Worker Sorting and the Gender Wage Gap

Giulia Vattuone
SOFI, Stockholm University

EEA-ESEM 2024 Rotterdam

August 26-30, 2024

Introduction

- Women more likely to work in firms that pay lower wages (Card et al., 2016)
  - → Gender differences in sorting across firms account for $\sim 20\%$ of the gender wage gap
    - ◇ This result is confirmed across multiple countries
    - ◇ Not due to lack of skills or experience
- Major debate: differences in employment opportunities or in preferences?
  - → Depending on the answer, distinct types of policies to enhance worker-firm matching

This paper

Research Question:

- Employment opportunities vs preferences

  → Quantify their relative importance in driving the sorting component of the gender wage gap

Data and Method:

- Matched employer-employee monthly data from **Ile-de-France** over 2015-2019
- **Revealed preference approach** in a random search framework
- Exploit information on **firm-to-firm transitions**
- Estimate model of wages and mobility in the spirit of Lentz, Piyapromdee, Robin (2023)
  - → Worker heterogeneity within *and* between genders to generate rich sorting patterns
  - → Quantify relative importance of **mobility into and out of employment** (left unrestricted)

Preview of Findings

- Conditional gender wage gap: 11 log points
- Gender differences in sorting across firms account for 20% of the gender wage gap
  - ◇ Differences in worker preferences account for over half of the sorting component
    - → More salient among high-wage, mid-experienced workers
  - ◇ Differences in job offer distribution after non-employment explain the other half
    - → Across all worker types and career stages
  - ◇ Differences in offer arrival rates do not contribute to the gender wage gap
    - → If anything, women more likely to receive offers for higher-paying positions
- Differences in firm sorting become relatively less important with experience:
  - ◇ 25% among junior workers, 16% among senior workers

Contribution

- *Literature that* ***quantifies the sorting component*** *of the gender wage gap*
  - ◇ (Card et al., 2016; Cardoso et al., 2016; Casarico and Lattanzio, 2024; Palladino et al., 2021)

$\Longrightarrow$ I gauge the relative importance of key mobility components driving it

- *Literature that* ***explains the sorting component*** *through a structural approach*
  - ◇ (Sorkin, 2017; Sorkin, 2018; Morchio and Moser, 2023)
  - ◇ pioneered revealed preference approach to extract info from firm-to-firm transitions

$\Longrightarrow$ I allow for **rich sources of worker heterogeneity**

- ◇ Female and male wages and mobility vary differently over their careers within a type of worker

$\rightarrow$ Differences in sorting across **different market segments** and at **different career stages**

Contribution

- *Literature that points out that gender wage gaps may materialise as a result of:*
  - Differences in job search behaviour (Braun and Figueiredo, 2022)
  - Employer discrimination in hiring (Neumark et al., 1996; Xiao, 2023; Kline et al., 2022)
  - Preferences for shorter commute (Le Barbanchon et al., 2021; Fluchtmann et al., 2024)
  - Preferences for flexibility (Mas and Pallais, 2017; Wiswall and Zafar, 2018)
  - Differences in risk preferences in job-finding behaviour (Cortés et al., 2023)

$\Longrightarrow$ I separate gender differences in offer distributions and preferences for a general economy

$\Longrightarrow$ I capture an overall bundle of characteristics valued by workers

- do not focus on a specific preference mechanism

Outline

1) Revealed preference approach and Statistical model

2) Data

3) Clustering results

4) Worker Sorting and the Gender Wage Gap

5) Conclusions

Outline

Revealed preference approach

- Data on observed firm-to-firm transitions are informative about:
  - Offer arrival rates (representing employment opportunities)
  - Acceptance rates (revealing worker preferences)
- A worker accepts an offer if poacher provides higher utility than incumbent
  - → Workers may value something beyond wages in a way that guides where they sort
- Denote j current firm and j' subsequent firm, with $j \neq j'$. Firm-to-Firm modelled as:

$$\Pr_i\left(j' \mid j\right) = \underbrace{\Pr_i\left(\text{Offer from } j'\right)}_{\text{employment opportunity}} \times \underbrace{\Pr_i\left(j' \succ j \mid \text{Offer from } j'\right)}_{\text{choice probability}}$$

- Identification challenge (**!**): cannot work at the i-j level

Identifying assumptions

1. Worker and firm unobserved heterogeneity is discretised
   - Workers are associated to a finite number of *types*, $l_i \in \{1, \ldots, L\}$
   - Firms are associated to a finite number of *classes*, $k_j \in \{1, \ldots, K\}$ and $k_0 = 0$

2. Workers of a given type have common preferences over firms of a given class
   - Consider workers of type l, with characteristics x, working in firms of class k. Their utility:

$$U_{ij} = \gamma_{lxk} + \epsilon_{ij}$$

   - ϵ_{ij} is an idiosyncratic utility draw, specific to worker i and firm j
     - $\rightarrow$ e.g.: choice to move is influenced by moving costs
   - When choosing between two firms j and j', worker i compares U_{ij} and $U_{ij'}$

The likelihood of (l, x)-type worker

Conditional on a classification C of firms into classes, on the initial characteristics x_{i1}, and on a value θ of the parameters, the complete likelihood of worker i's history is:

$$\mathcal{L}_i(\theta | l_i, x_{i1}, C) = \Pr\Big(l_i, k_{j(i,1)} \mid x_{i1}\Big) \times \prod_{t=1}^{T-1} \Bigg\{ \Pr\Big(k_{j(i,t+1)} \mid k_{j(i,t)}, l_i, x_{it}\Big)^{\mathbb{1}\{s_{it}=1\}} \times \Pr\Big(\text{no move} \mid k_{j(i,t)}, l_i, x_{it}\Big)^{\mathbb{1}\{s_{it}=0\}} \Bigg\} \times \prod_{t=1}^{T} f\Big(y_{it} \mid l_i, x_{it}, k_{j(i,t)}\Big)$$

Modelling mobility and identification

Estimation

- Workers and firms latent types are unobserved
    - **First step**: classify firms into classes using K-Means Algorithm
        - → firms in same class are similar in gender-specific wage distributions, female share, size
        - → number of classes $K = 15$
    - **Second step**: conditional on firm classes, EM algorithm to estimate parameters and classify workers
        - → Likelihood function is non-linear in the parameters
        - → EM + MM Algorithm (Lentz, Piyapromdee, and Robin, 2023)
        - → number of latent types $L = 3$
        - → Latent types interact with combinations of gender, experience, and tenure
        - → By interacting gender with time-varying characteristics, men's and women's wage profiles and mobility patterns can vary over time within a latent type

Data used to discretise Model fit

Summary of estimated parameters

- $m_0\left(l, k \mid x\right)$: initial worker-firm matching Go
- $\lambda_{lxk'}$: (l, x)-type worker's probability of receiving offer from k' Go
- γ_{lxk}: (l, x)-type worker's perceived quality of firm type k Go
- δ_{lxk}: (l, x)-type's probability of exiting k and going into non-employment Go
- $\psi_{lxk'}$: (l, x)-type's probability of moving to k' from non-employment Go
- μ_{lxk} and σ_{lxk}: (l, x, k)-specific wage distributions Go

Outline

1) Revealed preference approach and Statistical model

2) Data

3) Clustering results

4) Worker Sorting and the Gender Wage Gap

5) Conclusions

French Matched Employer-Employee Data

Data sources: *Déclarations Annuelles de Données Sociales* (DADS) 2015-2019

1. DADS-Postes
   - Universe of jobs in given year in France
     - $\rightarrow$ used to cluster firms
2. DADS-Panel
   - Panel of employed workers born in October
     - $\rightarrow$ used to model wages and mobility, and to cluster workers

- Sample selection
  - Workers in Île de France, employed in 01/2015, traced monthly until 12/19
  - N women $= 80,967$, N men $= 84,191$
  - Firms hiring both genders and active for five years (N firms $= 25,925$)

Outline

1) Revealed preference approach and Statistical model

2) Data

3) Clustering results

4) Worker Sorting and the Gender Wage Gap

5) Conclusions

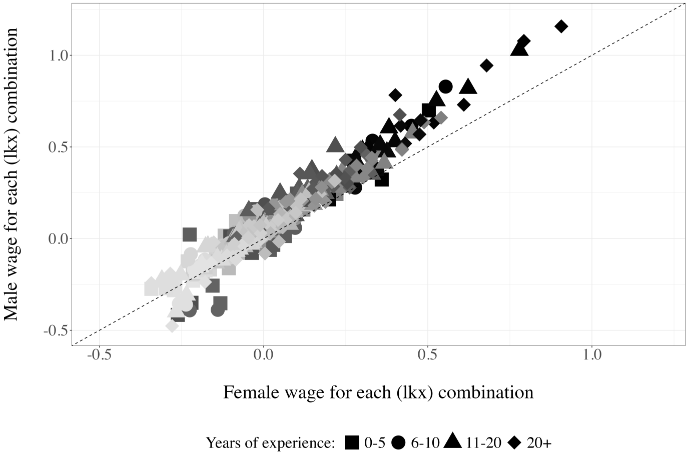
Latent types capture diverse career trajectories

Tenure:	Short						Long					
Worker type:	Low-wage		Mid-wage		High-wage		Low-wage		Mid-wage		High-wage	
Gender:	F	M	F	M	F	M	F	M	F	M	F	M
Unconditional log wage												
Experience 0-5	2.78	2.87	2.83	2.95	2.88	3.03	2.84	2.94	2.91	3.06	2.97	3.16
Experience 6-10	2.79	2.87	2.91	3.04	2.98	3.18	2.85	2.96	2.95	3.12	3.09	3.33
Experience 11-20	2.78	2.86	2.99	3.15	3.18	3.44	2.88	2.97	3.05	3.21	3.29	3.58
Experience 20+	2.74	2.86	3.04	3.19	3.30	3.54	2.88	2.98	3.11	3.25	3.43	3.71

- Stagnant wages irrespective of experience levels for low-wage worker types.
- Wages increase with experience for mid-wage and high-wage worker types.

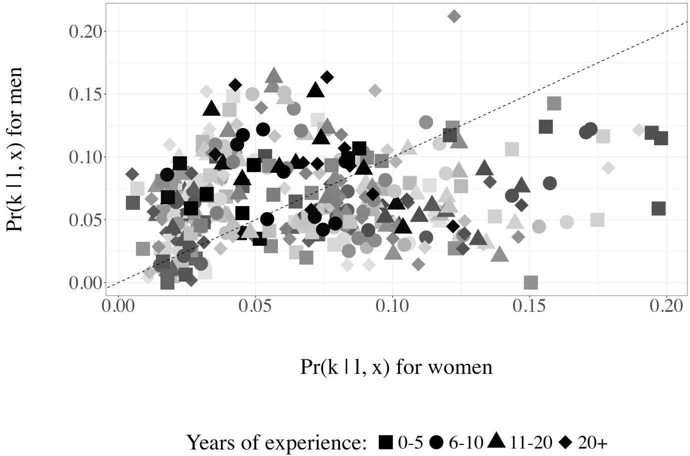
Ex-post tabulations on hours and mobility | Heterogeneity in worker preferences | Firm classification

Gender differences in wages



- **Lighter colors**: low-wage workers in low-paying firms.
- **Darker colors**: high-wage workers in high-paying firms.
- Strong correlation between female and male wages.
- Gender wage gaps are higher with experience.

Gender differences in worker-firm allocations



- Men and women are unequally distributed across firms.
- High-experience, high-wage men are more likely than their female counterparts to work at high-paying firms.
- No strong wage sorting for either gender. Stronger tendency of wage sorting for men.

Outline

1) Revealed preference approach and Statistical model

2) Data

3) Clustering results

4) Worker Sorting and the Gender Wage Gap

5) Conclusions

Worker Sorting

- Sorting is the stationary allocation of worker types and firm classes

$$\Pr^*\Big(k \mid l, x\Big)$$

- It is computed using the transition probabilities Go

Decomposing the gender wage gap

- Obtain distribution of matches using the sorting distribution and worker type frequencies

$$\Pr(l, x, k) = \Pr^*(k \mid l, x) \Pr(l, x)$$

More details

- Simulate cross-sectional dataset drawing wages using $\hat{\mu}_{lxk}$ and $\hat{\sigma}^2_{lxk}$
  - ◇ augment initial sample size by L
- Obtain counterfactual datasets under different scenarios:
  - ◇ if men and women had same offer arrival rates while employed
  - ◇ if they also had same preferences over firm classes
  - ◇ if they also had same exit rates while employed
  - ◇ if they also had same offer arrival rates while non-employed
- Compute the gender wage gap under different counterfactual scenarios

Worker Sorting and the Gender Wage Gap

Full sample

Cross-gender correlations of estimated parameters

Breakdown by experience groups

Outline

Conclusions

What the paper does:

- Administrative data + flexible model to decompose the sorting component of the gender wage gap first presented by Card, Cardoso, and Kline (2016).
    - ◇ Leveraging the finite mixture approach of Lentz, Piyapromdee, and Robin (2023)
    - ◇ Worker clustering instrumental in allowing for both within and between gender variation.

Findings:

- If women were distributed across firms as men are, the gender wage gap would ↓ by 20%.
- Differences in preferences explain over half of this sorting component of the wage gap.
    - ◇ Mainly driven by **high-wage workers** in child-rearing ages.
    - → Difficult to distinguish b/w preferences, norms, constraints (Akerlof and Kranton, 2000)
    - → These findings recall proposals entailing restructuring jobs so that a broader range of possibly constrained workers can reach them (Goldin and Katz, 2016; Wasserman, 2022)

References I

Akerlof, George A. and Rachel E. Kranton (2000). "Economics and Identity". *The Quarterly Journal of Economics* 115.3, pp. 715–753.

Braun, Christine and Ana Figueiredo (2022). "Labor Market Beliefs and the Gender Wage Gap". Working Paper.

Card, David, Ana Rute Cardoso, and Patrick Kline (May 2016). " Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women *". *The Quarterly Journal of Economics* 131.2, pp. 633–686.

Cardoso, Ana Rute, Paulo Guimarães, and Pedro Portugal (Jan. 2016). "What drives the gender wage gap? A look at the role of firm and job-title heterogeneity". *Oxford Economic Papers* 68.2, pp. 506–524.

Casarico, Alessandra and Salvatore Lattanzio (2024). "What Firms Do: Gender Inequality in Linked Employer-Employee Data". *Journal of Labor Economics* 42.2, pp. 325–355.

References II

Cortés, Patricia, Jessica Pan, Laura Pilossoph, Ernesto Reuben, and Basit Zafar (May 2023). "Gender Differences in Job Search and the Earnings Gap: Evidence from the Field and Lab*". *The Quarterly Journal of Economics* 138.4, pp. 2069–2126.

Fluchtmann, Jonas, Anita M. Glenny, Nikolaj A. Harmon, and Jonas Maibom (May 2024). "The Gender Application Gap: Do Men and Women Apply for the Same Jobs?" *American Economic Journal: Economic Policy* 16.2, pp. 182–219.

Goldin, Claudia and Lawrence F. Katz (2016). "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation". *Journal of Labor Economics* 34.3, pp. 705–746.

Kline, Patrick, Evan K Rose, and Christopher R Walters (June 2022). "Systemic Discrimination Among Large U.S. Employers*". *The Quarterly Journal of Economics* 137.4, pp. 1963–2036.

References III

Le Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet (Feb. 2021). "Gender Differences in Job Search: Trading off Commute against Wage*". *The Quarterly Journal of Economics* 136.1, pp. 381–426.

Mas, Alexandre and Amanda Pallais (Dec. 2017). "Valuing Alternative Work Arrangements". *American Economic Review* 107.12, pp. 3722–59.

Morchio, Iacopo and Christian Moser (Sept. 2023). *The Gender Pay Gap: Micro Sources and Macro Consequences*. Institute working paper 78. Federal Reserve Bank of Minneapolis. Opportunity and Inclusive Growth Institute.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort (1996). "Sex Discrimination in Restaurant Hiring: An Audit Study". *The Quarterly Journal of Economics* 111.3, pp. 915–941.

References IV

Palladino, Marco Guido, Alexandra Roulet, and Mark Stabile (Oct. 2021). *Understanding the role of firms in the gender wage gap over time, over the life cycle, and across worker types*. Discussion Paper 16671. CEPR.

Sorkin, Isaac (May 2017). "The Role of Firms in Gender Earnings Inequality: Evidence from the United States". *American Economic Review: Papers and Proceedings* 107.5, pp. 384–87.

— (Jan. 2018). "Ranking Firms Using Revealed Preference*". *The Quarterly Journal of Economics* 133.3, pp. 1331–1393.

Wasserman, Melanie (July 2022). "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents". *The Review of Economic Studies* 90.3, pp. 1535–1568.

Wiswall, Matthew and Basit Zafar (Aug. 2018). "Preference for the Workplace, Investment in Human Capital, and Gender*". *The Quarterly Journal of Economics* 133.1, pp. 457–507.

Xiao, Pengpeng (2023). *Equilibrium Sorting and the Gender Wage Gap*. Working Paper 144. VATT Institute for Economic Research.

Tables and Figures

Sample Description

BACK

Gender:	Women	Men
Mean Annual Earnings	40,874	53,178
Mean Hours	1,693	1,784
Share Part-time	16%	5%
Mean Age	41	41
Share doing JTJ	15%	15%
Share doing E-NE	19%	16%
Share doing NE-E	45%	38%
Mean job spell months	43	44
N workers	80,967	84,191

Additional breakdowns

Sample Description

Back

Women

	Women							
Tenure:	Short				Long			
Years of experience:	0-5	6-10	11-20	20+	0-5	6-10	11-20	20+
Mean Annual Earnings	26,692	30,161	34,122	37,358	32,900	34,721	40,174	44,925
Mean Hours	1,438	1,494	1,503	1,520	1,622	1,659	1,692	1,732
Share Part-time	14%	12%	15%	18%	12%	12%	18%	16%
Mean Age	29	31	38	48	29	31	38	48
Share doing JTJ	11%	9%	8%	8%	10%	13%	10%	7%
Share doing E-NE	14%	13%	12%	11%	12%	17%	12%	8%
Share doing NE-E	36%	32%	27%	24%	32%	31%	30%	20%
Mean job spell months	10	11	12	12	12	23	32	40
N workers	4,774	10,181	14,257	7,609	4,754	20,362	47,476	39,714

Sample Description

Back

Men

	Men							
Tenure:	Short				Long			
Years of experience:	0-5	6-10	11-20	20+	0-5	6-10	11-20	20+
Mean Annual Earnings	33,229	36,829	44,002	48,976	39,212	43,273	52,001	58,543
Mean Hours	1,510	1,596	1,628	1,648	1,695	1,747	1,790	1,810
Share Part-time	9%	6%	5%	6%	6%	5%	4%	4%
Mean Age	29	31	38	48	29	31	38	48
Share doing JTJ	9%	8%	8%	7%	11%	14%	10%	7%
Share doing E-NE	11%	10%	11%	10%	11%	13%	10%	8%
Share doing NE-E	33%	27%	25%	24%	33%	22%	20%	19%
Mean job spell months	11	11	12	12	13	23	32	41
N workers	4,135	8,872	13,711	9,011	4,310	18,640	48,211	44,958

Sample Description

Back

Occupations

Gender:	Women	Men
Share Managers	34%	42%
Share Intermediate	28%	20%
Share Employee non-manual	31%	17%
Share Employee manual	6%	21%
N workers	80,967	84,191

Firm classification

Back

Firm class:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N firms	570	1,973	1,763	1,576	1,001	1,992	1,121	1,360	2,298	1,017	2,772	1,480	1,939	2,799	2,264
Mean size	294	198	142	248	607	179	818	299	175	460	211	237	197	213	110
Female share	75%	72%	32%	28%	67%	24%	31%	80%	57%	55%	22%	69%	54%	27%	41%
Mean log hourly wage (EUR)	2.55	2.75	2.69	2.74	2.65	2.91	2.78	2.81	2.94	2.88	3.01	3.13	3.12	3.19	3.50
Mean hours	1,159	1,246	1,232	1,294	1,223	1,383	1,347	1,277	1,279	1,359	1,448	1,409	1,324	1,460	1,495
Share of women managers	11%	25%	16%	9%	6%	13%	6%	12%	29%	11%	10%	21%	28%	14%	21%
Share of men managers	7%	16%	29%	18%	4%	28%	13%	7%	26%	12%	25%	17%	28%	29%	32%
Share of women among managers	56%	62%	35%	30%	55%	30%	32%	64%	53%	47%	27%	54%	49%	31%	38%

Firm classification

Back

Firm class:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N firms	570	1,973	1,763	1,576	1,001	1,992	1,121	1,360	2,298	1,017	2,772	1,480	1,939	2,799	2,264
Mean size	294	198	142	248	607	179	818	299	175	460	211	237	197	213	110
Female share	75%	72%	32%	28%	67%	24%	31%	80%	57%	55%	22%	69%	54%	27%	41%
Share of firms in Hotel	9%	3%	2%	8%	17%	2%	14%	4%	1%	15%	4%	2%	1%	1%	1%
Share of firms in Admin Services	13%	6%	9%	15%	13%	6%	13%	6%	6%	6%	6%	6%	7%	4%	5%
Share of firms in Construction	0%	0%	3%	2%	0%	13%	4%	0%	1%	1%	17%	2%	2%	17%	9%
Share of firms in Commerce	21%	12%	12%	18%	15%	16%	19%	13%	9%	16%	21%	18%	13%	24%	28%
Share of firms in Education	8%	30%	3%	4%	4%	1%	2%	7%	22%	4%	1%	7%	11%	4%	2%
Share of firms in Managing	9%	11%	19%	9%	2%	12%	4%	5%	15%	6%	7%	15%	20%	10%	13%
Share of firms in Finance	2%	3%	2%	1%	0%	2%	1%	2%	6%	3%	2%	8%	12%	5%	11%
Share of firms in Pub Admin	2%	5%	0%	1%	28%	0%	3%	15%	2%	11%	1%	8%	1%	2%	2%
Share of firms in Health Accomm	10%	8%	2%	1%	10%	0%	2%	22%	2%	5%	0%	3%	1%	0%	1%

Latent types capture diverse career trajectories

Tenure:	Short						Long					
Worker type:	Low-wage		Mid-wage		High-wage		Low-wage		Mid-wage		High-wage	
Gender:	F	M	F	M	F	M	F	M	F	M	F	M
Unconditional log wage												
Experience 0-5	2.78	2.87	2.83	2.95	2.88	3.03	2.84	2.94	2.91	3.06	2.97	3.16
Experience 6-10	2.79	2.87	2.91	3.04	2.98	3.18	2.85	2.96	2.95	3.12	3.09	3.33
Experience 11-20	2.78	2.86	2.99	3.15	3.18	3.44	2.88	2.97	3.05	3.21	3.29	3.58
Experience 20+	2.74	2.86	3.04	3.19	3.30	3.54	2.88	2.98	3.11	3.25	3.43	3.71

- Stagnant wages irrespective of experience levels for low-wage worker types.
- Wages increase with experience for mid-wage and high-wage worker types.
- The gender wage gap widens with experience among high-wage types (from 15 to 28 log points).

Ex-post tabulations on hours and mobility

Worker types: Hours

Back

Tenure:	Short						Long					
Worker type:	Low-wage		Mid-wage		High-wage		Low-wage		Mid-wage		High-wage	
Gender:	F	M	F	M	F	M	F	M	F	M	F	M
Hours												
Experience 0-5	1,573	1,638	1,545	1,626	1,530	1,578	1,679	1,715	1,653	1,715	1,603	1,652
Experience 6-10	1,577	1,671	1,578	1,686	1,532	1,641	1,680	1,755	1,678	1,761	1,619	1,721
Experience 11-20	1,554	1,692	1,618	1,742	1,569	1,714	1,692	1,784	1,715	1,806	1,662	1,777
Experience 20+	1,523	1,703	1,654	1,769	1,617	1,746	1,718	1,800	1,758	1,823	1,705	1,792

Worker types: Firm-to-firm transitions

Back

Tenure:	Short						Long					
Worker type:	Low-wage		Mid-wage		High-wage		Low-wage		Mid-wage		High-wage	
Gender:	F	M	F	M	F	M	F	M	F	M	F	M
Share doing JTJ												
Experience 0-5	0.14	0.09	0.08	0.07	0.10	0.09	0.11	0.11	0.08	0.10	0.10	0.11
Experience 6-10	0.10	0.08	0.08	0.08	0.10	0.09	0.13	0.13	0.11	0.12	0.15	0.16
Experience 11-20	0.08	0.07	0.07	0.07	0.09	0.09	0.09	0.09	0.08	0.08	0.13	0.15
Experience 20+	0.08	0.07	0.08	0.08	0.07	0.07	0.06	0.06	0.06	0.06	0.10	0.10

- Within a latent type, workers tend to relocate less as experience accumulates.
- Across latent types, high-wage workers move more within specific experience groups.

Worker types: Employment to non-employment transitions

Back

Tenure:	Short						Long					
Worker type:	Low-wage		Mid-wage		High-wage		Low-wage		Mid-wage		High-wage	
Gender:	F	M	F	M	F	M	F	M	F	M	F	M
Share doing E-NE												
Experience 0-5	0.12	0.10	0.15	0.10	0.15	0.14	0.10	0.10	0.12	0.11	0.16	0.16
Experience 6-10	0.11	0.09	0.14	0.10	0.16	0.13	0.15	0.11	0.16	0.13	0.23	0.18
Experience 11-20	0.12	0.11	0.11	0.09	0.13	0.11	0.12	0.09	0.09	0.08	0.18	0.15
Experience 20+	0.14	0.11	0.10	0.10	0.10	0.10	0.08	0.08	0.05	0.05	0.13	0.12

- Within a latent type, workers tend to relocate less as experience accumulates.
- Across latent types, high-wage workers move more within specific experience groups.

Worker types: Non-employment to employment transitions

Back

Tenure:	Short						Long					
Worker type:	Low-wage		Mid-wage		High-wage		Low-wage		Mid-wage		High-wage	
Gender:	F	M	F	M	F	M	F	M	F	M	F	M
Share doing NE-E												
Experience 0-5	0.35	0.37	0.31	0.26	0.41	0.34	0.37	0.35	0.26	0.22	0.32	0.41
Experience 6-10	0.32	0.28	0.27	0.25	0.34	0.27	0.36	0.23	0.26	0.18	0.30	0.23
Experience 11-20	0.27	0.27	0.25	0.23	0.28	0.25	0.33	0.21	0.25	0.15	0.31	0.22
Experience 20+	0.26	0.25	0.23	0.25	0.23	0.23	0.22	0.22	0.15	0.14	0.22	0.19

- Within a latent type, workers tend to relocate less as experience accumulates.
- Across latent types, high-wage workers move more within specific experience groups.

Counts by Type, Experience, and Tenure

Back

Tenure:	Short			Long		
Worker type:	Low-wage	Mid-wage	High-wage	Low-wage	Mid-wage	High-wage
N workers in $t = 1$						
Experience 0-5	3,836	2,596	2,198	1,447	1,003	948
Experience 6-10	4,933	3,420	3,049	8,068	6,099	4,987
Experience 11-20	5,806	4,494	4,178	20,690	20,369	14,816
Experience 20+	2,930	2,284	2,089	15,564	17,288	12,068
N observations						
Experience 0-5	54,040	34,567	32,025	65,073	49,024	42,802
Experience 6-10	107,367	75,387	75,218	495,140	359,858	298,243
Experience 11-20	144,362	102,725	123,944	1,463,681	1,293,391	977,979
Experience 20+	78,188	56,918	71,314	1,388,600	1,481,206	1,038,429

Share of Women by Type, Experience, and Tenure

Back

Tenure:	Short			Long		
Worker type:	Low-wage	Mid-wage	High-wage	Low-wage	Mid-wage	High-wage
Experience						
Experience 0-5	0.51	0.53	0.59	0.49	0.51	0.54
Experience 6-10	0.50	0.48	0.56	0.51	0.49	0.57
Experience 11-20	0.47	0.47	0.47	0.48	0.50	0.51
Experience 20+	0.45	0.43	0.43	0.47	0.47	0.47

Worker types differ in their preferences over firm classes

Back

Tenure:	Short				Long			
Experience:	0-5	6-10	11-20	20+	0-5	6-10	11-20	20+
Panel A: mean rank deviation in preferences γ								
Men	2.40	2.13	2.44	2.13	2.76	2.40	2.76	1.64
Women	2.00	3.20	2.13	2.62	3.33	1.87	3.29	2.13
Panel B: $cor(\gamma_{lkx}, \mu_{lkx})$								
Men	0.01	0.49	0.44	0.27	0.03	0.54	0.56	0.49
Women	0.11	0.13	0.18	0.13	0.10	0.23	-0.09	0.55

Cross-gender correlations of estimated parameters

Back

Worker type:	Low-wage	Mid-wage	High-wage
Panel A: *Log wage* μ_{lkx}			
Experience 0-5	0.92	0.92	0.90
Experience 6-10	0.89	0.94	0.94
Experience 11-20	0.93	0.96	0.96
Experience 20+	0.92	0.95	0.96

- Strong correlations between wages of both genders over firm classes.

Cross-gender correlations of estimated parameters

Back

Worker type:	Low-wage	Mid-wage	High-wage
Panel B: *Job offer arrival rate* λ_{lkx}			
Experience 0-5	0.26	0.84	0.54
Experience 6-10	0.60	0.81	0.42
Experience 11-20	-0.09	0.35	0.28
Experience 20+	0.22	0.39	0.03

- Job offer rates b/w female and male high-experience, high-wage workers are orthogonal.

Cross-gender correlations of estimated parameters

Back

Worker type:	Low-wage	Mid-wage	High-wage
Panel C: *Preferences* γ_{lkx}			
Experience 0-5	0.41	0.02	0.12
Experience 6-10	0.68	0.30	0.05
Experience 11-20	0.80	0.16	-0.18
Experience 20+	0.77	0.62	0.32

- Female and male preferences tend to align strongly within latent types at higher experience.
- Zero or negative correlation for high-wage workers with 6-10 and 11-20 years of experience.

Cross-gender correlations of estimated parameters

Back

Worker type:	Low-wage	Mid-wage	High-wage
Panel D: *Exit rates* δ_{lkx}			
Experience 0-5	0.12	0.42	0.16
Experience 6-10	0.74	0.73	0.76
Experience 11-20	0.84	0.71	0.78
Experience 20+	0.77	0.76	0.39

- Strong correlations in the exit parameter between male and female workers.

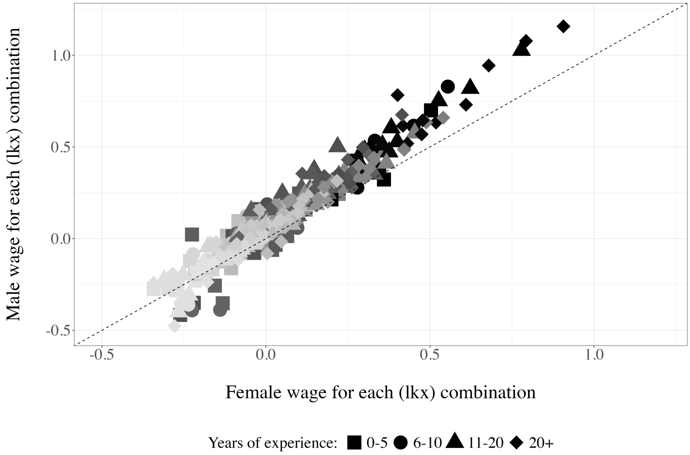
Cross-gender correlations of estimated parameters

Back

Worker type:	Low-wage	Mid-wage	High-wage
Panel E: *Entry rates* ψ_{lkx}			
Experience 0-5	0.24	-0.03	0.43
Experience 6-10	0.01	-0.25	0.10
Experience 11-20	-0.17	-0.01	0.01
Experience 20+	0.20	0.03	0.15

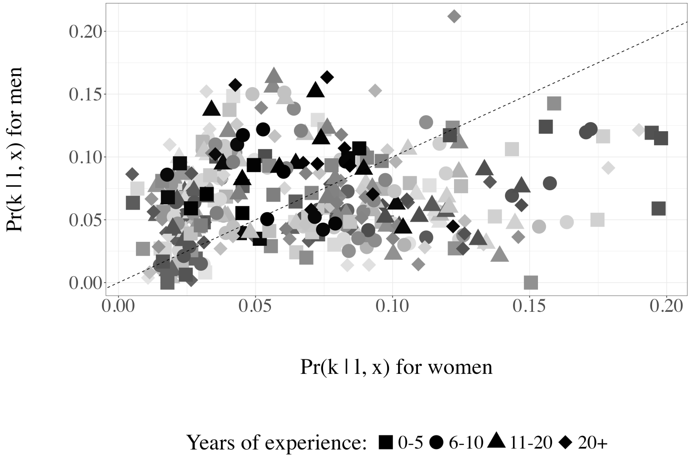
- Weak correlations between female and male entry rates across *all* worker types.
- Particularly among workers more likely to be in child-rearing ages.

Gender differences in wages



- **Lighter colors**: low-wage workers in low-paying firms.
- **Darker colors**: high-wage workers in high-paying firms.
- Strong correlation between female and male wages.
- Gender wage gaps are higher with experience.

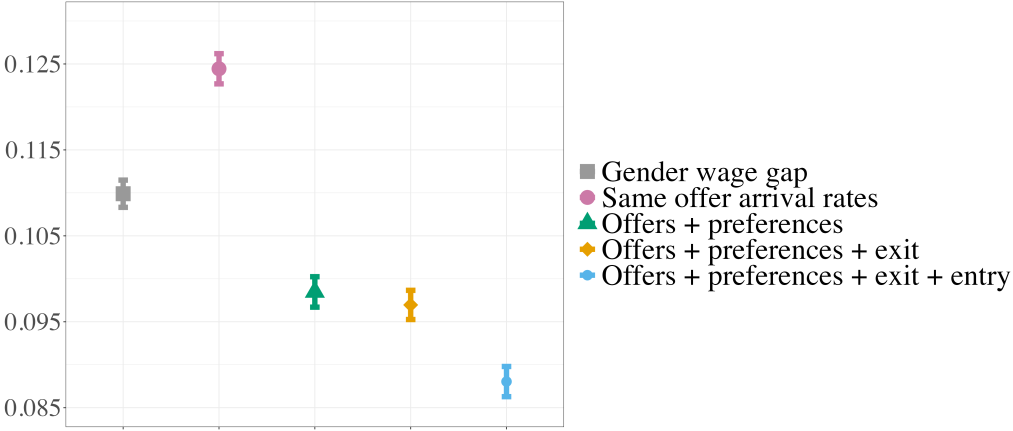
Gender differences in worker-firm allocations



- Men and women are unequally distributed across firms.
- High-experience, high-wage men are more likely than their female counterparts to work at high-paying firms.
- No strong wage sorting for either gender. Stronger tendency of wage sorting for men.

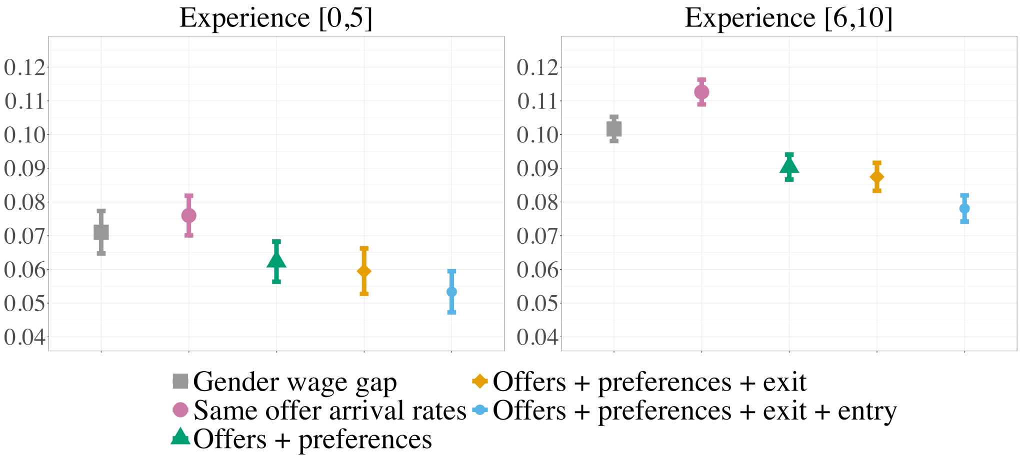
Worker Sorting and the Gender Wage Gap

Full sample



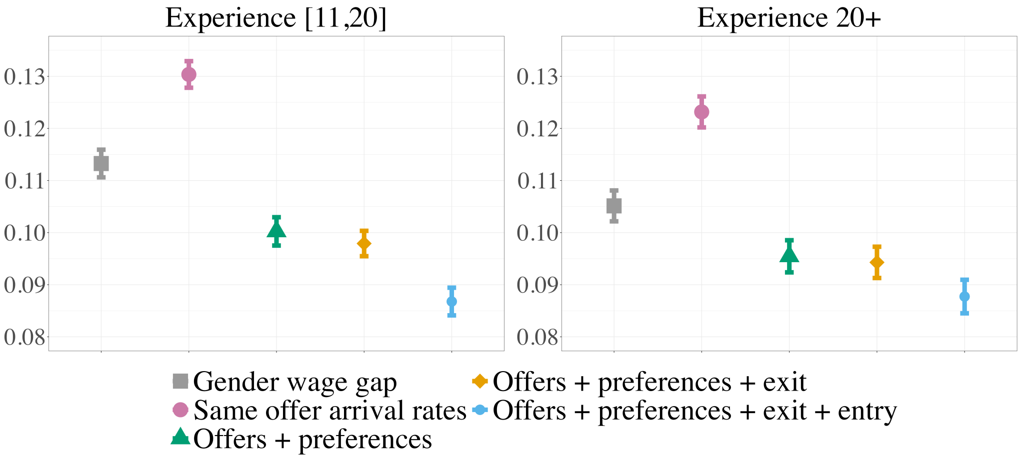
Worker Sorting and the Gender Wage Gap

By experience groups Back



Worker Sorting and the Gender Wage Gap

By experience groups Back



Appendix Tables and Figures

Workers' characteristics across selection steps

	Women							
	Earnings	Hours	Age	Experience	Tenure	% Part-Time	% Managers	N
Main job	20,678	1,225	41	17	5	29%	14%	1,378,721
Ile de France only	26,843	1,281	41	17	5	24%	24%	292,671
30+ days contract	27,030	1,289	41	17	5	24%	23%	289,443
+wages, +hours	27,021	1,292	41	17	5	23%	23%	287,211
Employed in Jan 2015	31,002	1,412	43	20	6	21%	26%	188,054
Never in Agriculture	31,005	1,412	43	20	6	21%	26%	187,926
Part-time & Full-time only	31,986	1,420	43	20	6	22%	27%	185,307
Aged 25-55	32,734	1,447	40	17	6	21%	29%	128,853
Never in seasonal/internship/domicile	34,359	1,493	41	17	6	19%	30%	117,314
Only in firms in DADS Postes	39,012	1,618	41	17	7	17%	34%	80,967

Workers' characteristics across selection steps

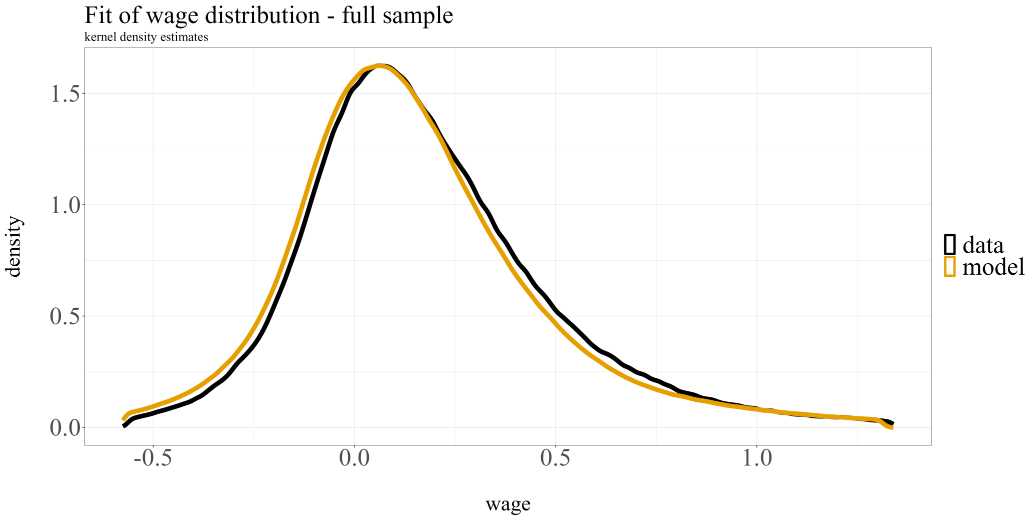
	Men							
	Earnings	Hours	Age	Experience	Tenure	% Part-Time	% Managers	N
Main job	27,820	1,343	40	17	5	14%	19%	1,441,594
Ile de France only	36,887	1,385	41	17	5	15%	31%	309,766
30+ days contract	37,177	1,396	41	18	5	14%	30%	306,025
+wages, +hours	37,152	1,400	41	17	5	14%	30%	303,896
Employed in Jan 2015	43,769	1,544	43	20	6	12%	34%	191,760
Never in Agriculture	43,793	1,544	43	20	6	12%	34%	191,482
Part-time & Full-time only	43,796	1,544	43	20	6	12%	34%	191,470
Aged 25-55	42,828	1,560	40	17	6	10%	35%	135,095
Never in seasonal/internship/domicile	45,794	1,623	41	18	6	8%	38%	122,378
Only in firms in DADS Postes	50,891	1,711	41	18	7	5%	42%	84,191

Firms' characteristics across selection samples

	Sample 1	Sample 2	Sample 3 (analysis)
N firms	100,424	48,667	25,925
N workers	73	145	250
N workers in Ile de France	27	53	92
Share of women	46%	41%	43%
Share of women in Ile de France	45%	40%	42%
Share of women that are managers	18%	26%	30%
Share of women in the board	36%	37%	39%
Share of men	54%	59%	57%
Share of men in Ile de France	55%	60%	58%
Share of men that are managers	25%	31%	36%
Share of men in the board	64%	63%	61%
Average earnings (EUR)	23,672	32,807	34,316
Median earnings (EUR)	22,363	30,545	31,688
Average hours	1,056	1,365	1,376
Median hours	1,106	1,457	1,498
Average hourly wages (EUR)	20	23	24
Median hourly wages (EUR)	19	21	22
Share with part-time contracts	7%	10%	9%
Share part-time and females	4%	6%	6%

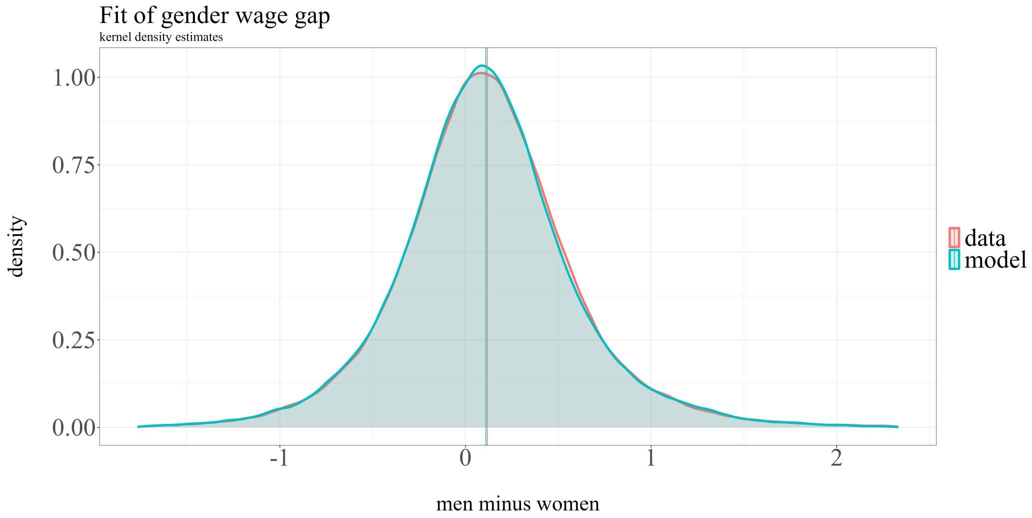
Model fit 1/2

BACK



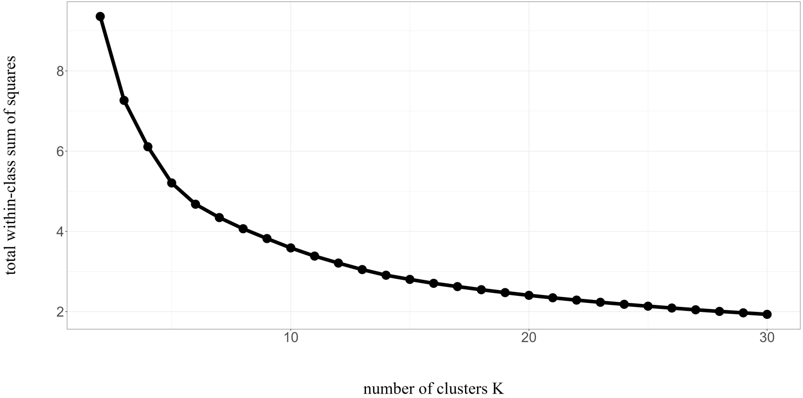
Model fit 2/2

BACK



Optimal number of clusters

BACK



- I choose K = 15, seeking a balance b/w minimising total intra-class variation and ensuring sufficient observations.

Stability of Firm Clusters

Firm class:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Avg Jaccard similarity	0.56	0.67	0.81	0.64	0.58	0.65	0.64	0.56	0.61	0.53	0.57	0.68	0.52	0.66	0.86

1. For a fixed K, I generate new firm-level datasets through random sampling with replacement, maintaining the original dataset's size.
2. I cluster the newly sampled data.
3. Using the Jaccard similarity, I then identify the most similar cluster in the new clustering for each cluster in the original classification, repeating this process 100 times.
4. The table shows the average similarity computed for each firm class across these repetitions.

Worker sorting and the gender wage gap

3 worker types and 15 firm classes

	(1)	(2)	(3)	(4)	(5)
	Full Sample	First-period years of experience:			
		0-5	6-10	11-20	20+
Gender wage gap	0.110	0.071	0.102	0.113	0.105
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offer arrival rates	0.124	0.076	0.113	0.130	0.123
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences	0.098	0.062	0.090	0.100	0.095
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences + exit	0.097	0.059	0.087	0.098	0.094
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences + exit + entry	0.088	0.053	0.078	0.087	0.088
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)

Worker sorting and the gender wage gap

3 worker types and 10 firm classes

	(1)	(2)	(3)	(4)	(5)
	Full Sample	First-period years of experience:			
		0-5	6-10	11-20	20+
Gender wage gap	0.114	0.070	0.103	0.115	0.114
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offer arrival rates	0.134	0.072	0.113	0.138	0.143
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences	0.102	0.062	0.093	0.103	0.102
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences + exit	0.102	0.059	0.090	0.102	0.102
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences + exit + entry	0.090	0.052	0.079	0.089	0.091
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)

Worker sorting and the gender wage gap

3 worker types and 5 firm classes

	(1)	(2)	(3)	(4)	(5)
	Full Sample	First-period years of experience:			
		0-5	6-10	11-20	20+
Gender wage gap	0.113	0.072	0.104	0.115	0.111
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offer arrival rates	0.125	0.078	0.117	0.130	0.120
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)
Same offers + preferences	0.104	0.063	0.094	0.105	0.103
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences + exit	0.103	0.061	0.092	0.104	0.103
	(0.001)	(0.003)	(0.002)	(0.001)	(0.002)
Same offers + preferences + exit + entry	0.090	0.051	0.077	0.089	0.093
	(0.001)	(0.004)	(0.002)	(0.002)	(0.002)

Appendix

Sample Selection

BACK

- Île de France

- Workers employed in 01/2015
  - 25-55 years old
  - drop agricultural sector
  - working in firms that hire both genders and are active for five years
  - full-time/part-time contracts lasting at least 30 days
  - drop apprenticeship/seasonal/domicile contracts
  - track them monthly using starting and ending dates of job contracts until 12/2019

- Outcome variable: hourly wages

Sample description

Statistical model

Objectives:

- Estimate the **sorting component** of the gender wage gap
  - ◇ share of gender wage gap due to gender differences in worker-firm allocations
- Quantify the relative **importance of mobility components** in explaining it
  - ◇ disentangle opportunities and preferences from firm-to-firm transitions
  - ◇ leave mobility in and out of employment unrestricted

To this purpose, I need to:

1. Predict mobility transitions of worker i across firms, and in and out of employment

   → Obtain stationary worker-firm allocations (sorting)

2. Predict average wage worker i would earn across firms

   → Analyse gender wage gap under counterfactual scenarios

Extracting information from firm-to-firm transitions

- Under revealed preference argument, firm-to-firm transitions are informative about:
  - ◇ Offer arrival rates (representing employment opportunities)
  - ◇ Choice to accept (revealing worker preferences)
- Denote j current firm and j' subsequent firm, with $j \neq j'$. Firm-to-Firm modelled as:

$$\Pr_i\left(j' \mid j\right) = \underbrace{\Pr_i\left(\text{Offer from } j'\right)}_{\text{employment opportunity}} \times \underbrace{\Pr_i\left(j' \succ j \mid \text{Offer from } j'\right)}_{\text{choice probability}}$$

- Identification challenge (**!**)
  - → Cannot work at the i-j level
  - → Need additional assumptions to disentangle the two channels

Identifying assumptions

1. Worker and firm unobserved heterogeneity is discretised
   - Workers are associated to a finite number of *types*, $l_i \in \{1, \ldots, L\}$
   - Firms are associated to a finite number of *classes*, $k_j \in \{1, \ldots, K\}$ and $k_0 = 0$

2. Workers of a given type have common preferences over firms of a given class
   - Consider workers of type l, working in firms of class k . Their utility is:

$$U_{ij} = \gamma_{lk} + \epsilon_{ij}$$

   - ϵ_{ij} is an idiosyncratic utility draw, specific to worker i and firm j
     - → e.g.: choice to move is influenced by moving costs
   - When choosing between two firms j and j', worker i compares U_{ij} and $U_{ij'}$

Data structure

Observed

- Workers $i \in \{1, \dots, N\}$; Firms: $j \in \{0, 1, \dots, J\}$; $j = 0$ is non-employment
- Worker characteristics: experience and tenure (x_{it}), gender $\left(g_i = \{F, M\}\right)$
- Sequence of firm identifiers: $\left(j(i,1), \dots, j(i,T)\right)$
- Sequence of log-hourly wages: $\left(y_{i1}, \dots, y_{iT}\right)$
- Sequence of mobility indicators: $\left(s_{i,1}, \dots, s_{i,T-1}\right)$
  - $s_{i,t} = 1$ if $j(i,t) \neq j(i,t+1)$, else $s_{i,t} = 0$. Three types of transitions:
    - $\rightarrow$ *Firm to Firm*: $j \neq j'$, with $j, j' \in \{1, \dots, J\}$ [informative of opportunities and preferences]
    - $\rightarrow$ *Employment to Non-employment*: $j \neq j'$, with $j \in \{1, \dots, J\}$ and $j' = 0$ [**unrestricted**]
    - $\rightarrow$ *Non-employment to Employment*: $j \neq j'$, with $j = 0$ and $j' \in \{1, \dots, J\}$ [**unrestricted**]

Observed mobility

- $s_{i,t} = 1$ if $j(i,t) \neq j(i,t+1)$, else $s_{i,t} = 0$
- Denote j as current firm, and j' as subsequent firm
- Worker i can make three types of transitions:

1. *Firm to Firm*: $j \neq j'$, with $j, j' \in \{1, \ldots, J\}$

   $\rightarrow$ informative of offer arrival rates and worker preferences

2. *Employment to Non-employment*: $j \neq j'$, with $j \in \{1, \ldots, J\}$ and $j' = 0$

   $\rightarrow$ exit rates, left **unrestricted**

3. *Non-employment to Employment*: $j \neq j'$, with $j = 0$ and $j' \in \{1, \ldots, J\}$

   $\rightarrow$ entry rates, left **unrestricted**

Modelling mobility BACK

Firm-to-Firm transitions

- Denote k the current firm class, k' the subsequent firm class

$$\Pr\Big(k' \mid k, l, x\Big) = \underbrace{\lambda_{lxk'}}_{\text{offer probability}} \times \underbrace{\Pr_{lx}(k' \succ k)}_{\text{choice probability}} = \lambda_{lxk'} \times \frac{\gamma_{lxk'}}{\gamma_{lxk} + \gamma_{lxk'}}$$

- γ_{lxk} captures (l, x)-type worker perceived value of class k
- Worker perceived values guide the choice probability
  - $\rightarrow$ The higher the perceived value of the poacher with respect to the incumbent, the higher the probability the worker chooses to move
- Under discretisation of heterogeneities, λ_{lxk} and γ_{lxk} are separately identified
  - $\rightarrow$ Identification

Modelling mobility

BACK

Into and out of non-employment

Denote k the current firm class, k' the subsequent firm class

- From $k \in \{1, \ldots, K\}$ to $k' = 0$: $\Pr\left(0 \mid k, l, x\right) = \delta_{lxk}$
- From $k = 0$ to $k' \in \{1, \ldots, K\}$: $\Pr\left(k' \mid 0, l, x\right) = \psi_{lxk'}$
- Simple frequencies

Modelling mobility

BACK

Probability of staying in current firm class/employment status

- For employed workers, $k \geq 1$, the probability of staying with the same firm is:

$$\Pr\big(\text{no move} \mid k, l, x\big) = 1 - \delta_{lxk} - \sum_{k'=1}^{K} \left(\lambda_{lxk'} \frac{\gamma_{t'}}{\gamma_{lxk} + \gamma_{lxk'}} \right)$$

- The probability of staying into non-employment is:

$$\Pr\big(\text{no move} \mid 0, l, x\big) = 1 - \sum_{k'=1}^{K} \psi_{lxk'}$$

Initial matching distribution

BACK

- At $t = 1$, worker i with observed x_{i1} starts being employed
- Initial observed heterogeneity determines distribution of initial matches

$$\Pr\left(l_i, k_{j(i,1)} \mid x_{i1}\right) = m_0\left(l, k \mid x\right)$$

- Simple frequencies
- At each $t \geq 1$, I observe whether the worker separates or not
  - $\rightarrow$ Model mobility process as described above
- Recall that x contains interactions of gender, experience, and tenure

Modelling wages

BACK

- Hourly wages are drawn from a static worker-firm-specific log-normal distribution:

$$\ln f\left(y_{it} \mid l, x, k\right) \sim N\left(\mu_{lxk}, \sigma_{lxk}\right)$$

- Allows for wage complementarities between workers and firms
- Note: wages are previously residualised on occupational dummies
  - effects estimated on female sample only
  - model part of wages unexplained by returns to skills

Identification BACK

Simple example

- Suppose there is one type of worker and two classes of firms, A and B.
- Take all transitions *within class* A:
    - ◇ In expectations, workers are indifferent between firms of class A
    - ◇ Assume that workers employed in A accept offers from class-A firms with prob. $\frac{1}{2}$
    - → **Expected number of offers from class-A firms**: $2 \times$ number of *within class*-A moves
- Take all transitions *within class* B:
    - ◇ Get **expected number of offers from class-B firms** in similar manner
- Take *between-class* transitions:
    - ◇ Given expected number of offers, get **expected share of accepted offers**
- Revealed preference argument:
    - ◇ Suppose share of accepted class-A offers $>$ share of accepted class-B offers
    - → workers prefer firms in class A

Identification 1/

BACK

- Note: vectors γ_{lx} and γ'_{lx} equivalent if one scalar multiple than the other

  $\rightarrow$ normalise $\sum_{k=1}^{K} \gamma_{lxk} = 1$

- Under discretisation of heterogeneities, $\lambda_{lxk'}$ and γ_{lxk} separately identified using:
  - Frequencies of transition probabilities $\Pr\left(k' \mid k, l, x\right)$
  - Normalisation $\sum_{k=1}^{K} \gamma_{lxk} = 1$

Identification 2/ BACK

- First, pin down λ_{lxk} for any (l, x, k) match using within-firm-class variation

$$\Pr\Big(k' = k \;\Big|\; k, l, x\Big) = \lambda_{lxk}\ \frac{1}{2}$$

  - ◇ Under assumption of no empty cell
  - ◇ Given structure on choices, no loss of generality in setting it $= \frac{1}{2}$

Identification 3/ BACK

- Second, pin down $\Pr_{lx}\left(k' \succ k\right)$ using info on frequencies and knowledge of λ_{lxk}

$$\Pr\left(k' \neq k \mid k, l, x\right) = \lambda_{lxk'} \Pr_{lx}\left(k' \succ k\right)$$

- Finally, pin down $\frac{\gamma_{lxk'}}{\gamma_{lxk}}$ using:

$$\frac{P_{lx}\left(k' \succ k\right)}{P_{lx}\left(k \succ k'\right)} = \frac{\gamma_{lxk'}}{\gamma_{lxk}}$$

$$\sum_{k=1}^{K} \gamma_{lxk} = 1$$

Discretise worker and firm heterogeneity

Back

- Finite number of firm latent *classes*:
  - Classify firms into classes using data on firms' characteristics
  - Firms in a given class share similar characteristics
    - gender-specific wage distributions, female shares, size
- Finite number of worker latent *types*:
  - Classify workers into types using data variations in wages and mobility
  - Latent types interact with combinations of gender, experience, and tenure
    - $g_i \in \{F, M\}$, $\text{Exp}_{it} \in \{\text{0-5},\ \text{6-10},\ \text{11-20},\ \text{20+}\}$, $\text{Short-tenure}_{it} = \mathbb{1}\left\{\text{Tenure}_{it} \leq 2 \text{ years}\right\}$
  - In expectations, (l, x)-type workers earn similar wages and have similar mobility patterns
  - By interacting gender with time-varying characteristics, men's and women's wage profiles and mobility patterns can vary over time within a latent type

Sorting Parameter

Back

- Build transition matrix using the estimated structural parameters

$$\hat{M}_{lx} = \begin{bmatrix} \hat{m}_{lx}(1,1) & \hat{m}_{lx}(1,2) & \dots & \hat{m}_{lx}(1,K) & \hat{\delta}_{lx}(1,0) \\ \hat{m}_{lx}(2,1) & \dots & \dots & \dots & \hat{\delta}_{lx}(2,0) \\ \dots & \dots & \dots & \dots & \dots \\ \hat{m}_{lx}(K,1) & \hat{m}_{lx}(K,2) & \dots & \hat{m}_{lx}(K,K) & \hat{\delta}_{lx}(K,0) \\ \hat{\psi}_{lx}(0,1) & \hat{\psi}_{lx}(0,2) & \dots & \hat{\psi}_{lx}(0,K) & 0 \end{bmatrix}$$

- Normalise $\hat{M}_{lx}$ and solve: $\hat{M}_{lx}^T \, s^* = s^*$
- s^* : $(K+1) \times 1$ vector of $\Pr\Big(k \mid l, x\Big) \; \forall \, k \in \{0, 1, \dots, K\}$

Empirical distribution of matches

Back

$$\Pr\big(l, x, k\big) = \Pr^*\big(k \mid l, x\big)\ \Pr\big(l,\, x\big)$$

$$\Pr\big(l,\, x\big) = \sum_{i=1}^{N} p_i\big(l \mid \theta^{(m)}, x_{i1}, C\big) \sum_t \mathbb{1}\{x_{it} = x\} \quad \big(\text{normalised to make it a probability}\big)$$

EM Algorithm

BACK

- E step: For given parameters $\theta^{(m)}$ and a firm classification C, compute the posterior probability that worker i is of type $l = 1, \ldots, L$

$$p_i\big(l \mid \theta^{(m)}, x_{i1}, C\big) = \frac{\mathcal{L}_i\big(\theta^{(m)} \mid l_i, x_{i1}, C\big)}{\sum_{l=1}^{L} \mathcal{L}_i\big(\theta^{(m)} \mid l_i, x_{i1}, C\big)}$$

- M step: Maximise the expected log-likelihood with respect to the parameter of interest θ

$$\theta^{(m+1)} = \text{argmax}_\theta \sum_i \sum_l p_i\big(l \mid \theta^{(m)}, x_{i1}, C\big)\ ln\mathcal{L}_i\big(\theta \mid l_i, x_{i1}, C\big)$$

EM Algorithm

BACK

The wage segment of the expected log-likelihood writes:

$$W = \sum_i \sum_l p_i\big(l \mid \theta^{(m)}, x_{i1}, C\big) \sum_k \sum_{t=1}^{T_i} \mathbb{1}\{k_{j(i,t)} = k, x_{it} = x\} \ln f(y_{it}|l, x, k)$$

EM Algorithm

BACK

Taking derivatives with respect to μ_{lxk} and σ_{lxk}

$$\mu_{lxk}^{(m+1)} = \frac{\sum_i p_i(l \mid \theta^{(m)}, x_{i1}, C) \sum_{t=1}^{T_i} \mathbb{1}\{k_{j(i,t)} = k, x_{it} = x\} y_{it}}{\sum_i p_i(l \mid \theta^{(m)}, x_{i1}, C) \sum_{t=1}^{T_i} \mathbb{1}\{k_{j(i,t)} = k, x_{it} = x\}}$$

$$\sigma_{lxk}^{(m+1)} = \sqrt{\frac{\sum_i p_i(l \mid \theta^{(m)}, x_{i1}, C) \sum_{t=1}^{T_i} \mathbb{1}\{k_{j(i,t)} = k, x_{it} = x\}(y_{it} - \mu_{lxk})^2}{\sum_i p_i(l \mid \theta^{(m)}, x_{i1}, C) \sum_{t=1}^{T_i} \mathbb{1}\{k_{j(i,t)} = k, x_{it} = x\}}}$$

EM Algorithm

BACK

$$m_0\left(l, k \mid x\right) = \frac{\sum_i p_i\left(l \mid \theta^{(m)}, x_{i1}, C\right) \mathbb{1}\left\{k_{j(i,1)} = k, x_{i1} = x\right\}}{\sum_l \sum_k \sum_i p_i\left(l \mid \theta^{(m)}, x_{i1}, C\right) \mathbb{1}\left\{k_{j(i,1)} = k, x_{i1} = x\right\}}$$

EM Algorithm

BACK

For $k, k' \in \{0, 1, \ldots, K\}$, define:

- $n^{(m)}_{lxk\neg} = \sum_i p_i\big(l \mid \theta^{(m)}, x_{i1}, C\big) \sum_{t=1} \mathbb{1}\big\{k_{j(i,t)} = k, x_{it} = x, s_{it} = 0\big\}$
- $n^{(m)}_{lxkk'} = \sum_i p_i\big(l \mid \theta^{(m)}, x_{i1}, C\big) \sum_{t=1} \mathbb{1}\big\{k_{j(i,t)} = k, k_{j(i,t+1)} = k', x_{it} = x, s_{it} = 1\big\}$

EM Algorithm

BACK

For a given l, the segment of the expected log-likelihood to update ψ is:

$$n_{lx0\neg}^{(m)} \ln\left(1 - \sum_{k'=1}^{K} \psi_{lxk'}\right) + \sum_{k'=1}^{K} n_{lx0k'}^{(m)} \ln(\psi_{lxk'})$$

Taking derivatives we obtain the M-step updating formula for ψ

$$\psi_{lxk'}^{(m+1)} = \frac{n_{lx0k'}^{(m)}}{n_{lx0\neg}^{(m)} + \sum_{k'=1}^{K} n_{lx0k'}^{(m)}} \tag{1}$$

EM Algorithm

BACK

- For a given l, the remaining segment of the expected log-likelihood is:

$$\sum_{k=1}^{K} n_{lxk\neg}^{(m)} \ln(M_{lxk\neg}) + \sum_{k=1}^{K} n_{lxk0}^{(m)} \ln(M_{lxk0}) + \sum_{k=1}^{K}\sum_{k'=1}^{K} n_{lxkk'}^{(m)} \ln(M_{lxkk'})$$

- Under the parametric specification provided above, this segment of the expected log-likelihood is not linear in the parameters of interest (specifically, the one related to job-to-job transitions). I therefore consider the minorising function proposed by Lentz et al. (2023).

EM Algorithm

BACK

$$H(M|\theta^{(m)}) = \sum_{k=1}^{K} n^{(m)}_{lxk\neg} \ln(\underline{\mathrm{M}}_{lxk\neg}) + \sum_{k=1}^{K} \sum_{k'=0}^{K} n^{(m)}_{lxkk'} \ln(M_{lxkk'})$$

Given $\theta^{(m)}$ obtained at the m-step of the EM algorithm, I update δ, γ, λ by maximising $H(M|\theta^{(m)})$ using an iterative procedure.

EM Algorithm

BACK

First define:

- $\tilde{n}^{(s)}_{lxkk'} = n^{(m)}_{lxk\neg} \dfrac{\lambda^{(s)}_{lxk'}(1 - P^{(s)}_{lxkk'})}{M^{(s)}_{lxk\neg}}$ the predicted number of lg-type stayers that receive an offer from k' but prefer to stay in k.

- $\hat{n}^{(s)}_{lxk} = n^{(m)}_{lxk\neg} \dfrac{1 - \delta^{(s)}_{lxk} - \sum\limits_{k'=1}^{K} \lambda^{(s)}_{lk'}}{M^{(s)}_{lxk\neg}}$ the predicted number of lg-type stayers that stay because they receive no offer/layoff.

EM Algorithm

BACK

$$\gamma_{lxk}^{(s+1)} = \frac{\sum_{k'=1}^{K} (\tilde{n}_{lxkk'}^{(s)} + n_{lxk'k}^{(m)})}{\sum_{k'=1}^{K} \left(\frac{\tilde{n}_{lxkk'}^{(s)} + n_{lxkk'}^{(m)} + \tilde{n}_{lxk'k}^{(s)} + n_{lxk'k}^{(m)}}{\gamma_{lxk}^{(s)} + \gamma_{lxk'}^{(s)}} \right)}$$

EM Algorithm

BACK

$$\lambda_{lxk'}^{(s+1)} = \frac{\sum_{k=1}^{K} \left(\tilde{n}_{lxkk'}^{(s)} + n_{lxkk'}^{(m)} \right)}{\sum_{k=1}^{K} n_{lxk0}^{(m)} + \sum_{k=1}^{K} \hat{n}_{lxk}^{(s)} + \sum_{k=1}^{K} \sum_{k'=1}^{K} \tilde{n}_{lxkk'}^{(s)} + \sum_{k=1}^{K} \sum_{k'=1}^{K} n_{lxkk'}^{(m)}}$$

EM Algorithm

BACK

$$\delta_{lxk}^{(s+1)} = \frac{n_{lxk0}^{(m)}\left(1 - \sum_{k'=1}^{K} \lambda_{lxk'}^{(s+1)}\right)}{n_{lxk0}^{(m)} + \hat{n}_{lxk}^{(s)}}$$

EM Algorithm

BACK

For given value of $\theta^{(m)}$, the sequence $H(M|\theta^{(m)})$ increases at each iteration step s of the MM algorithm. It is thus not strictly necessary to wait for convergence, the algorithm can be stopped at any time. I iterate the MM algorithm 200 times before it delivers the updated values $\delta^{(m+1)}$, $\gamma^{(m+1)}$, and $\lambda^{(m+1)}$.

MM for Bradley-Terry model - Hunter (2004)

$$P\big(\text{individual } i \text{ beats individual } j\big) = \frac{\gamma_i}{\gamma_i + \gamma_j}$$

We observe a number of pairings among individuals and we want to estimate $\gamma_1, \ldots, \gamma_N$ using MLE. The log-likelihood is:

$$\mathcal{L}(\gamma) = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} \big[ln\gamma_i - ln(\gamma_i + \gamma_j) \big]$$

- ω_{ij} is the number of times i beats j
- we cannot separate the components of γ due to $ln(\gamma_i + \gamma_j)$

MM for Bradley-Terry model - Hunter (2004)

BACK

The concavity of the logarithm implies for positive x and y:

$$-lnx \geq 1 - lny - \frac{x}{y} \qquad \text{with equality if } x = y$$

Fix $\gamma^{(s)}$ and define the function:

$$Q_s(\gamma) = \sum_i \sum_j \omega_{ij} \left[ln\gamma_i + 1 - ln(\gamma_i^{(s)} + \gamma_j^{(s)}) - \frac{\gamma_i + \gamma_j}{\gamma_i^{(s)} + \gamma_j^{(s)}} \right]$$

- $Q_s(\gamma) \leq \mathcal{L}(\gamma)$ with equality if $\gamma = \gamma^{(s)}$
- iterative algorithm with $\gamma^{(s+1)} = \text{argmax}\, Q_s(\gamma)$
- maximization of $Q_s(\gamma)$ equal to maximization of each γ_i separately
- $\gamma_i^{(s+1)} = W_i \left[\sum_{i \neq j} \frac{N_{ij}}{\gamma_i^{(s)} + \gamma^{(s)}} \right]^{-1}$. Renormalise at each step s.t. $\sum_i \gamma_i = 1$