How workers sorting into organizations and occupations amplifies between- and within-demographic group wage differences

UNIVERSITY OF TRENTO
Doctoral School of Social Sciences
SORBONNE UNIVERSITÉ
Gemass

Lucas SAGE

INED Big_Stat
16-11
-2021

From Between- to Within-Group Wage Differences

- Traditional accounts of wage inequality focus on **between** group wage differences:
  - Age
  - Gender
  - Education…
- At best we explain half of the wage variance (vanHeuvelen 2018a, 2018b).
- Residual wage variance, differs between groups: higher among men, higher educated, older groups (heteroscedasticity).

Sociology as a population science

- (Goldthorpe 2016, Xie 2007) distinguish **Typological Thinking** vs sociology as a **Population Science** view:
  - **Typological Thinking:** the regression coefficient is considered as the **true value** of a mechanism and the residual as a **measurement error.**
  - **Population Science**: the residual is not an error but **part of the phenomenon of interest.** No reason to focus more on the mean than on the variance.

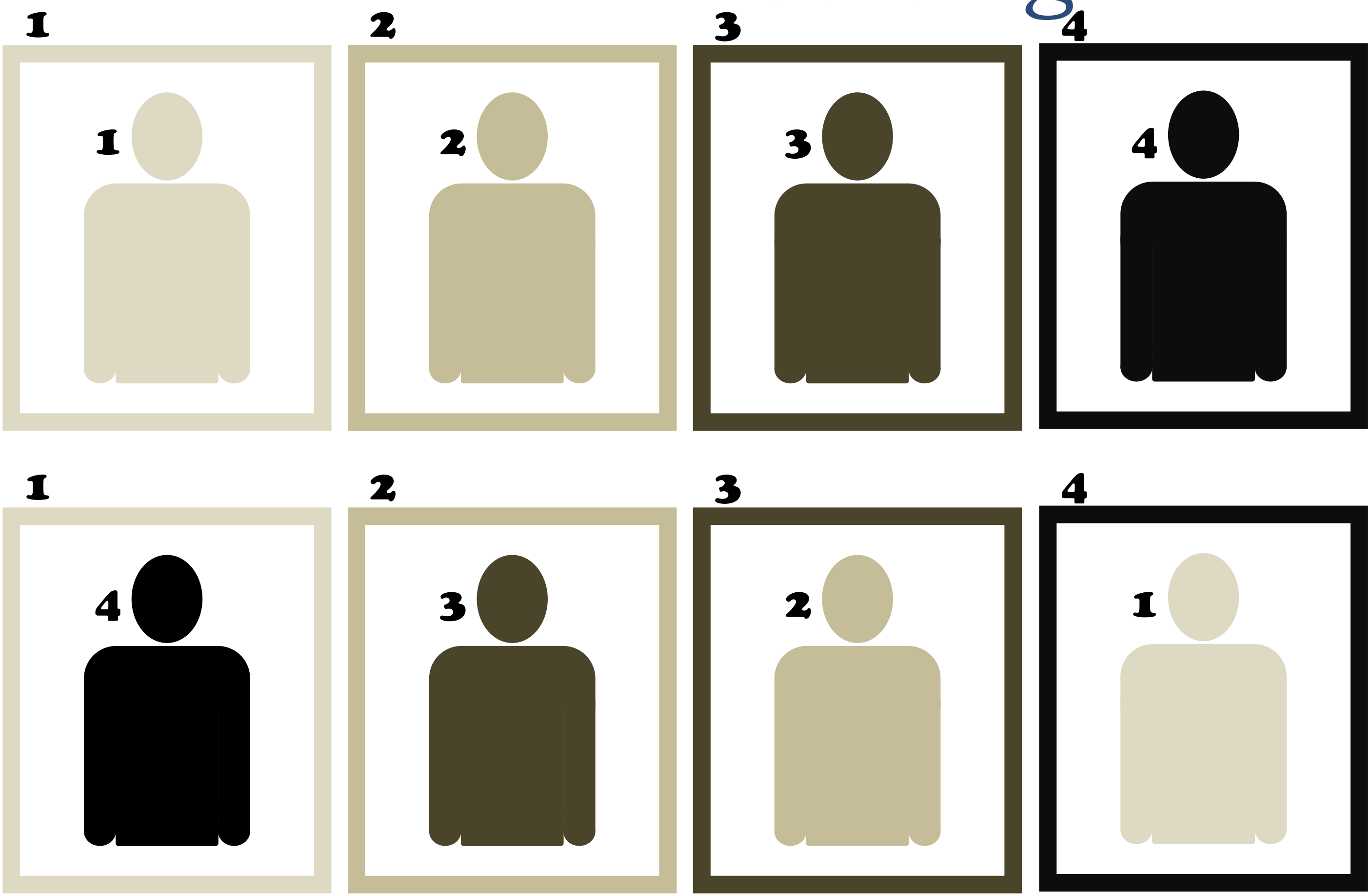
From Between- to Within-Group Wage Differences

- The literature has mainly been interested in temporal variations of BGWI and WGW
  - Institutional context (VanHeuvelen 2018a, 2018b)
  - Macroeconomic factors (Juhn et al. 1993)
  - Demographic changes (Lemieux 2006)
- They are generally analyzed as two separate phenomena
- The reasons **why similar workers are paid differently**, and **why this varies across demographic groups** in the first place are still unclear…

Theoretical explanation

1. Highly able worker (workers that **would get paid more** no matter **where** they work and **what** they do)…
2. …are overrepresented in highly paying organizations, and occupations (**sorting**).
3. Ability is correlated with observable characteristics like education…
4. …then, **sorting** will **exacerbate** both **between** and **within** group wage differences.

Sorting



- Var=0
- Var=6.67

Methods

- 3 dimensional fixed effect model:

$$y_{it} = \alpha_i + org_{J(it)} + occ_{K(it)} + X_{it}\beta + \varepsilon_{it}$$

Log hourly wage

Individual FE

Organization FE

Occupation FE

Time varying controls (exp, exp^2, year FE)

Residual

Methods

$$\overbrace{Var(y_{it})}^{Wage} = \overbrace{Var(\alpha_i)}^{Worker} + \overbrace{Var(org_{J(it)})}^{Organization} + \overbrace{Var(occ_{K(it)})}^{Occupation}$$

$$+ \underbrace{2Cov(\alpha_i, org_{J(it)})}_{Organizational\ sorting} + \underbrace{2Cov(\alpha_i, occ_{J(it)})}_{Occupational\ sorting}$$

$$+ \underbrace{2Cov(occ_{K(it)}, org_{J(it)})}_{Consolidation} + \underbrace{Var(\varepsilon_{it})}_{Volatility}$$

Methods

- To reduce the number of organizational FE (Bonhomme et al. 2019):

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

  - $\hat{F}_j$ the empirical cumulative distribution function of log earnings in firm j
  - n_j the number of employees in firm j,
  - µ a discrete or continuous measure,
  - k(1),…,k(J) a partition of firms into K classes,
  - H1,…,HK are cumulative distribution functions

Data

- Déclaration annuelle des données sociales (DADS):
  - DADS-postes (2010)
    - All jobs within organizations (SIREN id) in the economy, between 18-60 y.o. (≈**25 million jobs**)
    - Information about **gross hourly wage**, industry, age, sex
  - DADS-panel (2010-2014)
    - 1/12 of the French workers (born in october) (≈**2 million individuals**)
    - Information about hourly hourly wage, occupation (2 digits PCS), organizations (SIREN), experience, age, sex
- Echantillon démographique permanent (EDP)
  - Born on the first 4 days of october (≈**250,000 individuals**)
  - Information about education (*no schooling; highschool; highschool + 2; highschool +3 or more*)

Descriptive results

Firm Cluster	Number of individual observations	Number of firms	Average firm size	Average log hourly wage	Variance log hourly wage	Skewness	Kurtosis
1	402,229	350978	7.11	1.46	0.10	1.37	5.38
2	358,688	190400	40.63	2.20	0.02	0.49	43.20
3	677,947	145541	131.17	2.24	0.06	-1.25	14.58
4	452,941	44674	420.93	2.26	0.23	0.06	3.36
5	312,570	136421	19.56	2.33	0.01	-1.28	68.96
6	1,053,898	71071	1638.56	2.33	0.07	0.65	16.17
7	1,150,597	64494	1391.90	2.41	0.07	1.36	19.03
8	292,136	98167	42.91	2.41	0.01	0.68	55.92
9	807,106	54263	328.69	2.48	0.14	0.43	6.79
10	1,361,963	52972	5360.07	2.48	0.07	1.47	13.46
11	166,803	80562	40.66	2.48	0.01	1.81	76.82
12	1,501,688	37128	56838.48	2.54	0.05	1.89	15.46
13	1,416,862	45051	1320.71	2.57	0.11	1.03	8.57
14	274,774	78221	70.54	2.58	0.03	1.89	31.31
15	546,416	46095	726.51	2.62	0.22	0.14	4.36
16	1,898,634	41821	5696.63	2.63	0.10	1.32	11.30
17	503,175	73357	654.50	2.69	0.06	1.44	16.09
18	1,856,253	30597	2227.96	2.70	0.12	1.12	7.18
19	1,511,305	25821	8353.55	2.78	0.13	0.89	7.63
20	1,012,046	29110	54708.29	2.79	0.19	0.93	7.79
21	591,384	19637	162.96	2.82	0.18	-0.21	4.58
22	745,063	54096	3358.20	2.82	0.08	1.13	10.28
23	1,443,633	20989	32844.12	2.88	0.16	0.65	6.37
24	1,210,118	34853	13558.19	2.94	0.13	0.72	9.25
25	805,121	22283	4424.57	3.05	0.28	0.22	5.43
26	818,033	37794	4462.44	3.07	0.14	0.47	9.02
27	1,140,572	30091	6908.48	3.24	0.25	0.08	7.02
28	542,583	44430	1709.36	3.53	0.34	0.09	7.41
Total	24,854,538	1,960,917					
Average			3485.92	2.67	0.13		

Table 1: descriptive statistics firm clustering algorithm BLM. Data: DADS-postes 2010.

Results regression

Covariances y1 and:	y1	Percentage explained
y1	0.22	100
Individual FE	0.08	34.74
Organization FE	0.02	6.95
Occupation FE	0.03	14.10
Labor market experience	0.07	33.44
Year FE	0.00	-0.10
Residual	0.02	10.88

Table 5: Covariance between log hourly wage and components of the wage equation 3. Because all terms are positive, except for the covariance with year FE which is negligible, we can roughly interpret cov(a,y)/var(y) as an indication of the contribution of a to the wage variance.

Results regression

Theoretical explanation

- Highly able worker (workers that **would get paid more** no matter **where** they work and **what** they do)…
2. Are overrepresented in highly paying organizations, and occupations (**sorting**).
3. Ability is correlated with observable characteristics like education…
4. Then, **sorting** will **exacerbate** both **between** and **within** group wage differences.

Covariances y1 and:	**y1**	**Percentage explained**
y1	0.22	100
Individual FE	0.08	34.74
Organization FE	0.02	6.95
Occupation FE	0.03	14.10
Labor market experience	0.07	33.44
Year FE	0.00	-0.10
Residual	0.02	10.88

Table 5: Covariance between log hourly wage and components of the wage equation 3. Because all terms are positive, except for the covariance with year FE which is negligible, we can roughly interpret cov(a,y)/var(y) as an indication of the contribution of a to the wage variance.

Results

Theoretical explanation

- Highly able worker (workers that **would get paid more** no matter **where** they work and **what** they do)…
- Are overrepresented in highly paying organizations, and occupations (**sorting**).
- Ability is correlated with observable characteristics like education…

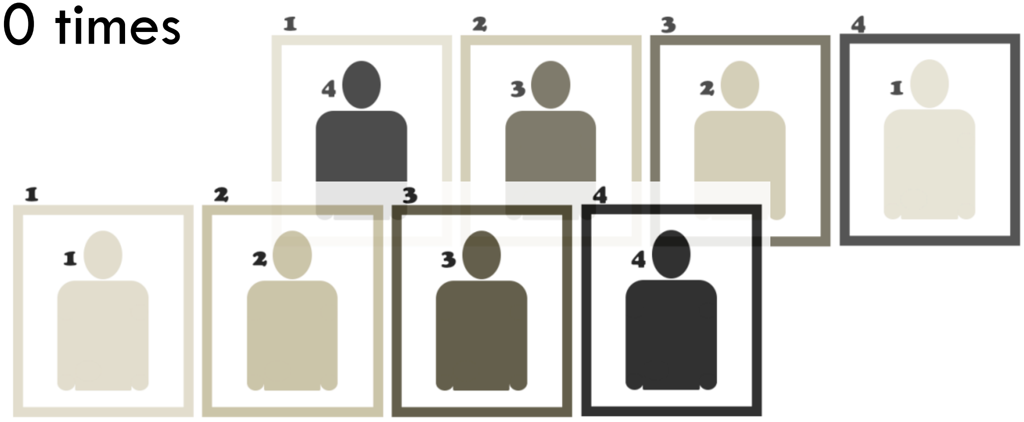
4. Then, **sorting** will **exacerbate** both **between** and **within** group wage differences.

$$Cor\left(\alpha_i, org_{J(it)}\right) > 0$$
$$Cor\left(\alpha_i, occ_{K(it)}\right) > 0$$

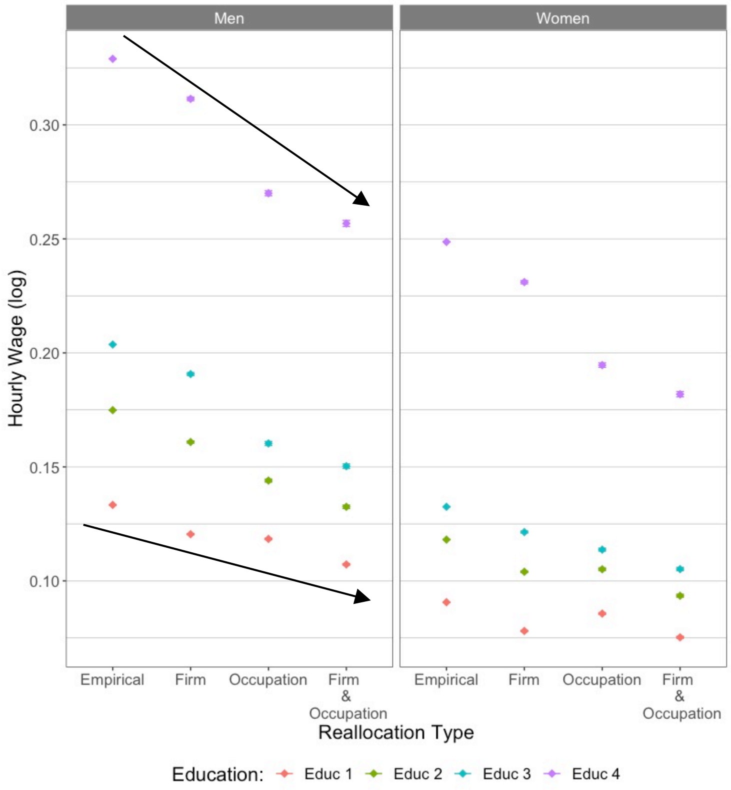
Counterfactual reallocations

$$y_{it} = \widehat{\alpha}_i + \widehat{org_{J(it)}} + \widehat{occ_{K(it)}} + \hat{\beta}X_{it} + \widehat{\varepsilon_{it}}$$

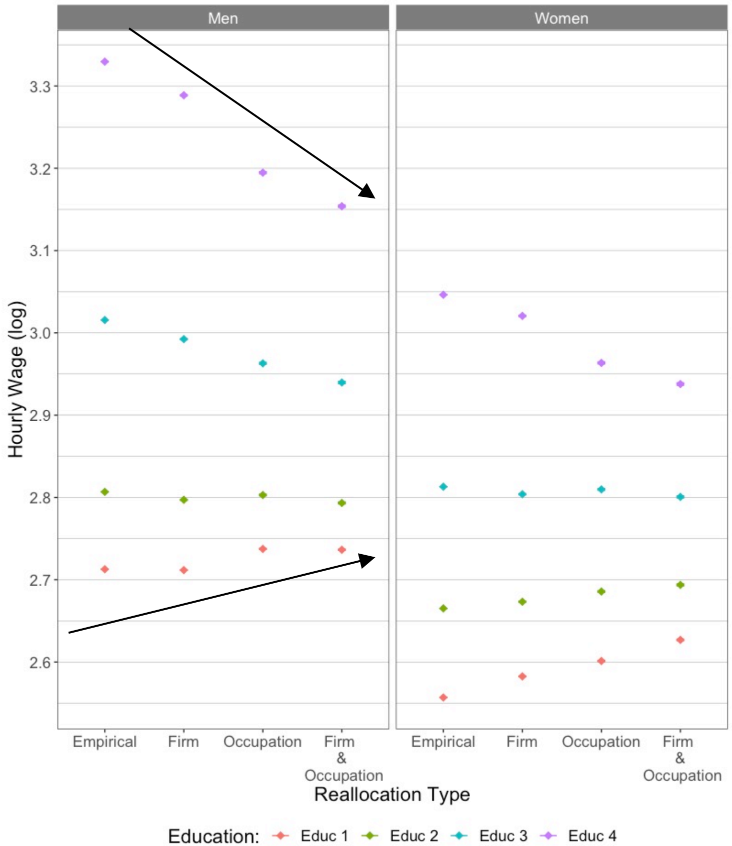
- **Randomly reallocate** workers into **occupations** and/or **organizations** holding constant their marginal distributions
- Compute a **counterfactual** wage for each individual
- Repeat 2500 times



Counterfactual: random reallocation: WGWD



Counterfactual: random reallocation: BGWI



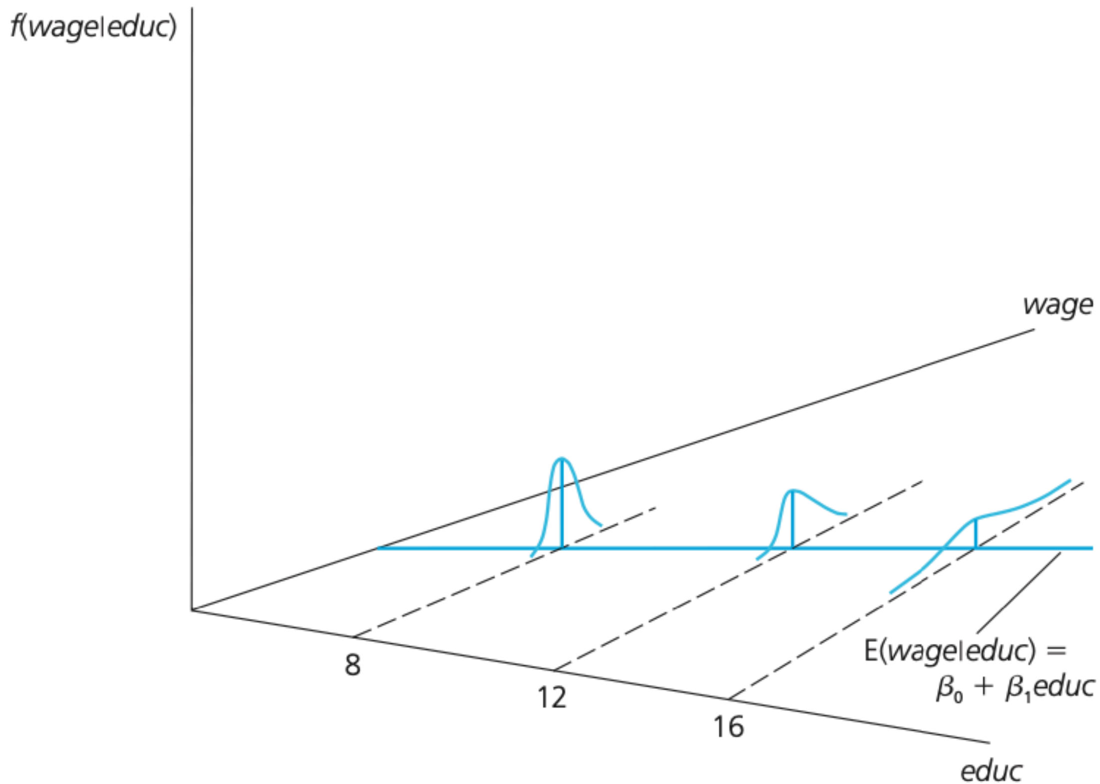
- Reduction by 32% for men, 37% for women

Summary

- **Occupational** and **organizational sorting** partly explain:
  - why similar workers are paid differently
  - why this differs across sociodemographic groups
  - why there is inequality between social groups
- They act as a **multiplier** of **pre-existing inter-individual differences** (Matthew effect).
- Nuance: it is because the correlation between ability and occupational and organizational FE is **positive but not perfect** that it contributes to both between and within group wage differences.

Thank you for you attention!

From Between- to Within-Group Wage Differences



Sources of wage differences

- Processes of **sorting**: What are the interplays between the three aspects?
- Mechanisms possibly generating sorting:
  – Recruitment procedure of organizations
  – Social networks
  – Influence of co-workers on individual ability

Methods

- Computationally difficult (too many parameters)
- Organizational and occupational FE are estimated using the movement of some workers across occupations and organizations
- Graph were nodes are organizations or occupations, and they are connected by the movement of workers (in both directions)

Methods

- Limited mobility bias: too few moves to estimate all organizations and occupations FE
- Biased in favor of organizations (more dummies)

Descriptive results

Occupation	Number of individual-year observations	Proportion of Men	Average labor market experience (years)	Average Age	Standard deviation Age	Proportion Education 1	Proportion Education 2	Proportion Education 3	Proportion Education 4
Ouvriés non qualif artisanal	368,707	0.60	14.87	37.10	13.64	0.79	0.14	0.04	0.03
Perso services directs aux part	599,043	0.32	12.29	34.98	13.31	0.67	0.21	0.07	0.04
Ouvriers agricoles et assimilés	13,683	0.75	13.03	35.81	13.05	0.69	0.21	0.07	0.04
Employés de commerce	638,457	0.33	11.38	33.73	12.03	0.54	0.29	0.12	0.05
Ouvriers non qualifies indus	475,014	0.67	13.87	35.95	12.28	0.75	0.17	0.05	0.03
Employés civils	261,238	0.22	15.88	39.44	12.29	0.59	0.22	0.16	0.03
Policiers, militaires	79,600	0.85	16.57	39.92	11.56	0.67	0.21	0.07	0.05
Chauffeurs	305,279	0.90	19.51	42.82	11.67	0.80	0.14	0.04	0.02
Clergé, religieux	649	0.65	24.49	50.40	11.89	0.30	0.14	0.22	0.34
Ouvriers qualif indus	554,807	0.94	16.02	37.85	11.96	0.75	0.19	0.04	0.02
Employés admin	901,155	0.22	14.57	37.96	11.97	0.28	0.31	0.27	0.13
Ouvriers qualif manutention	183,665	0.88	16.81	39.13	11.11	0.71	0.21	0.06	0.02
Professeurs écoles	78,638	0.53	15.43	39.88	12.77	0.22	0.27	0.23	0.28
Ouvriés qualif indus	529,369	0.84	18.40	40.43	11.32	0.71	0.21	0.06	0.02
Prof interm santé & social	257,647	0.25	14.55	38.27	12.20	0.16	0.23	0.49	0.13
Prof. inter admin fct pu	31,696	0.42	22.46	45.99	10.62	0.42	0.36	0.15	0.07
Techniciens	327,736	0.83	16.44	39.14	11.20	0.28	0.25	0.34	0.12
Prof inter admin et com d'entr	492,749	0.43	17.04	40.44	11.03	0.27	0.26	0.30	0.17
Artisans	4,873	0.84	24.27	48.26	9.73	0.62	0.20	0.11	0.06
Commerçants et assimilés	13,588	0.74	23.04	47.95	10.65	0.46	0.22	0.17	0.15
Contremaîtres & maîtrise	140,460	0.90	20.92	43.37	10.25	0.55	0.22	0.17	0.05
Cadres function publique	20,377	0.59	23.30	47.21	9.92	0.20	0.24	0.23	0.33
Prof. libérales	12,548	0.25	14.31	39.52	11.69	0.03	0.05	0.07	0.84
Prof. de l'info, arts, spectacles	63,531	0.59	16.40	41.50	11.67	0.20	0.22	0.23	0.35
Ingé & cadres tech d'entr	414,373	0.80	16.65	40.44	10.34	0.10	0.10	0.20	0.59
Professeurs, prof. scienti	49,509	0.45	18.89	44.69	12.91	0.03	0.04	0.09	0.84
Cadres admin et comm d'entr	530,813	0.56	18.91	42.98	10.36	0.15	0.16	0.23	0.46
Chefs d'entr	44,687	0.81	24.55	49.33	9.90	0.28	0.19	0.21	0.32
Average	264,068	0.56	15.78	38.73	12.17	0.48	0.22	0.17	0.14

Table 3: Descriptive statistics for occupations (pcs 2 digits). Occupations are ranked in ascending order according to their average wage. Source: DADS-Panel matched with EDP.

Sorting on observables

	Occupations	Firm groups
Education	0.1637	0.0573
Gender	0.4736	0.1248

Table 4: Entropy index for education and dissimilarity index for gender across occupations and firm clusters.

Sorting on observables

Figure 2: Educational sorting across firm groups. Source DADS-panel matched with EDP.

Counterfactuals

1. No organizational effect:
$$y_{it}^{CF1} = y_{it}^{F} - \hat{k}_{ij}$$
2. No occupational effect:
$$y_{it}^{CF2} = y_{it}^{F} - \widehat{occ}_{i}$$
3. None of the two:
$$y_{it}^{CF3} = y_{it}^{F} - \hat{k}_{ij} - \widehat{occ}_{i}$$

Counterfactuals

Log wage distribution	Log wage variance	Reduction
Empirical	0.224	
No organizations	0.197	11.97%
No occupations	0.171	23.89%
No organizations and occupations	0.148	33.76%

Table 6: Log wage variance of counterfactual log wage distributions where the roles of organizations and/or occupations are netted out.

Counterfactual: random reallocation

- Assessing the specific contributions of both forms of sorting on BGWI, WGWD and overall wage variance:

$$y_{it} = \widehat{\alpha}_i + \widehat{k_{ij}} + o\widehat{cc}_i + X_{it}\hat{\beta} + \widehat{\varepsilon_{it}}$$

Counterfactual: random reallocation

Distribution	Wage variance	99% bounds of estimated values	Reduction
Empirical	0.2242		
Random organization reallocation	0.2017	[0.2016; 0.2018]	10.04%
Random occupation reallocation	0.1803	[0.1802; 0.1805]	19.56%
Random firm and occupation reallocations	0.1625	[0.1623; 0.1626]	27.52%

Table 7: Random reallocation exercises repeated 2500 times.

Counterfactual: random reallocation

Distribution	**Wage variance**	**99% bounds of estimated values**	**Reduction**
Empirical	0.2242		
Random organization reallocation	0.2017	[0.2016; 0.2018]	10.04%
Random occupation reallocation	0.1803	[0.1802; 0.1805]	19.56%
Random firm and occupation reallocations	0.1625	[0.1623; 0.1626]	27.52%

Table 7: Random reallocation exercises repeated 2500 times.

Counterfactual: random reallocation: gender

Distribution:	Empirical	Organization	Occupation	Firm & Occupation
Average gap	0.1779	0.1568	0.1479	0.1268
Reduction		11.86%	16.86%	28.72%
Variance difference	0.0556	0.0552	0.0422	0.0415
Reduction		1%	24%	25%

Table 8: effect of random reallocation exercises on the gender wage gap and on the within group wage variance (WGWD).

Thank you!

UNIVERSITY OF TRENTO

Doctoral School of Social Sciences

SORBONNE UNIVERSITÉ

Gemass

F

- T

Sources of wage differences

- Processes of sorting
  - What are the interplays between the three aspects?