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

Ester Faia
Goethe University

Gianmarco Ottaviano
Bocconi University

Saverio Spinella
New York University

November 15, 2022

# Overview: What We Do

- **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**: vertical (SBTC, RBTC) and horizontal (CBTC) specialization
- We analyze the impact of **robot adoption** on wage inequality through SBTC/RBTC/CBTC **in Italy**
- We exploit the universe of contracts between firms and workers across Italian **local economies** ('provinces') from 1983 to 2020 + Robot acquisition from the International Federation of Robotics (IFR)

# Overview: Why Italy?

- Italy is an **interesting case**: G7 country, large industrial sector (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)
- **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

# Overview: Descriptive Statistics

- All measures 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 30 percent respectively

- Simple wage variance decomposition across firms, occupations ('tasks') and sectors shows 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, the **between-firm, within-sector and within-task** component is the single **most important** component of overall wage variance

- Hence, for a given task in a given sector, the main driver of wage inequality is **match heterogeneity** for the same task across firms within the same sector

# Overview: Research Question

- These findings do not necessarily imply an increase in the **assortativity** between workers' specialized skills and firms' specific tasks
- 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**

# Overview: Research Strategy

- **Two stages** involving the estimation of **how sorting evolves** through time and the assessment of the **causal effect of robot adoption** on that evolution

- The **first stage** separately **identifies sorting** from firm and worker characteristics in matched employer-employee wage data allowing for complementarity-induced non-linearities and solving the incidental parameter bias ('low mover bias') in the wake of **Bonhomme, Lamadon and Manresa (2019; BLM)**

- To reduce the number of estimated 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

# Overview: Research Strategy (1st Stage)

- Estimation is then performed with **firm class fixed effects** rather than individual firm fixed effects
- **Worker** heterogeneity is captured through **random fixed effects** after reducing its dimensionality by approximating the workers' distribution via a finite support population density
- This specification is used to collapse worker heterogeneity in a limited number of **probabilistic 'types'**
- The end 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

# Overview: BLM Extension 1

- 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

- Each point estimate requires a **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 (BLM use a **four-year window** for their dynamic model)

- Exploiting the longer time series dimension of our data, we obtain time varying estimates of unobserved firm and worker characteristics re-estimating them every second year over **partially overlapping 4-year intervals**

# Overview: BLM Extension 2

- 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 compute the **probabilities** that workers belong to the different worker types and associate them with their highest probability type. Worker types are assigned to provinces by the **addresses** of the workers they include

# Overview: Research Strategy (2nd Stage)

- 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 instrument 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**)

# Overview: Results (1st Stage)

- In the first stage we find that the **correlation** between unobserved firm-class and worker-type characteristics is **positive** ('sorting') 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 both **'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

# Overview: Results (1st Stage - Cont.)

- 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

# Overview: Results (1st Stage - Cont.)

- 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, **horizontal task specialization**
- 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**

# Overview: Results (2nd Stage)

- In the second stage we find that **specialization is caused by robot adoption** as our shift-share instruments foster both sorting and segregation
- We can therefore conclude that our econometric analysis reveals the presence of **both** 'routine-biased technological change' (RBTC), new technology decreases the relative demand for workers in traditional routine tasks, and by CBTC, accoridng to which new technology requires workers with specialized knowledge independently of their tasks being more or less routine intensive.

# Overview: Conceptual Framework

- We show 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, Postel-Vinay and Jean-Marc Robin (2006) and Bagger and Lentz (2019)

- Extended with endogenous job search and logsupermodular flow surplus (complemetarity). We model the **impact of automation** as strengthening production complementarities between workers' specialized skills and firms' specific tasks and disproportionately reducing the search frictions for higher worker types

- We show that **stronger complementarities and lower search frictions increase wage dispersion** with a relevant role played by both between and within firm class dispersion

# Related literature and contribution

- Firm and worker effects on wages: Abwod, Kramarz and Margolis 1995.

$$y_{ijt} = \psi_i + \theta_j + \varepsilon_{ijt} \tag{1}$$

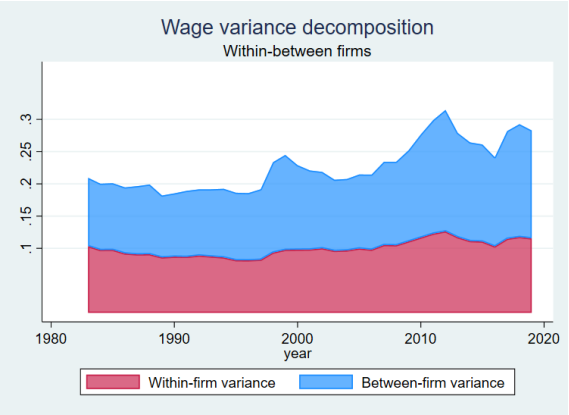
- Andrews et al. 2012 underscores AKM estimates instability. They rely heavily on the connectedness of the network

- New methodology: Mixture-models (random-fixed) Bonhomme, Lamadom and Manresa 2019

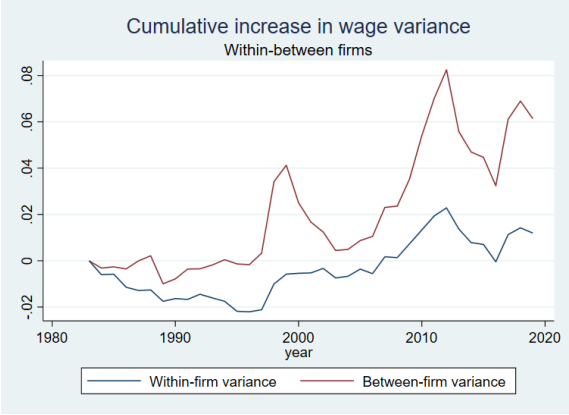
- Literature on automation (see Acemoglu and Restrepo 2020 among others) on role of capital substitution. Our focus is on specialization (task-biased)

- Model: build on Cahuc, Postel-Vinay and Robin 2006

# Motivating evidence - Between-within firm inequality

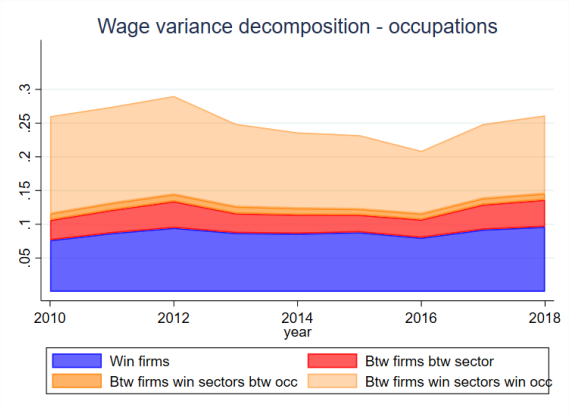


# Motivating evidence - Between-within inequality, Cumulative



# Motivating evidence - the role of tasks

Further cut into sectors and occupations → most of the between-firm variance takes place across occupations.



# Two-sided heterogeneity - BLM

- **Mincer residualization**:

$$y = \alpha + \beta_1(\text{age} - 40)^2 + \beta_2(\text{age} - 40)^3 + \beta_3(\text{tenure}) + I(\text{sector} = s) + \varepsilon \quad (2)$$

- **Finite mixture maximum likelihood in two stages**: - Stage 1: k-means to cluster firms -Stage 2: Estimate random-fixed effects interacted model

$$y_{ijt} = \theta_j + \psi_{j(i,t)}\Theta_i + \bar{X}_{it}\bar{\beta} + \varepsilon_{ijt} \quad (3)$$

  $\Theta_i$ worker effect, and the vector $[\theta_j, \psi_{j(i,t)}]$ is the two-dimensional firm effect
- Estimates worker random effects by MLE
- Eliminates incidental parameter bias and accounts for non-linearities

# Two-sided heterogeneity - BLM

- Create time-series cross section of sorting estimates
- **Bayesian assignment**:

$$\arg\max_{\alpha*} p(\alpha = \alpha^*|y_i, k_i) = \arg\max_{\alpha*} \frac{f_{k_i,\alpha^*}(y_i)\ q_{k_i}(\alpha^*)}{\sum_j^J f_{k_i,\alpha_j}(y_i)\ q_{k_i}(\alpha_j)} \tag{4}$$

$\alpha$ worker type, $f(.)$ distribution of wages, $q(.)$ estimated proportions of workforce-type

# Decomposition of Effects

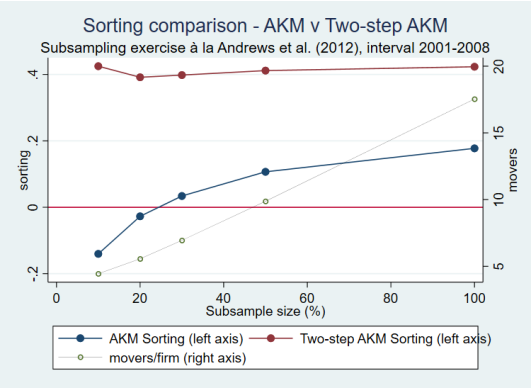
- Variance of effects and **sorting**: $y$ log weekly wage, $\psi$ worker FE, $\theta$ firm FE

$$v(y) = v(\psi) + v(\theta) + v(\varepsilon) + 2\text{corr}(\psi, \theta)\text{sd}(\psi)\text{sd}(\theta) \tag{5}$$

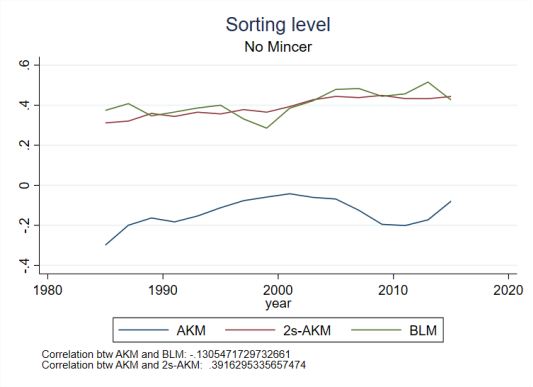
- Alternatively:

$$v(y) = v(\psi - \bar{\psi}^j) + v(\varepsilon) + v(\theta) + 2\text{cov}(\bar{\psi}^j, \theta) + v(\bar{\psi}^j) \tag{6}$$

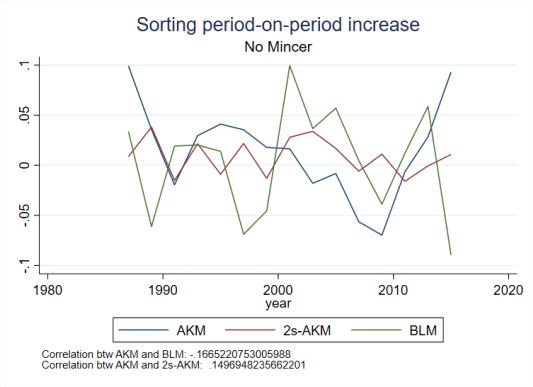
# AKM issues - Andrews et al 2012



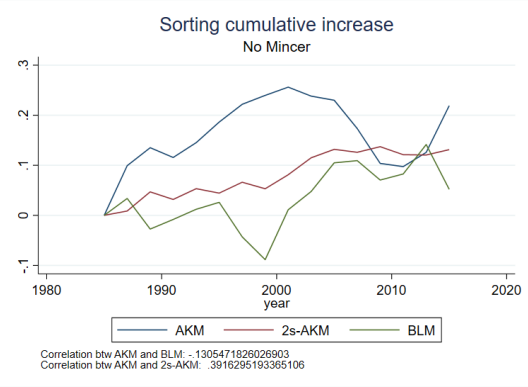
# AKM issues - Subsampling exercise



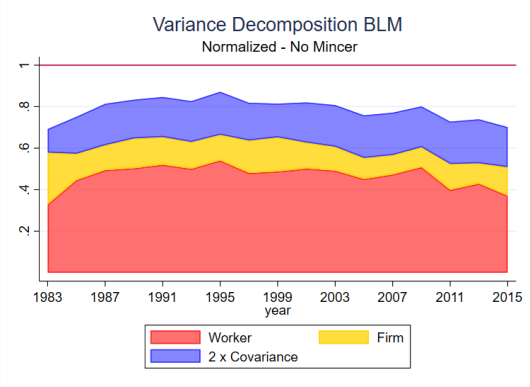
# AKM issues - Subsampling exercise



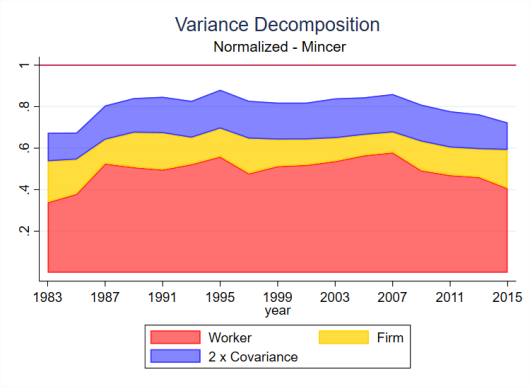
# AKM issues - Subsampling exercise



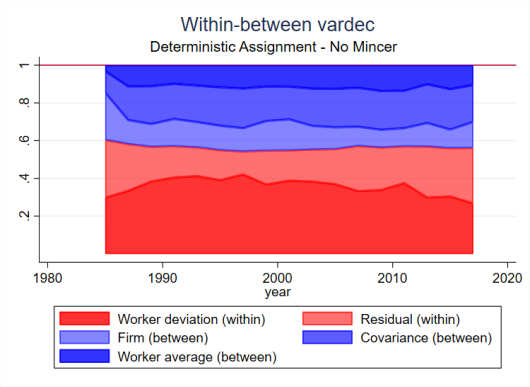
# BLM results - variance decomposition



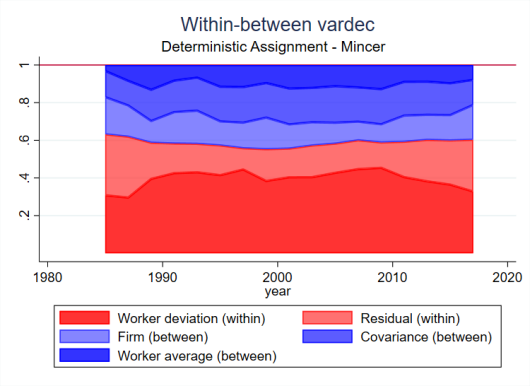
# BLM estimates - variance decomposition - Mincer



# BLM results - variance decomposition - No Mincer

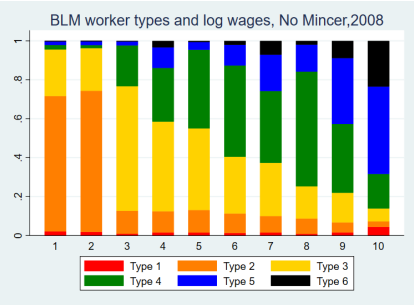


# BLM Estimates - Variance decomposition with segregation - Mincer

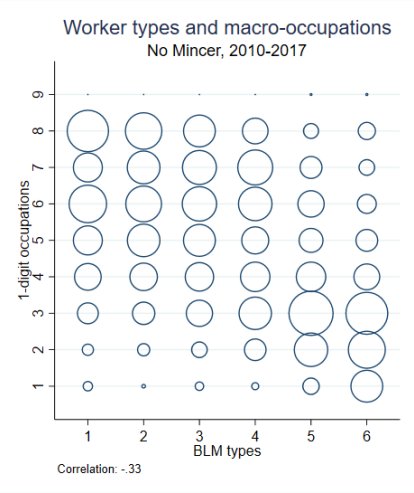


# BLM Estimates - Workforce composition

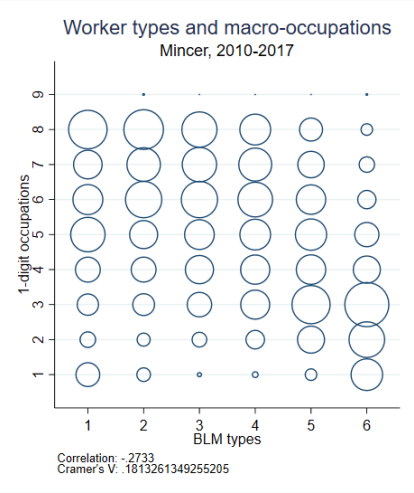
Workforce composition across firm clusters signal positive sorting



# BLM - worker types and tasks/occupations



# BLM - worker types and tasks/occupations



# Second Stage

$$\text{corr}(\psi, \theta)_{p,\tau} = \alpha + \beta(\text{automation}_{p,\tau-1}) + \sum_{h}^{\{25,50\,75\}} \gamma_h \text{HHI}^h_{p,\tau} + \delta_1 \text{share manuf} + \delta_2 \text{share serv} + \eta_\tau \times \zeta_{p\in m} + \varepsilon_{p,\tau}$$

$\text{HHI}^h$ is HHI of employment workers of the (sector x province); $\text{HHI}^h$ is $h$-th percentile. $\eta_\tau$ and $\zeta_{p\in m}$ are period and macro-area fixed effects.

# Measures of Automation

- **International Federation of Robotics**, 1995 to 2016, most world economies, level of two-digits NACE sectors:

$$\text{automation}_p = \sum_{s \in S} L_s^p \underbrace{\left( \frac{dM_s}{L_s} - \frac{dY_s}{Y_s} \frac{M_s}{L_s} \right)}_{\text{Sector's exposure}} \tag{7}$$

  $L$ employment, $\frac{dY_s}{Y_s}$ growth rate of value added, $M$ and $dM$ stock of robots and the  of installations

- **Instrument**:

$$\text{automation}_p = \frac{1}{N} \sum_{s \in S} L_s^p \sum_{c=1}^{N} \left( \frac{dM_s^c}{L_s^c} - \frac{dY_s^c}{Y_s^c} \frac{M_s^c}{L_s^c} \right) \tag{8}$$

  $c$ indexes 24 countries (all Europe, US, JP)

- **Automation meaasures** for all industries, with and without automobile. **IVs** with all countries, only US

# Second Stage - Result - No Mincer, All Industries

| 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Automation | 0.0237*** | 0.00262 | 0.00365 | 0.00329 | 0.0187*** | 0.00716* | 0.00947** | 0.0106** | 0.00184 | 0.0122 | 0.0117 | 0.0162* |
| | (0.00343) | (0.00270) | (0.00279) | (0.00283) | (0.00437) | (0.00398) | (0.00408) | (0.00422) | (0.0111) | (0.00958) | (0.0100) | (0.00987) |
| Observations | 883 | 883 | 883 | 883 | 883 | 883 | 883 | 883 | 883 | 883 | 883 | 883 |
| $R^2$ | 0.282 | 0.574 | 0.605 | 0.641 | 0.280 | 0.573 | 0.604 | 0.639 | 0.250 | 0.570 | 0.603 | 0.634 |
| 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 | | | | | 541 | 332.3 | 323 | 319 | 12.79 | 13.53 | 12.33 | 13.08 |

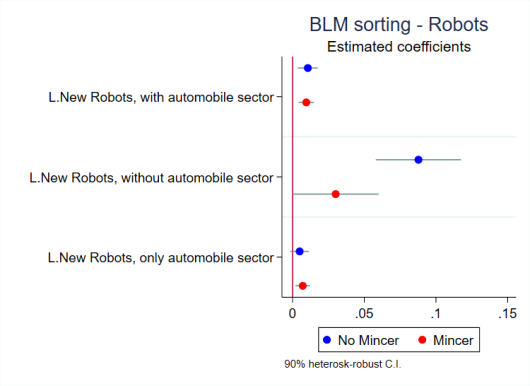
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

# Second Stage - Shift-share Robot Installations measure

Here we look at two moments of the distributions of our regressor.

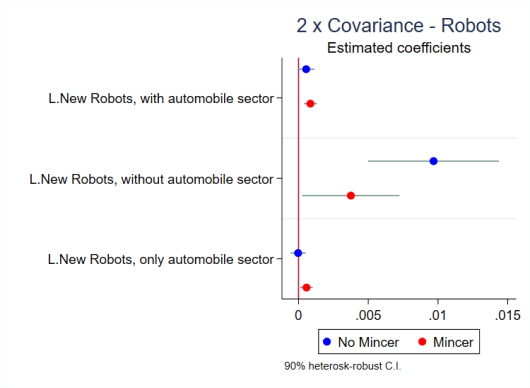
| | n obs | mean | var | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| All industries | 984 | .634 | .638 | .033 | .179 | .439 | .777 | 1.336 |
| Excluding automobile | 984 | .284 | .068 | .022 | .048 | .252 | .429 | .600 |
| Automobile only | 984 | .351 | .567 | -.042 | .009 | .086 | .380 | .988 |

# Automation and Complemenatarities

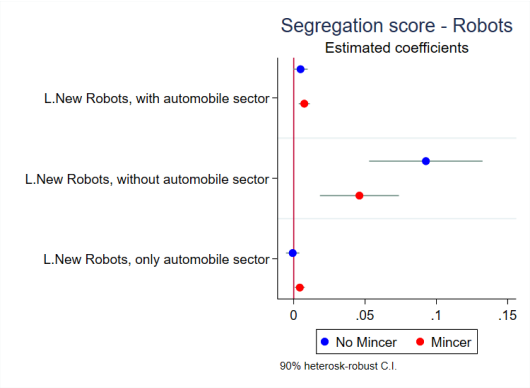



Coefficient for 90%, heteroskedasticity-robust confidence intervals. Three shift-share measures, two measures of sorting from raw earnings and Mincer wages.

# Automation and Covariance of Wage Decomposition



# Automation and Segregation Score




**Segregation Score:** segregation score $= \dfrac{\mathrm{var}(\theta^{\bar{j}})}{\mathrm{var}(\theta_i)}$

# Counterfactual for tech adoption

If we were to move a province from the 25th percentile to the 75th percentile of robot installations per worker, ceteris paribus, this would lead to:

- 9.0 % increase in average sorting
- 14.5 % increase in average covariance
- 19.2 % increase in average segregation score

# Model

- **Role of complementarities** in production in Cahuc, Postel-Vinay and Robin (2006)
- Two-sided heterogeneity, random matching, **on-the-job search** and Bertrand poaching and bargaining with employed workers as in Cahuc, Postel-Vinay and Robin (2006), extended with endogenous search intensity (endogenous mover choice) and log-supermodular production function (complementarties)
- Derive **wage equation and distribution**: sorting (interaction term) emerges from **rebargaining surplus**
- Model supports **BLM empirical design** for the estimation of unobserved firm and worker characteristics
- **W**e show that robot adoption, by rising complementarities, it affects the *interaction term*

# Conclusions

- Role of **core-task biased** technological change for wage distribution
- Rationalize the **between firm component** observed in data with sorting of workers based on specialization
- Identify robust **worker and firm effects** and relate estimated sorting and segregation to robots shift shares
- ***Automation*** *increases* ***specialization****, which in turn increases* ***wage dispersion***