019). Earnings inequality and workers' skills in Italy. *Structural Change and Economic Dynamics*, 51:215 – 224.

Goldschmidt, D. and Schmieder, J. F. (2017). The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure*. *The Quarterly Journal of Economics*, 132(3):1165–1217.

Graetz, G. and Michaels, G. (2018). Robots at Work. *The Review of Economics and Statistics*, 100(5):753–768.

Katz, L. F. and Autor, D. H. (1999). Changes in the Wage Structure and Earnings Inequality. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics, vol. 3A*, pages 1463–1555.

Smith, B. (2018). The Role of Labor Market Entry and Exports in Sorting: Evidence from West Germany. Working paper.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2018). Firming Up Inequality. *The Quarterly Journal of Economics*, 134(1):1–50.

A Non Parametric Tests of Conditional Random Mobility

This section provides evidence on the conditional random mobility assumption. Following Card et al. (2013), there are three main channels through which conditional random mobility may be violated. First, workers employed at firms that are experiencing negative shocks may decide to move to firms that are experiencing positive shocks: this generates correlation between firm components of wages and the probability that worker i is employed at firm j at time t. If this is the case, workers would experience a drop in wages before the move, and a sudden rise in pay after. Figure A.1 shows that this is not the case. Specifically, I build a sample of moves and compute mean daily wages associated with changes from the first and the last quartile of firm effects.[1] There are no changes in the evolution of mean wages before or after the move.

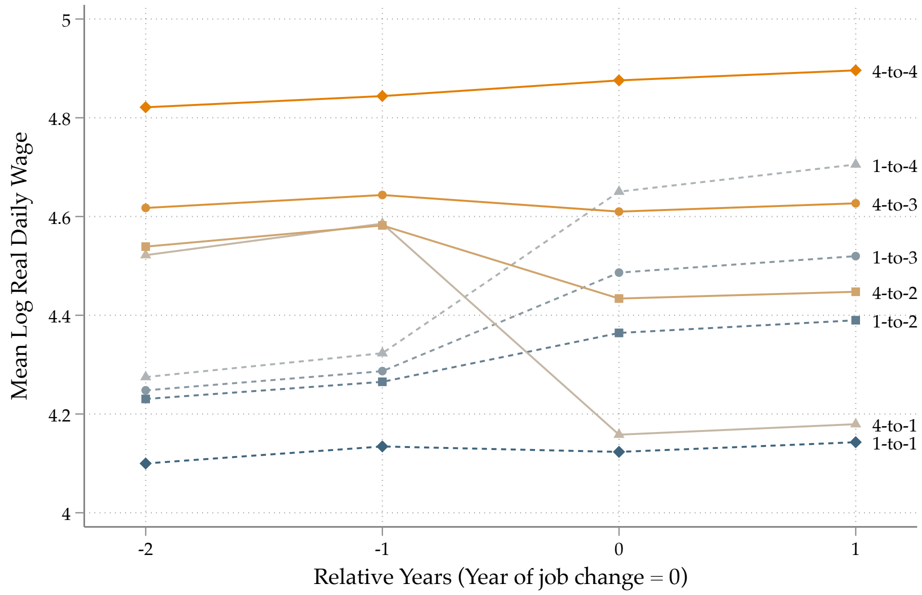
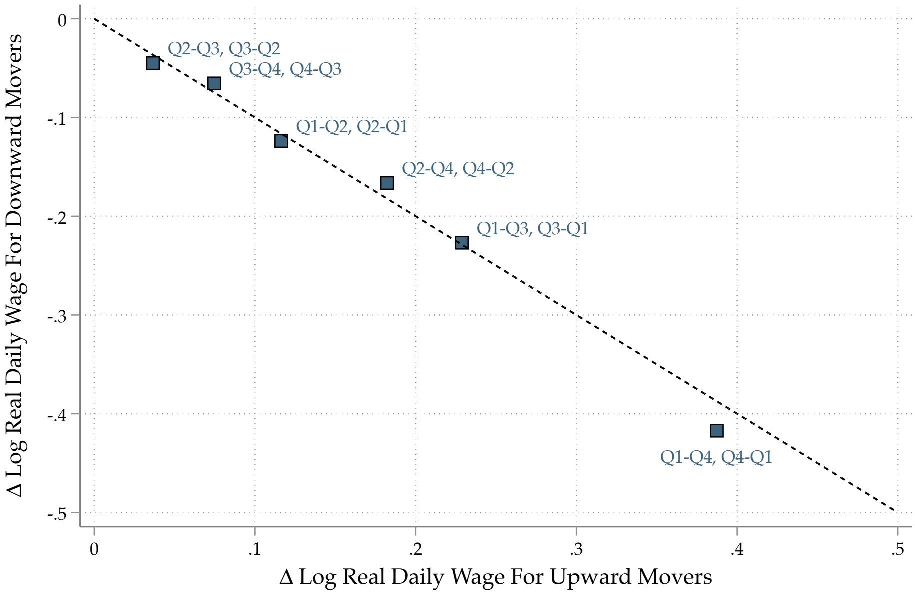


Figure A.1: Mean weekly earnings of movers across firm effects quartiles

A second threat to identification comes from the presence of match effects, if workers decide to move because they think that joining a new firm would deliver a better match between their personal characteristics and the firm characteristics compared to the firm of origin. This violation implies that the match component is correlated with the probability that worker i is employed at firm j at time t. In the presence of correlation, movers would experience in any case a wage gain, irrespective of whether they move from a high-wage to a low-wage firm, or the opposite. On the other hand, if match effects are unimportant in determining mobility, then the wage gain associated with moves from low- to high-wage firms should be roughly comparable in magnitude to the earnings loss for moves in the

[1] I identify low-wage and high-wage firms on the basis of the quartiles of the estimated firm effects. I then assign each job mover to the corresponding quartile of the origin and destination firm. In this way I identify sixteen cells of movers, each one corresponding to the pair origin-destination quartile (4×4 cells). Within each cell, I compute the mean log daily wages of movers. I just retain movers that are continuously observed in the two years prior to the move and in the two years after.

Figure A.2: Adjusted change in wages of symmetric job moves across firm effects quartiles



opposite direction. This symmetry in gains and losses with each opposite move is better assessed examining the magnitudes of such changes over the entire 4 year period under analysis and for all possible moves, looking at the difference in wages from the first period considered (2 years prior to the move) to the last period (one year after). This boils down to comparing the overall wage change (wages one year after *minus* wages two years before) for opposite moves.[2] The comparisons are displayed in Figure A.2, which plots the adjusted wage changes[3] for downward movers against the adjusted wage changes for upward movers. Opposite moves display the expected degree of symmetry, that is, they are in all cases of opposite sign. Moreover, all scatter points cluster very close to the 45 degrees line, meaning that each symmetric move, both upward and downward, generates a wage change of a similar magnitude, supporting the assumption of symmetry.

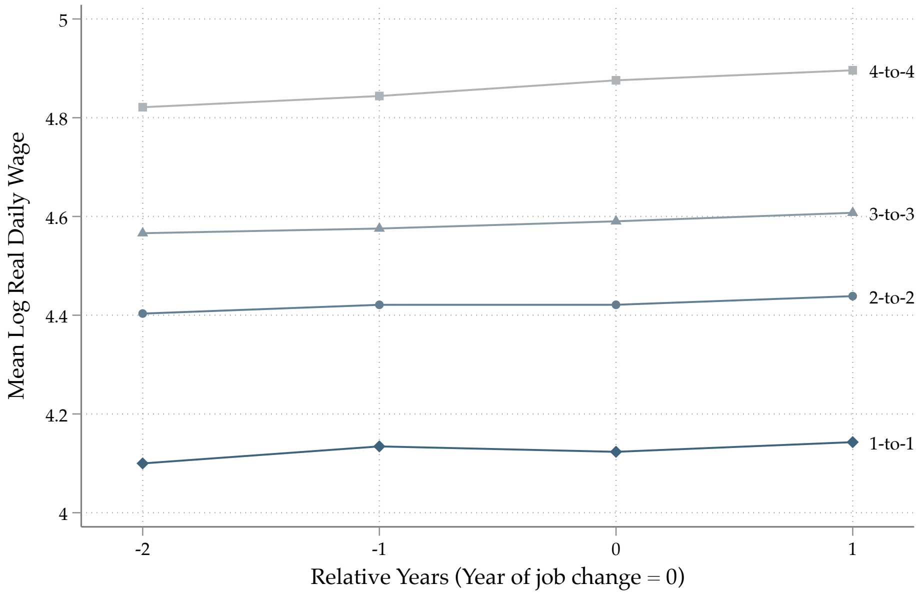
As an additional check, Figure A.3 reports the wage evolution for movers within the same quartile in the origin and destination firms. If it is true that there are no match effects in mobility, then these movements should be characterised by almost no wage gains. This is indeed the case: both panels show that the wage evolution is basically flat for within-quartile movements. This is clearly inconsistent with specific worker-firm match gains related to job changes.

A last threat to the identification of firm effects comes from individual transitory shocks, that generate correlation between the transitory component of wages and the probability that worker i is employed at firm j at time t. If workers are experiencing an increase in their earnings before the move because of some productivity premium associated to a transitory

[2]Opposite moves are those from quartile k to quartile j, and the other way around.

[3]Adjusted wage changes equal raw wage changes minus the wage change for within-quartile movers: that is, I subtract the change for movers from quartile q to quartile q from the raw change for movers from quartile q to quartile q', with $q \neq q'$.

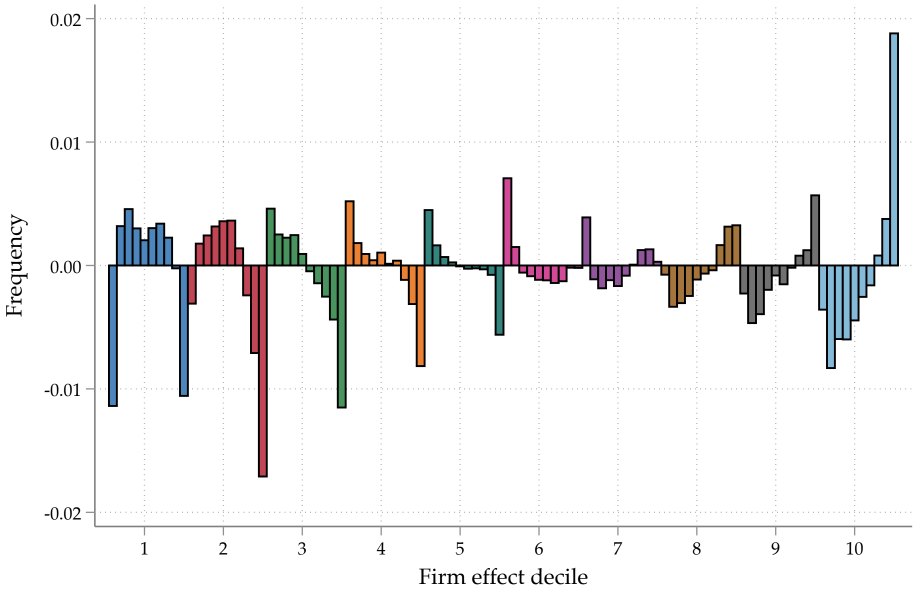
Figure A.3: Mean daily wages of movers within same firm effects quartiles



change in their characteristics or to some of their skills showing up after an accumulation period, then they might move to other firms that reward these characteristics more, with a larger gain from the move compared to that obtained in the origin firm. On the other hand, if the transitory shock is negative, workers might experience a wage decline in their origin firm and therefore move to firms that would limit such decline, because better suited to reward their characteristics. We can refer again to Figure A.1, where, if mobility is driven by individuals recognising their higher (lower) productivity we should see unusual wage growth (decline) before the move for people moving towards the top and unusual wage decline (growth) for people moving in the opposite direction. Nothing like that happens in the data.

As a final check, following again Card et al. (2013), I examine residuals from model (1) for different groups of individual effects in different groups of firm effects. Namely, I define deciles of both person and firm effects and compute the mean estimated AKM residuals in each of the 100 cells defined by the combination of worker and firm deciles. If the empirical model is incorrectly specified, because, for instance, it is missing some important match component between specific individuals and firms, one would expect to find high mean residuals in those cells that are most threatened by misspecification. Figure A.4 plots the mean residuals for each of the person-firm cells. The deviations are really small in magnitude and exceed 1 log point only in few cases. Overall, there is no evidence against the conditional random mobility assumption.

Figure A.4: Mean AKM residuals across deciles of person and firm effects



B Additional Tables and Figures

Table B.1: Sectors included in IFR data

Sector	Nace	Group
Agriculture, forestry, fishing	A-B	1
Mining and quarrying	C	2
Food and beverages	10, 11, 12	3
Textiles	13, 14, 15	4
Wood and furniture	16	5
Paper	17, 18	6
Pharmaceuticals, cosmetics	19	7
Other chemical products n.e.c.	20, 21	8
Rubber and plastic products (non-automotive)	22	9
Chemical products, unspecified	229	9
Glass, ceramics, stone, mineral products (non-automotive)	23	10
Basic metals	24	11
Metal products (non-automotive)	25	12
Electronic components/devices	260	13
Semiconductors, LCD, LED	261	13
Computers and peripheral equipment	262	13
Info communication equipment, domestic and prof.	263	13
Medical, precision, optical instruments	265	13
Household/domestic appliances	275	14
Electrical machinery n.e.c. (non-automotive)	271	14
Electrical/electronics unspecified	279	14
Industrial machinery	28	15
Metal, unspecified	289	15
Automotive	29	16
Other vehicles	30	17
Electricity, gas, water supply	E	18

Notes. The table reports the sectors included in IFR data, together with their classification in Nace Rev. 2 and how the sectors have been grouped for the purpose of the present paper.

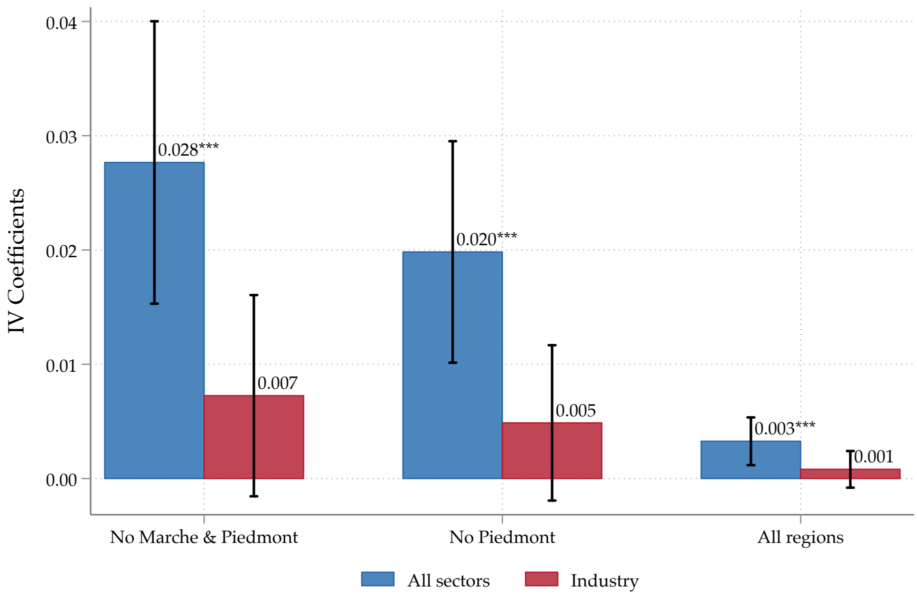


Figure B.1: Effects of automation on sorting, excluding Piedmont and Marche from the sample