Institut d'études politiques de Paris
École Doctorale de Sciences Po
Programme doctoral en économie
Département d'économie
Doctorat en Sciences économiques

Quantitative Insights into Labor Dynamics

Riddhi KALSI

Thesis supervised by
Jean-Marc ROBIN, Professeur des Universités, Sciences Po

defended in Paris on June 11, 2026

Jury:
Clément de CHAISEMARTIN, Professeur des Universités, Sciences Po
Gregory JOLIVET, Professor of Economics, University of Bristol *(reviewer)*
Costas MEGHIR, Professor of Economics, Yale University
Hélène PÉRIVIER, Directrice de recherche OFCE -Sciences Po
Fabien POSTEL-VINAY, Professor of Economics, University College London *(reviewer)*
Jean-Marc ROBIN, Professeur des Universités, Sciences Po

To the certainty in all (my) potential outcomes, Saagnik.

Acknowledgments

My deepest thanks go first to my supervisor, Jean-Marc Robin, who introduced me to the wonderful world of labor econometrics and gave me the support to grow here. Jean-marc, your expertise, and tough but gentle nature have pushed the boundaries of the work I have pursued. I am equally appreciative of the support you extended during some of the more difficult periods of these years. I am sincerely grateful to Clément de Chaisemartin. I learned a lot from working with you and from your sharp insights. Your support throughout this journey — both intellectual and professional — has meant a great deal to me. *Descriptives, first!* is now a shared ethos. I greatly benefited from the guidance and mentorship of Costas Meghir. The visit to Yale has shaped this thesis in more ways than I can list here. Thank you, Costas, for giving me a concrete sense of the landscape of this field — the clarity of your thinking and articulation has helped me immensely. I also owe a very special thanks to Rawi Abdelal. My dream of academia began when I was inspired by your teaching during my time as a consultant, and I remain deeply appreciative of the encouragement, guidance, and support you extended throughout the journey leading to this point.

I am lucky to have benefited from the expertise of Pierre Cahuc, who has always taken the time to address my (many) questions about administrative data and the French labor market. This work was not possible without your help. A special note of thanks goes to Florian Oswald for his generous comments and sustained encouragement over the course of the PhD. I wish to convey my deepest appreciation to Olivier Godechot — I have thoroughly enjoyed our interdisciplinary conversations about inequality that brightened the long days in Reims. Thank you for also aiding me in the data hunt for studying privatization! Although it is not part of this thesis, the pursuit to understand it better still remains. I wish to acknowledge the generosity and support of Hélène Périvier who took an interest in the production of this dissertation from early on and provided invaluable guidance through its development.

Conducting research and communicating it effectively are two separate yet equally essential skills. I am especially grateful to Pierre-Philippe Combes, Ghazala Azmat, Thierry Mayer, Antoine Ferey, and Claire Montialoux for helping me refine the latter, which ultimately also improved the former. Thank you for your generous and constructive feedback. I would be remiss not to thank Cormac O'Dea — I learned a lot from your teaching style, and I also thank you for preparing me for interviews about my research.

Paris gave me more than an academic home; it also gave me friendships that became family. I am particularly grateful to Fatimetou for her steadfast friendship and support during the many highs and

lows of these years. My life would not have been the same without you. *Ni le master, ni le doctorat, ni la vie*. You, Hassan, and Emin are family. Vallari, whose sisterhood has been one of the great gifts of these past years since we met in 2021, thank you for your presence, generosity, intelligence, and humor. I am still not sure how you endured all my long rants about economics, but I am grateful that you did. Kunal has been, in many ways, the younger brother I never had. It has been a privilege to witness your growth over these years. Special thanks goes to Vrinda, who has stepped up to help me in many times of need. Anmol, spending time with you has always been equal parts intellectually stimulating and immensely fun. You are responsible for my having any social life at all. Qinci, thank you for your quiet warmth, sincerity, and companionship. Each of you, in your own way, made this city feel less foreign and this journey far less lonely.

My PhD journey has been enriched by amazing, bright colleagues who are also good friends. Nathaniel, our many conversations about macroeconomics, banking, and politics have been among the intellectual highlights of these years. Thank you for also introducing Kate and me. Felipe and Dani, each conversation with you has carried warmth; I have felt seen, heard, and cared for. Sophia and Niklas, thank you for helping me navigate moments when I felt completely lost and when I was caught between a rock and a hard place with respect to data access. Mattis, you may not remember but the hour you spent on the FLFS to look up the two hours variables validated the hours story I saw through the DADS! More importantly, our many conversations over food, work, and life brought levity and perspective to this PhD journey. Andrei, I did not expect to make a great friend when going to the econometric society conference, but I am grateful for your unwavering friendship. I have been quite fortunate to have made amazing friends whilst at Yale. Big shout-out to my fellow Foxes: Vandita, Shucheng, Mwiti, Zişan, and Prita. At the economics department, I had the privilege of meeting Chi, Amelie, Sevin, and Annica, who became not only fellow economists but also close friends. I miss exploring the US and navigating economics together — thankfully, we will still have the latter.

I am also grateful for the friendships I built at Sciences Po, which made the institution feel warmer, livelier, and far more human. Ye, thank you for your friendship and for all the fun moments we shared in Paris before your own PhD journey took you to Cambridge. Dima, our conversations about imperialism, feminism, and politics remain among memorable discussions I have had over these years. The long days in Reims would have been far harder without Nelly's companionship and solidarity. Max, I am grateful not only for the solidarity forged through work in Reims, but also for the many moments of friendship and laughter outside of it. Constantin is the warmest desk neighbor one could ask for — thanks for always bringing levity into otherwise stressful days. Olivia, our conversations about feminism, academia, and life came at key points. Leonard, Ségal, and Myléne, although you were initially my seniors, I am very grateful that our paths evolved into friendship. Special thanks go to Pilar who is more popularly known for her role in administration, but she was actually my first friend when I started the PhD program. Philippe, Fatumata and Professor Traoré, thank you for the recharging coffee breaks!

I also wish to thank many colleagues at Sciences Po whose presence enriched these years in large and small ways: Daniel, Juan, Constance, Gustave, Alexandre, Norbert, Agathe, Swann, and Valentin. Thank you for the camaraderie, coffees, complaints about data access, and moments of collective survival that make doctoral life less isolating. Valentin, merci aussi d'avoir patiemment pratiqué le français avec moi over the years. Swann, it is our secret who hides information from INSEE! I am also thankful to Johanna, Olivia, Jakob, Davide, Diego, Gabrielle, Nourhan, Mathias, Paloma, and Viktor for the many conversations and shared moments that accompanied me throughout these years. Academic life is shaped as much by the people around us as by the work itself, and I feel fortunate to have shared these years with all of you.

A special note of gratitude goes to Guillaume Sarrat de Tramezaigues, Maria Öhlund, Sandrine Le Goff, Stéphanie Berrebi, Lucie Desramaut, and Melissa Mundell for their invaluable administrative support throughout these years. Navigating academia often requires navigating bureaucracy, and your efficiency, patience, and kindness made that process infinitely smoother. I am especially grateful to Hadjila Nezlioui-Serraz for her calmness, reassurance, and extraordinary patience in the face of the blizzard of administrative matters at the School of Research. Your steady presence brought clarity to moments that otherwise felt overwhelming.

Beyond academia, I have been incredibly fortunate to be surrounded by friends whose presence kept me grounded throughout this journey. Arjun, Samyukta, Dimpi, Sana, Mayank, Maryam, Aastha, Karan, Udai, Jaya, Meeta Ma'am, Saptarshi, Sahil, Pranjal, and Devina: thank you for your friendship across time zones, life stages, and long periods of absence caused by the PhD. Anirudh, I am so happy since your move to Europe and I'm looking forward to jamming together again. The continued presence of each of you in my life reminded me that there exists a world beyond deadlines, revisions, and job market papers.

To my parents, no acknowledgement could adequately capture what I owe you. Mimi and Appa, thank you for the sacrifices you made, the risks you took, and the unwavering faith you placed in me long before I had any reason to place it in myself. Everything I have achieved rests on the foundation you built for me with love and resilience.

To my brother, Aaditya, thank you for being a constant source of love, humor, acuity, and perspective throughout my life. No matter the distance or the chaos surrounding this PhD, your presence has always grounded me.

Finally, to my partner, Saagnik, thank you for walking beside me through every stage of this journey. Your patience, love, encouragement, and belief in me sustained me through moments of uncertainty, exhaustion, and doubt. You celebrated every small victory, carried me through the difficult days, and made even the hardest moments feel lighter. I cannot imagine this dissertation — or these years — without you and I am happy that I do not have to — I cannot wait to start our life together in Manchester!

Note to the Reader

The three chapters of this dissertation are self-contained research articles and can be read separately, although they contribute to an overarching theme. They are preceded by an introduction which summarizes the research presented in this dissertation. The terms "paper" or "article" are used to refer to the chapters. Chapter 2 is co-authored, which explains the use of the "we" pronoun.

Contents

List of Figures 1

List of Tables 5

Introduction 9

Chapter 1. Evaluating the Public Pay Gap by Gender 15
- 1.1 Data 18
- 1.2 Stylized facts 20
- 1.3 Model 24
- 1.4 Model Fit 29
- 1.5 Results: Public and Gender Premia in Lifetime Earnings 33
- 1.6 Discussion 39
- 1.7 Conclusion 40

Chapter 2. Job Mobility and Career Values 43
- 2.1 Data and Methodology 47
- 2.2 Estimation 52
- 2.3 Measuring Sorting: Variance and Information Decomposition 57
- 2.4 Model Fit 60
- 2.5 Results 64
- 2.6 Conclusion 83

Chapter 3. Signals, Noise, and Sorting 89
- 3.1 Introduction 89
- 3.2 The Geometry of Noise Amplification 92
- 3.3 Bias Correction Methods 97
- 3.4 Monte Carlo Evidence 102
- 3.5 Discussion 110
- 3.6 Conclusion 111

Appendix to Chapter 1 115
- A1.1 Panel Construction 115
- A1.2 Descriptives 116
- A1.3 Model 118
- A1.4 Estimating the mobility and income parameters using sequential EM 120
- A1.5 Model Fit 124

A1.6 Estimated model parameters 129
A1.7 Results: Figures 132

Appendix to Chapter 2 133
A2.1 Data 133
A2.2 Methodology Details 134
A2.3 Additional Results 138
A2.4 Relation to the AKM Literature 148
A2.5 Validation using CEM on Simulated Data 154

Appendix to Chapter 3 163
A3.1 Relevance of the AKM Methodology 163
A3.2 Proofs 163
A3.3 Additional Monte Carlo Tables 168

Résumé en Français 177

List of Figures

1.1 Distributions of Hourly and Annual Wage Changes 21

1.2 Annual wage dispersion by sector, sex and sector, over experience 23

1.3 Observed latent wage class distribution 31

1.4 Observed and predicted wage densities 33

1.6 Gender premium in lifetime earnings 37

1.5 Loss in lifetime earnings with selection Lifetime earnings percentage difference of job-for-life case and mobility case (the latter takes into account the probability of transitioning to non-employment or part-time work). 37

1.7 Percentage difference of lifetime earnings in the public and private sectors by education levels and sex 39

2.1 Log-likelihoods across the grid of (X, Y) pairs. The sub graphs look at the projections of the likelihood surface along the xy plane, yz plane, and xz plane. 55

2.2 Simulated vs. actual transition shares. Each point represents an origin-destination pair (panels a–b) or a firm-type/occupation cell (panel c). The 45-degree line indicates perfect fit. 61

2.3 Standardized within-cell wage distributions for the four largest (x, y, k) cells, pooling across gender. Each panel plots the empirical density of standardized residuals $(w - \mu_{x,y,k})/\sigma$ against the model's predicted $\mathcal{N}(0, 1)$. Cell sizes range from 950,458 to 2,454,572 spells. 62

2.4 Firm classification accuracy by firm size. Accuracy is defined as the share of firms whose assigned type under the full likelihood (wages, hazards, and destinations) coincides with their best-fit type. Bins are defined by the number of observed spells at each firm. 63

2.5 Net present value (NPV) of lifetime earnings by worker type x and firm type y and occupation k. Each cell reports the mean NPV for workers of type x employed at firms of type y in occupation k, averaged across genders and age bins. 65

2.6 Firm type profiles. Panel A: mean daily wage (firm types ordered by wage). Panel B panels show number of firms in each firm type and Panel C plots the mean firm size by firm type. 66

2.7 Additive variance decomposition of wages (Panel A) and NPV (Panel B). Each bar shows the contribution of a variance or covariance term to total variance (percentage points). Positive covariances are green, negative red, the interaction term yellow, and within-cell residual grey. Dotted and dashed reference lines mark the additive total and between-cell total respectively. 74

2.8 Mutual information decomposition: wages versus NPV. *Panel A:* unconditional MI (faded dot) versus full conditional MI (solid dot) for each factor; arrows indicate the synergy (+) or

redundancy (−) shift. *Panel B:* sum of unconditional MIs (stacked bar) versus joint MI (solid bar); the gap is net synergy. 74

2.9 Distribution of Job-to-Job Transition Gains. *Top panels*: density of NPV gains (left) and wage gains (right); red dashed line marks zero. *Bottom panels*: excess transition probability relative to the marginal destination distribution (green = above random, red = below random), binned by NPV gain (left) and wage gain (right). Excess probability leans rightward for wages but is roughly symmetric for NPV; both panels spike at zero (lateral mobility). 79

2.10 Flow Quadrant Decomposition of job Transitions. *Panel A*: quadrant shares by age bin; harvest moves (w+ opt−) dominate throughout. *Panel B*: same decomposition by gender. *Panel C*: scatter of ΔV_{wage} against ΔV_{opt} for all employed-to-employed transitions; the strong negative correlation ($r = -0.73$) confirms a systematic wage–option trade-off. 82

3.1 Average node degree as a function of per-period mobility and workers per firm under all designs split by two mobility types and two firm size distributions. The vertical axis represents the average of the effective realized node degree per Definition 2 105

3.2 Average bias over all design cells as a function of the average node degree $\bar{d}$. The right panel uses log-log axes, revealing a power-law relationship between the bias factor B_ψ and the average node degree $\bar{d}$. Each point corresponds to a unique combination of mobility rate, firm size distribution, number of firms, and workers-per-firm ratio. The tight clustering of points along a single curve confirms that $\bar{d}$ fully captures the influence of these design parameters on the bias. 106

A1.1 Share of women in the Public and Private Sectors over 2012-19 116

A1.2 Histogram of log net annual wages by sex and sector for full-time work 117

A1.3 Predicted latent wage class distribution 126

A1.4 Wage Densities by employment state: Observed vs Predicted 127

A1.5 Observed and predicted wage densities 128

A1.6 Observed and predicted wage densities by employment state 128

A1.7 Observed and predicted wage densities and wage classes 129

A1.8 Public premium in lifetime earnings: "job for life" counterfactuals with and without selection 132

A2.1 Predicted vs. observed separation probabilities across 2,731 cells defined by (k, female, age bin, x, y). Cells with fewer than 20 spells are excluded. The 45-degree line indicates perfect fit. 143

A2.2 Distribution of workers by number of employment spells within the 5-year panel window: data (dark) vs. model simulation (light). 143

A2.3 Mean spell duration by (y, k) cell: simulated vs. actual. Each point is one of 96 populated cells. 144

A2.4 Mean daily separation rate by age bin: data vs. simulation. 145

A2.5 Within-cell wage dispersion. Panel (a): distribution of the ratio of empirical cell standard deviation to the model's common σ. Panel (b): scatter of cell standard deviation against cell mean log wage, showing modest heteroscedasticity (lower-wage cells have higher dispersion). 146

A2.6 Distribution of maximum posterior probabilities across 18.5 million workers. With $X = 3$ latent types, the uniform prior assigns $p = 0.33$; values near 1.0 indicate near-certain classification. 146

A2.7 Additional simulated vs. actual transition shares: (a) firm-type destination conditional on origin (y, k); (b) occupation destination conditional on origin (y, k); (c) non-employment to employment destination distribution across (y, k) pairs. 147

A2.8 On original dataset. 160

A2.9 On uniform simulated dataset. 160

A2.10 On Zipf simulated dataset. 160

A3.1 Google Scholar citations of the Abowd et al. (1999) article 163

List of Tables

1.1 Descriptive Statistics by Employment Type and Sector 19

1.2 Transition Probabilities (Aggregate, Men, Women) 21

1.3 Gender Differences in Compensated Hours by Sector and Employment Type 22

1.4 Key Regression Results (Log Wages) 24

1.5 Share of individuals by Sex, Education, Transition Class and Income Class 30

2.1 Summary Statistics by Occupational Category 48

2.2 Worker Type Profiles 64

2.3 Firm Type Profiles (ordered by mean log wage) 65

2.4 Occupation Profiles 67

2.5 Wage Variance Shapley Decomposition 67

2.6 Additive Wage Variance Decomposition 68

2.7 Additive Wage Decomposition: Effect of Adding Occupation 70

2.8 Wage Mutual Information Decomposition (% of $H(\tilde{w}) = 4.322$ bits) 71

2.9 NPV Variance Shapley Decomposition 72

2.10 Additive NPV Variance Decomposition 72

2.11 NPV Mutual Information Decomposition (% of $H(\text{NPV}) = 4.280$ bits) 73

2.12 Pairwise Sorting Mutual Information by Age Bin (bits) 75

2.13 Decomposition by Age Bin (% of subgroup variance or entropy) 76

2.14 Within-y Mobility Among J2J Movers 78

2.15 Destination Choice and NPV Regression Results 80

2.16 Job to job Flow Decomposition: NPV vs Wage Quadrants 82

2.17 Share of J2J Moves Uphill by NPV, by Worker Type and Gender 82

3.1 Variance-share distortions from true shares at 5% mobility (Zipf firm sizes, size-weighted mobility) 108

A1.2 Classification of French Degrees and U.S. Equivalents 115

A1.3 Distribution of Education Levels by Sector, Sex and Age Groups at first observed spell 117

A1.4 OLS regressions on real net wages - alternative specification 118

A1.5 Demographic and Educational Composition by Transition Class (km) 123

A1.6 Transition Probabilities by Latent Transition Class (km) – Observed Data 124
A1.7 Summary of Latent Classes 125
A1.8 Predicted Transition Probabilities (Aggregate, Men, Women) 126
A1.9 Transition Probabilities (Aggregate, Men, Women) 127
A1.10 Parameters of unobserved mobility heterogeneity (mlogit models) 129
A1.11 Parameters of unobserved income heterogeneity (mlogit models) 129
A1.12 Parameters of state mobility (mlogit models) 130
A1.13 Parameters of cross-sectional income means and standard deviations 131
A1.14 Parameters of income mobility 131

A2.15 4-Factor Decomposition: Age Bin 1 (ages 20–27) 139
A2.16 4-Factor Decomposition: Age Bin 2 (ages 28–34) 139
A2.17 4-Factor Decomposition: Age Bin 3 (ages 35–41) 139
A2.18 4-Factor Decomposition: Age Bin 4 (ages 42–49) 140
A2.19 4-Factor Decomposition: Age Bin 5 (ages 50–60) 140
A2.20 Pairwise Sorting MI by Age Bin: All Six Factor Pairs (bits) 140
A2.21 Additive NPV Variance Decomposition 141
A2.22 Wage vs. NPV: Summary Comparison Across Decomposition Methods 141
A2.23 Heterogeneity and Non-Employment 142
A2.24 Structural Log-Share (Conditional Logit) 142
A2.25 Career-level comparison: data vs. simulation. 144
A2.26 Move-type composition: data vs. simulation. 145
A2.27 Daily separation rates by age bin. 145
A2.28 Within-cell standard deviation diagnostics. 146
A2.29 Worker classification by posterior entropy. 147
A2.30 Firm classification accuracy by size (full likelihood). 147
A2.31 Simulation vs. data: transition correlations. 148
A2.32 AKM variance decomposition. 150
A2.33 Projections of AKM effects onto CEM types. 151
A2.34 Mover event study: mean wage change by origin and destination $\hat{\psi}$-quartile. 153
A2.35 Scalar parameter recovery. 155
A2.36 Monte Carlo validation: four-factor additive variance decomposition. 156
A2.37 Monte Carlo validation: Shapley mutual information decomposition. 157
A2.38 AKM variance decomposition: original vs simulated data. 158

A3.1 Variance-share distortions at 10% mobility (Zipf firm sizes, size-weighted mobility) 168

A3.2 Variance-share distortions at 20% mobility (Zipf firm sizes, size-weighted mobility) 169
A3.3 Variance-share distortions at 40% mobility (Zipf firm sizes, size-weighted mobility) 170
A3.4 Variance-share distortions at 5% mobility (Uniform firm sizes, size-weighted mobility) 171
A3.5 Variance-share distortions at 10% mobility (Uniform firm sizes, size-weighted mobility) 172
A3.6 Variance-share distortions at 20% mobility (Uniform firm sizes, size-weighted mobility) 173

Introduction

Wage inequality remains one of the most persistent and extensively studied phenomena in labor economics. Disentangling the sources of wage dispersion requires understanding a complex interplay of institutional frameworks, firm-level wage policies, worker characteristics, and the assortative matching between different sides of the labor market. Historically, the literature has approached wage differentials through the lens of human capital theory, attributing variation primarily to observable traits such as education, experience, and skill (Mincer, 1974; Becker, 1964). However, empirical evidence consistently demonstrates that observable characteristics explain only a fraction of total wage variance, leaving a substantial residual gap driven by unobserved heterogeneity and institutional structures.

This dissertation investigates the multifaceted nature of wage inequality by tracing an analytical path from institutional labor market boundaries to the structural econometrics of worker-firm-occupation networks. The research presented here unfolds as an intellectual progression, moving from a specific empirical puzzle regarding sectoral wage differentials to a statistical model of multidimensional sorting, and then to foundational questions about the econometric tools most widely used to decompose wage variance. By evaluating the public-private pay gap, estimating a flexible multidimensional model of lifetime earnings, and subsequently diagnosing the reliability of canonical variance decomposition estimators, this thesis contributes to answering how labor market sorting operates and how it is measured.

Institutional Boundaries: The Public-Private Pay Gap

The initial inquiry of this dissertation examines how broad institutional boundaries shape long-term economic outcomes. The division between the public and private sectors serves as a labor market duality, particularly in coordinated market economies. The first chapter, *Evaluating the Public Pay Gap by Gender*, addresses an empirical puzzle in the public-private wage literature: the persistence of contradictory conclusions regarding whether public employment yields a wage premium or a penalty. While cross-sectional analyses of hourly wages frequently suggest penalties for highly educated workers, other methodologies hint at institutional premiums, creating a fragmented understanding of sectoral compensation.

Resolving this puzzle requires shifting the analytical focus from static, cross-sectional hourly wages to dynamic, lifetime earnings. Utilizing rich French administrative panel data spanning from 2012 to 2019, the first chapter establishes a set of undocumented yet intuitive stylized facts concerning wage dynamics, sectoral mobility, and gender disparities. The core finding reveals that modest hourly

wage gaps systematically conceal substantial disparities in lifetime earnings and employment stability. Evaluating compensation over a career horizon demonstrates that lifetime earnings premia differ radically based on both observable and unobserved characteristics.

Crucially, the analysis identifies distinct winners and losers within the public-private sector framework. Women and low-educated workers gain a significant lifetime earnings advantage in the public sector. This premium is driven not merely by base pay, but by institutional features that foster higher retention, better-compensated part-time work, and more equitable annual hours compared to the private sector. In the private sector, gender gaps remain starkly wider, particularly for individuals possessing higher levels of education. Conversely, highly educated men experience a pronounced lifetime earnings penalty in public employment, a direct consequence of rigid, centralized wage structures that compress the upper tails of the wage distribution and stifle the returns to advanced human capital.

By dynamically modeling sectoral transitions, movements into and out of employment, and earnings heterogeneity using an Expectation-Maximization algorithm approach inspired by Postel-Vinay and Turon (2007), the first chapter demonstrates that both premiums and penalties depend systematically on gender, education, and accumulated labor market experience. The analysis reveals that significant unobserved heterogeneity remains in wage dynamics, even after controlling for a rich set of observables. This chapter unifies previously conflicting narratives by providing a comprehensive account of sectoral differences, highlighting the dual role of public sectors: mitigating gender inequality through employment stability while simultaneously imposing ceilings on wage progression for certain groups.

The presence of profound unobserved heterogeneity driving lifetime earnings differentials naturally invites a broader question. If unobserved characteristics dictate such divergent outcomes across sectors, what are the deeper structural dimensions—worker quality, firm wage policies, and the nature of the jobs themselves—along which sorting operates in the labor market as a whole?

Multidimensional Sorting and Lifetime Value

Answering this question requires a framework that captures the full dimensionality of the labor market. Since the seminal contribution of Abowd et al. (1999), the literature studying how individual wages vary across workers and firms has centered on additive fixed-effects models that decompose wage variance into worker and firm components. This approach, extended by bias-correction methods (Andrews et al., 2008; Kline et al., 2020), has been enormously productive for understanding the bipartite structure of worker-firm wage determination. Yet by construction, additive worker-firm models treat the labor market as a two-sided matching problem. In practice, a job is not merely a worker-firm pair: it is a triplet comprising a worker, a firm, and an occupation. Workers change occupations—both within and across firms—and these transitions carry rich information about career trajectories and lifetime value. In bipartite frameworks, this occupational dimension is necessarily absorbed into worker effects, firm effects, or the residual, rather than being isolated as a distinct source of sorting.

The second chapter, *Job Mobility and Career Values*, co-authored with Jean-Marc Robin, addresses this gap by jointly estimating three-way sorting among workers, firms, and occupations. To achieve this without imposing any behavioral assumptions characteristic of structural search models, the chapter implements a flexible, reduced-form Classification Expectation-Maximization (CEM) algorithm (Lentz et al., 2023) on French administrative data. By defining a job as a triplet comprising a worker type, a firm type, and an occupation, the analysis recovers three discrete latent worker types and twelve latent firm types. This discretization inherently bypasses the incidental parameter problem associated with estimating millions of fixed effects, while allowing for the nonlinear interactions between worker quality, firm wage policy, and occupational choice that additive models rule out.

Building on the finite mixture approach of Bonhomme et al. (2019) and the structural mobility model of Lentz et al. (2023), the CEM algorithm classifies firms using the full information of both wage distributions and mobility patterns, rather than wages alone. Occupational heterogeneity enters through eight occupation categories based on French PCS (Professions et Catégories Socioprofessionnelles) codes, interacted with age and gender, allowing the model to capture rich variation in how different types of workers sort into different types of jobs at different stages of their careers. Furthermore, the chapter introduces Shapley value decompositions to the wage variance literature. This methodological innovation allows for an exact, order-independent attribution of the contribution of each sorting dimension, providing a principled solution to the correlation problem that arises whenever worker types, firm types, and occupations are jointly determined.

Three profound findings emerge from this multidimensional framework. First, the analysis identifies occupation as the single largest contributor to both wage variation and lifetime-value variation—a dimension largely absorbed or obscured in traditional worker-firm models. The full interaction of occupations, gender, and the latent worker and firm types explains nearly 64% of age-adjusted log-wage variance. Within this explained variance, occupation contributes the largest share (26.6%), followed closely by firm types (23.0%), while latent worker types (7.8%) and gender (6.3%) provide smaller marginal contributions.

Second, the chapter extends the analysis from current wages to lifetime value by computing the net present value (NPV) of each job using exact backward induction on the estimated Markov chain of worker transitions. The findings reveal a striking divergence: while the worker-firm-occupation triplet explains 71% of NPV entropy, it explains only 17% of age-adjusted wage entropy. This indicates that NPV effectively integrates over a forward-looking transition structure governed heavily by latent types, capturing a dimension of career progression invisible in cross-sectional wage variance.

Finally, the chapter evaluates the drivers of worker mobility. The analysis demonstrates that workers' destination choices respond strongly to the capitalized current-wage component of a job's value, but exhibit almost no response to the option-value component. Transitions are overwhelmingly wage-augmenting, providing compelling evidence that the career-building or "stepping stone" value of an employer change is significantly less important to worker mobility than immediate wage gains. These

findings highlight the immense informational content of occupational heterogeneity and redefine the understanding of why workers sort across the labor market.

The Reliability of Additive Variance Decomposition

Having established that a multidimensional, discrete-type approach reveals occupation as the dominant contributor to wage variance and uncovers nonlinear interactions that additive models cannot capture, a natural complementary question arises: under what conditions do additive variance decompositions—the workhorse of the matched employer-employee literature—reliably recover their target parameters? The Abowd et al. (1999) framework provides an elegant and parsimonious decomposition of log-wage variance into worker effects, firm effects, and a sorting covariance. For the bipartite question it was designed to answer—how much of wage variance is attributable to workers versus firms—it remains the canonical tool. Yet as research leveraging administrative datasets has expanded, mounting evidence suggests that the variance shares produced by AKM estimators are sensitive to sample structure, particularly regarding the degree of worker mobility across firms (Andrews et al., 2008; Kline et al., 2020).

The third chapter, *Signals, Noise, and Sorting*, directly engages with this question. This chapter investigates the structural conditions under which additive variance decomposition is reliable. While the existing literature identifies limited mobility as a source of incidental parameter bias, the chapter reconceptualizes this issue through the lens of spectral graph theory. The analysis demonstrates that the bias inherent in plug-in estimators is fundamentally driven by the spectral properties of the graph Laplacian, which characterizes the firm-firm network of worker mobility.

A critical innovation of this chapter is identifying the average node degree ($\bar{d}$) as a primary, easy-to-compute diagnostic for estimation fragility. The analysis reveals that distortions in sorting covariance and the artificial inflation of worker variance shares are not merely artifacts of limited sample sizes. Rather, they arise from the interaction of idiosyncratic noise with the weighting structure of the bipartite graph. When the average node degree is low—a mathematical symptom of low average firm size and sparse worker transitions—the Laplacian's spectrum collapses toward zero. Under these conditions, idiosyncratic noise destroys the positive sorting signal, heavily distorting all plug-in variance components.

Through extensive Monte Carlo simulations, the chapter evaluates the performance of state-of-the-art bias correction methods, including leave-out estimators and split-sample techniques. The results indicate that while these corrections successfully reduce limited mobility bias, they remain highly susceptible to noise-amplification when the network's spectral density is low. This phenomenon systematically attenuates evidence of positive assortative sorting, casting doubt on findings from sparse networks. The split-sample estimator emerges as the most robust alternative, as it uniquely breaks the spectral noise-amplification channel while removing bias through subtraction-based recovery.

Ultimately, the third chapter establishes that the reliability of additive variance decomposition is fundamentally constrained by the structural connectivity of the underlying mobility network. By identifying $\bar{d}$ as a sufficient statistic for the severity of spectral collapse, this research provides a diagnostic framework that allows researchers using AKM-type models to assess whether their network supports credible inference. The contribution is thus to the AKM literature itself: strengthening the conditions under which its decomposition can be trusted.

Progression and Contributions

Read sequentially, the chapters of this dissertation capture an intellectual progression. The journey begins with an empirical question: how do public and private sector lifetime earnings differ? The discovery that these differences are highly heterogeneous and dependent on unobserved worker quality motivates a deeper investigation into the structural dimensions of wage variance.

This leads directly to the second chapter's multidimensional analysis. By estimating a flexible CEM model with discrete latent types, the chapter maps the three-way sorting of workers, firms, and occupations. Reintroducing occupation as a central pillar of wage determination—and expanding the horizon from cross-sectional variance to lifetime net present value—reveals that the labor market is richer than a bipartite worker-firm world suggests. Occupational mobility, nonlinear interactions between worker and firm quality, and the distinction between current wages and forward-looking career values all emerge as first-order features of inequality.

The third chapter then turns to the econometric foundations of the existing workhorse. Understanding the spectral properties of the graph Laplacian explains why the literature has struggled to reach consensus on the degree of assortative matching: in sparse networks, the very structure that researchers attempt to analyze can undermine the estimators used to do so. By providing a diagnostic framework based on network connectivity, this chapter strengthens the toolkit available for additive variance decomposition while clarifying its boundaries.

While each chapter stands as an independent contribution to the literature, together they form a comprehensive narrative. They demonstrate that understanding wage inequality requires attending to institutions and the multidimensional reality of worker choices and firm policies, and rigorously evaluating the econometric tools we use to measure them. Through this progression, this dissertation contributes to the ongoing effort to map, measure, and ultimately understand the distribution of earnings.

Bibliography

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

Martyn J Andrews, Leonard Gill, Thorsten Schank, and Richard Upward. High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A*, 171(3):673–697, 2008.

Gary S. Becker. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. Number beck-5 in NBER Books. National Bureau of Economic Research, Inc, none edition, August 1964. doi: None. URL https://ideas.repec.org/b/nbr/nberbk/beck-5.html.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer-employee data. *Econometrica*, 83(3):699–739, 2019.

Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898, September 2020.

Rasmus Lentz, Suphanit Piyapromdee, and Jean-Marc Robin. The anatomy of sorting—evidence from danish data. *Econometrica*, 91(6):2409–2455, 2023. doi: https://doi.org/10.3982/ECTA16425.

Jacob A. Mincer. The human capital earnings function. In *Schooling, Experience, and Earnings*, NBER Chapters, pages 83–96. National Bureau of Economic Research, Inc, August 1974. URL https://ideas.repec.org/h/nbr/nberch/1767.html.

Fabien Postel-Vinay and Hélène Turon. The public pay gap in britain: Small differences that (don't?) matter. *Economic Journal*, 117(523):F422–F442, 2007. doi: 10.1111/j.1468-0297.2007.02067.x.

CHAPTER 1

Evaluating the Public Pay Gap by Gender

A Comparison of Public and Private Sector Wages in France

Abstract

This paper resolves an empirical puzzle in the public-private wage literature: why studies using similar data reach contradictory conclusions about wage premiums and penalties. Utilizing rich French administrative panel data (2012-2019), this study has two main contributions: first, it presents a set of new, intuitive yet previously undocumented stylized facts about wage dynamics, sectoral mobility, and gender differences across sectors. Second, the results reveal that the modest hourly wage gaps conceal substantial disparities in lifetime earnings and employment stability. Women, in particular, gain a significant lifetime earnings advantage in the public sector, driven by higher retention, better-compensated part-time work, and more equitable annual hours compared to the private sector, where gender gaps remain larger, especially for those with higher education. In contrast, highly educated men experience a lifetime penalty in public employment due to rigid wage structures. By dynamically modeling sectoral transitions, transitions into and out of employment, and earnings heterogeneity using an Expectation-Maximization algorithm, this study shows that both premiums and penalties depend systematically on gender, education, and labor market experience. The analysis reveals that significant unobserved heterogeneity remains in wage dynamics. These findings unify prevailing narratives by providing a comprehensive, descriptive account of sectoral differences in transitions, part-time work and wages by gender.

Why do some studies find wage penalties in the public sector while others report premiums? The literature on public-private wage differentials poses an empirical puzzle: even when using similar data, researchers often reach contradictory conclusions, with some documenting sizable penalties and others finding substantial premiums. This empirical puzzle stems from valuable but distinct methodological approaches and sample selection criteria to studying public-private wage differentials. Most studies fall into two distinct camps: those using fixed-effects regressions on hourly wages[1] and those employing dynamic lifetime approaches on monthly or annual wages[2]. However, neither approach simultaneously incorporates the key elements that explain contradictory empirical findings. At its heart, the public-private wage gap question is: "For a given person, what is the effect of working in the public sector on their wage, compared to if they had worked in the private sector?". The epistemological problem is that we can never observe this counterfactual. We cannot see the same person in both sectors at the same time. Therefore, any statistical method is an attempt to construct a *plausible* counterfactual group to compare against.

[1] (Bargain and Melly, 2008) and more recent work (Bargain and Melly, 2008)
[2] (Beffy and Kamionka, 2010; Dickson et al., 2014)

The choice of method dictates the nature of that counterfactual and thus defines what we are actually claiming to know. Fixed effects regressions estimate the wage gap for people who switch sectors, averaging over switchers their wages before and after the sectoral change while controlling for innate, time-invariant traits. Lifetime earnings approaches with selection models estimate the expected gap for the entire population, using statistical techniques to simulate what a typical person would earn in either sector by accounting for job loss risk and the preexisting differences that lead people to choose one sector over the other – this paper takes inspiration from the lifetime earnings literature, in particular the approach first estimated in the UK by Postel-Vinay and Turon (2007) and across European countries in Dickson et al. (2014).

The answer to this empirical conundrum thus lies in the methodology used. Fixed-effects studies on hourly wages, while controlling for unobserved heterogeneity, focus on hourly wages and miss the differential impact of hours worked—particularly relevant given women's higher propensity for part-time work. More importantly, fixed effects wage regressions conceal a key element: if we take into account the pre-trend of annual wages for sector switchers, we find an Ashenfelter dip in annual earnings prior to switching — as opposed to a near-parallel pre-trend in hourly wages for sector switchers compared to stayers[3]. Meanwhile, lifetime approaches account for job security through employment transitions but have either been limited to male-only samples, or have not allowed for free movement between the sectors. The former method is used solely in section 1.2. The latter is adopted for the main results because it delivers superior in- and out-of-sample performance, accommodates job stability and annual wage variation arising from heterogeneous labor supply, and enables joint statistical modeling of sectoral stayers and movers.

This study makes two key contributions. *First*, the study builds on the current literature by reconciling conflicting empirical evidence and simultaneously examining gender dynamics, unobserved heterogeneity[4], unrestricted sectoral mobility, and lifetime earnings. Women disproportionately sort into stable public roles due to both observable traits like education and latent factors shaping career trajectories[5]. *Second*, I demonstrate that seemingly modest hourly wage gaps between public and private sectors mask profound disparities in lifetime earnings, with the public sector exhibiting gender gaps 13% lower than

[3]More in subsection 1.2.1.

[4]A critical issue in estimating the public-private wage gap is selection bias, which arises because public sector jobs often attract individuals with different preferences, skills, or motivations compared to those in the private sector. (Heckman and Singer, 1984) famously introduced the selection correction technique, which allows researchers to account for such unobserved differences and has been extended and applied by (Beffy and Kamionka, 2010) in looking at the public pay gap in France. For instance, public sector workers may have a stronger preference for job security or non-wage benefits like pensions, which can complicate comparisons with private sector employees who may prioritize higher pay or career mobility. Studies that fail to account for this selection bias risk over- or under-estimating the true wage differential.
Based on the above analysis, one could argue that there is self-selection of women into the public sector; that is, women prefer (or "select" themselves) to be in the public sector wherever possible. Therefore, comparing these jobs held by women in the public sector to private sector employment will mechanically give a positive public premium as agents are rational. On looking more closely, one realizes that *a priori* there is no reason to assume that only women self-select into the public sector. With rational agents in society, we expect that any individual is selecting the sector most suitable for them. Hence, the same argument can be made for men's self-selection to the public sector, and to that effect we should find similar public premia by gender and earnings percentiles, which is not the case in this setting as it is higher for women extensively and intensively. This is not a complete measure of self-selection but of the different outcomes women have compared to men.

[5]This is echoed in (Fougère and Pouget, 2003; Garibaldi et al., 2021; Gomes and Kuehn, 2025)

the private sector on average across earnings quantiles through better retention and compensation for part-time work—a mechanism overlooked in prior hourly wage comparisons.

By utilizing a sequential Expectation-Maximization (EM) algorithm on French panel data that include men, women, full-time and part-time work, and allows for sector-switching, I find that public sector employment offers a significant lifetime earnings premium for women but imposes penalties for highly educated men. The methodology simultaneously estimates individual income trajectories, employment dynamics, and selection into sectors, allowing for unobserved heterogeneity in earnings patterns and sector choice. Considering lifetime values accounts for not just wage levels but also the security of income over an entire career, particularly in the face of job loss risk. (Garbinti et al., 2023) highlight the importance of incorporating lifetime earnings into gender pay gap analyses, showing that short-term wage differences underestimate long-term disparities. (Guvenen et al., 2021) further argue that income volatility and employment uncertainty significantly shape utility-based job preferences, particularly for risk-averse individuals who may select into the public sector for its greater stability.

This work unifies two main strands of literature on France. While (Bargain et al., 2018; Bargain and Melly, 2008) report hourly wage penalties for educated workers, (Beffy and Kamionka, 2010; Dickson et al., 2014) find lifetime premia using monthly earnings and accounting for selection. This paper shows that hourly wages mask profound lifetime disparities driven by part-time work and unobserved selection. France's institutional features—centralized wage bargaining, regulated hours, and strong job protection—position it as a prototype of coordinated market economies where similar dynamics operate.

The public sector's role as a mitigator of gender inequality is increasingly relevant as fiscal pressures and privatization debates intensify. The gender pay gap is a well-documented phenomenon, consistently showing that women earn less than men across various sectors and countries (for example, (Blau and Kahn, 2017; Goldin, 2014; Petersen and Morgan, 1995), etc.). In high-income countries, the compatibility of women's career and family goals, facilitated by family policy, cooperative fathers, favorable social norms, and flexible labor markets, has become key drivers of fertility and participation in the labor force (Bertrand et al., 2010). (Goldin, 2014) argues that while gender convergence in labor market outcomes has progressed significantly, workplace flexibility remains a key obstacle, a factor that is particularly relevant in explaining the persistent gender wage gap between public and private sectors - similarly, this study echoes the disparities in compensated hours by gender. Extensive research has explored multiple dimensions of this disparity, including differences in occupation, industry, work experience, education, and discrimination. However, a crucial aspect that warrants further attention is the variation in gender pay gaps between the private and public sectors over a lifetime. This article contributes to the existing body of literature by demonstrating that the private sector imposes a larger gender pay gap when compared to the public sector when considering lifetime earnings. Prior work highlights shrinking gender wage gaps in countries with large public sectors (Ponthieux and Meurs, 2015), but few studies examine lifetime earnings or part-time work comprehensively.

This paper resolves existing contradictions by modeling annual earnings and incorporating part-time employment—critical for women, who constitute 75% of part-time workers in France.

The analysis reveals a critical equity-efficiency trade-off: while the public sector insulates workers from volatility, it imposes lifetime earnings penalties on highly educated workers due to rigid wage structures. The core idea of this study is to provide a nuanced understanding of public-private earnings differences; that penalties or premiums are not universal features of either sector.

This article proceeds as follows: Section 1.1 describes the data and briefly, some institutional context. Section 1.2 provides descriptive evidence and some stylized facts that support the statiscal model that is later specified in section 1.3. Section 1.4 presents the unobserved heterogeneity analysis along with model fit both in- and out-of-sample. Section 1.5 documents the main results. After a discussion on points for future research in section 1.6, section 1.7 concludes.

1.1 Data

1.1.1 Data description

This study uses the DADS-EDP tous salariés panel for 2012-2019, combining the DADS "all-employees" panel with the Échantillon Démographique Permanent (EDP). The DADS panel provides comprehensive employee and job information from annual employer declarations and state payroll files. The EDP adds demographic data from civil status records and surveys for individuals born on October 1st and 4th. The datasets are matched via registration index numbers, allowing education levels to be determined from survey responses.

A key development in the data is the integration of DSN (Déclaration Sociale Nominative) information starting in 2016, which streamlined social declarations for French companies [6]. By 2017, the DSN became mandatory for most private sector companies, and by 2018, 99% of private companies had adopted it. This transition expanded the dataset's scope to include both private and public sector employees, as well as employees of individual employers, enabling more comprehensive analysis of employment and wages across the French economy.

1.1.2 Sample selection and statistics

The panel used in this analysis consists of 276,514 individuals (training data) which is a randomly selected quarter of the larger sample consisting of 829,232 individuals (used to check for out-of-sample model fit). The latter is a random sample that is representative of the French working population. The sample size is purposely kept large to better inform the dynamic model of the heterogeneity in mobility and income over time. The total number of panel observations including spells of non-employment total 2,065,902. Individuals aged 18-60 who have completed their education prior to or during the

[6] https://www.insee.fr/en/information/4195367?sommaire=4195376

observed time-periods are kept to avoid peculiarities (apprenticeship for young workers, and early retirement in the public sector, for example). Employment spells that total atleast 6 months or 180 days in a year are used for the analysis [7]. The log of real net annual wages or "salaire net fiscal en euros constants" is used throughout. Using real net wages provides a more accurate representation of workers' purchasing power by accounting for all social contributions and eliminating inflation effects. This measure offers a standardized basis for comparison across sectors and time periods, reflecting the real economic situation of workers more effectively than gross salary figures.

Education is coded as a categorical variable with the three levels: high (some university qualification), medium (some vocational training post high school), and low (at most a high school degree). Table A1.2 tabulates this classification and and provides U.S. equivalents. In the sample, approximately 69% of individuals in the panel had a low education level, 16% medium, and 15% high.

Table 1.1. Descriptive Statistics by Employment Type and Sector

		Full-time spells (79%)		Part-time spells (21%)	
		Public	Private	Public	Private
Employment share (%)		24	76	26	74
Log real net annual wage	Mean	10.19	10.14	9.58	9.32
	(sd)	(0.35)	(0.49)	(0.69)	(0.87)
Log real net hourly wage	Mean	2.69	2.65	2.56	2.42
	(sd)	(0.32)	(0.42)	(0.32)	(0.41)
Demographics					
Share of Women (%)		61	36	84	77
Low Education (%)		52	68	56	74
Med Education (%)		21	17	23	14
High Education (%)		28	15	21	12
Ages ≤ 30 (%)		11	18	11	16
Ages 31-45 (%)		41	43	44	39
Ages 45+ (%)		49	39	44	45
# panel observations		1,142,309		299,721	
# individuals		276,514			

Table 1.1 presents descriptive statistics for full-time and part-time employment spells, broken down by public and private sectors. Full-time employment accounts for 79% of spells, with a larger share in the private sector (76%) compared to the public sector (24%). Part-time employment constitutes 21% of spells, with similar sectoral proportions (26% public, 74% private). The mean log real net annual wage is slightly higher in the public sector than in private. This pattern persists for part-time workers, with mean log annual wages of 9.58 in the public sector and 9.32 in the private sector. Similarly, hourly wages are marginally higher in the public sector for both full-time and part-time workers.

Demographic differences are notable. Women dominate public sector employment (61% for full-time and 84% for part-time) compared to the private sector (36% and 77%, respectively). Public sector employees also exhibit higher levels of education, with 28% of full-time and 21% of part-time workers

[7] This construction includes employment spells across multiple employers within a sector, allowing for a more flexible approach than typical full-year, single-employer restrictions. If total employment in a year is less than 6 months, that year is coded as a non-employment spell. See Appendix A1.1 for further details on panel construction.

having high education, versus 15% and 12% in the private sector. Private sector workers tend to be younger, with 18% under 30 in full-time roles, compared to 11% in the public sector.

Women dominate part-time employment, accounting for 79% of part-time spells across both public and private sectors, which represents 17% of aggregate employment during the observed period. Among women's employment spells, part-time work varies inversely with education level: 37% for low-educated, 28% for medium-educated, and 24% for highly educated women. This pattern suggests an inverse relationship between educational attainment and the propensity for part-time work among women.

1.2 Stylized facts

In these sub-sections, I present some novel evidence from the panel that motivates the estimation strategy presented in the next section. These nuances are key to understanding public-private differences and are seldom consolidated in a single narrative. The findings are intuitive, but are nonetheless important to quantify.

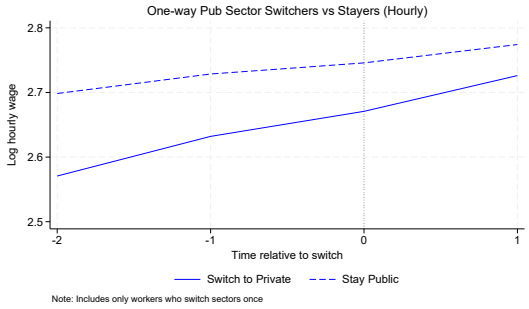
1.2.1 Sector-switching is rare...but not random

Table 1.2 above highlights how rare sector-switching is, particularly when considering full-time employment. Year-on-year a mere 1.2% switch sectors in full-time jobs. In prior works on the subject leveraging fixed-effects, identification relies on the assumption that sector switching is random - more specifically, that changes in wage are orthogonal to switching sectors. This assumption appears plausible when examining hourly wage changes, as shown in figures 1.1a and 1.1b , where the wage distribution across sectors shows limited patterns of systematic variation among movers. The pre-trend of stayers and movers is similar. However, when analyzing annual wage changes, a different narrative emerges. Figures 1.1c and 1.1d suggest that sector switching is systematically associated with an increase in earnings in both directions (although a smaller increase when transitioning into the public sector). Here, we can observe a strong Ashenfelter dip in earnings for switchers. This pre-trend of falling wages for switchers whose earnings then increase on switching sectors suggests that sector-switching is not orthogonal to wages. This discrepancy between hourly and annual wages gives reason to revisit the assumption of randomness, as movers may be selecting sectors based on broader income considerations than hourly wage rates alone.

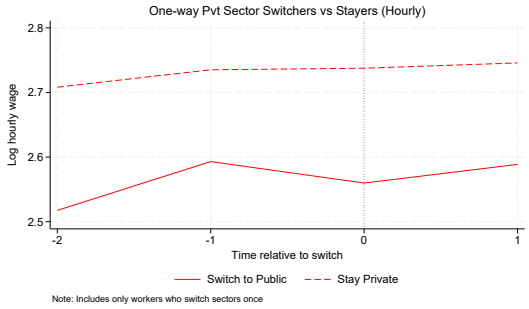
1.2.2 Women and men move differently in the labor market

Table 1.2 presents transition probabilities across employment states, comparing non-employment, private, and public sector (full-time and part-time) employment segmented by gender[8]. Key findings are: first, the higher persistence in employment status within both public and private sectors, as

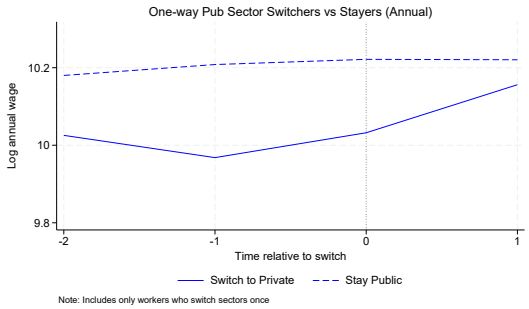
[8]Note that these figures imply an inactivity rate of about one-third (referred to as non-employment in this study), and private:public sector employment ratio of approximately 75:25 as share of all employment. These figures are in line with national statistics published here and here.



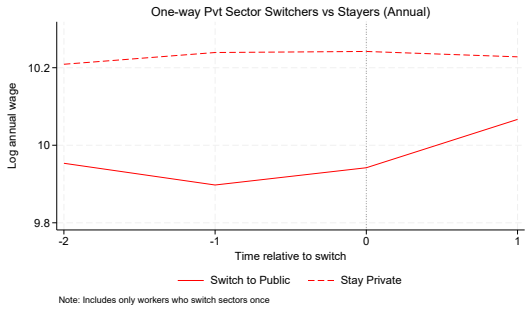
(a) Hourly Wage Changes - Public Sector Switchers and Stayers



(b) Hourly Wage Changes - Private Sector Switchers and Stayers



(c) Annual Wage Changes - Public Sector Switchers and Stayers



(d) Annual Wage Changes - Private Sector Switchers and Stayers

Figure 1.1. Distributions of Hourly and Annual Wage Changes

evidenced by high probabilities for remaining in the same employment type (e.g., 0.89 for private full-time and 0.93 for public full-time).

Table 1.2. Transition Probabilities (Aggregate, Men, Women)

From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
			Aggregate		
Non-employment (NE)	0.82	0.11	0.02	0.04	0.01
Private full-time (Pvt FT)	0.08	0.89	0.002	0.03	0.0004
Public full-time (Pub FT)	0.02	0.006	0.93	0.0015	0.04
Private part-time (Pvt PT)	0.17	0.12	0.004	0.7	0.005
Public part-time (Pub PT)	0.10	0.01	0.13	0.01	0.75
Total	*0.31*	*0.42*	*0.13*	*0.10*	*0.04*

	Men					**Women**				
From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
NE	0.83	0.13	0.01	0.02	0.005	0.81	0.08	0.02	0.06	0.02
Pvt FT	0.08	0.90	0.002	**0.02**	0.0003	0.09	0.86	0.003	**0.05**	0.001
Pub FT	0.02	0.006	0.95	0.001	**0.02**	0.03	0.006	0.92	0.002	**0.05**
Pvt PT	0.22	**0.22**	0.004	**0.55**	0.004	0.15	**0.09**	0.004	**0.74**	0.01
Pub PT	0.15	0.02	0.15	0.01	**0.67**	0.09	0.01	0.13	0.01	**0.76**
Total	*0.31*	*0.53*	*0.10*	*0.05*	*0.01*	*0.31*	*0.31*	*0.16*	*0.16*	*0.06*

Notably, men exhibit greater stability in private full-time roles (0.90) compared to women (0.86). Second, transition rates into part-time roles show a gender disparity: women are more than twice as likely as men to move into part-time jobs (0.05 vs. 0.02) from full-time positions in both sectors.

They also **remain in part-time roles more often**, with less than half the probability of transitioning to private full-time jobs compared to men. Men are more likely to move from non-employment to private full-time roles, underscoring gender disparities in workforce re-entry.

1.2.2.1 The Part-Time Paradox: Women's Dominance in Flexible Work

A significant advantage of using the DADS data is that it allows us to observe the hours workers are compensated for, including paid time off. In other data sources, respondents are asked how many hours they work on average, or in that week, like in the Labor Force survey. In a study like this, where we want to be able to predict transitions and wages over a lifetime, having accurate and administrative records is crucial to avoid recall bias and measurement errors that can arise from self-reported data. This precision ensures a more reliable estimation of lifetime earnings trajectories and sectoral transitions, which are key to understanding long-term labor market dynamics.

Women's labor supply significantly differs from men's across public and private employment sectors, as evidenced by disparities in compensated hours captured through administrative data. Table 1.3 highlights key differences in average annual compensated hours between men and women across employment types. In both full-time public and private sectors, men consistently have more compensated hours on average than women, with the gap being more pronounced in the private sector. For example, in the full-time private sector, men are compensated for 63.57 more hours annually than women (roughly translates to two weeks), while in the public sector, this difference narrows to 20.65 hours. These trends suggest that the public sector offers more equitable compensation structures for women compared to the private sector.

Table 1.3. Gender Differences in Compensated Hours by Sector and Employment Type

		Observations		Average Annual Hours		Difference
Type	Sector	Men	Women	Men	Women	
Full-time	Private	552,026	315,664	1,833.399	1,769.83	63.569***
	Public	107,911	165,376	1,817.262	1,796.615	20.647***
Part-time	Private	51,769	170,255	1,111.697	1,226.44	-114.743***
	Public	10,546	64,053	1,110.931	1,287.2	-176.27***

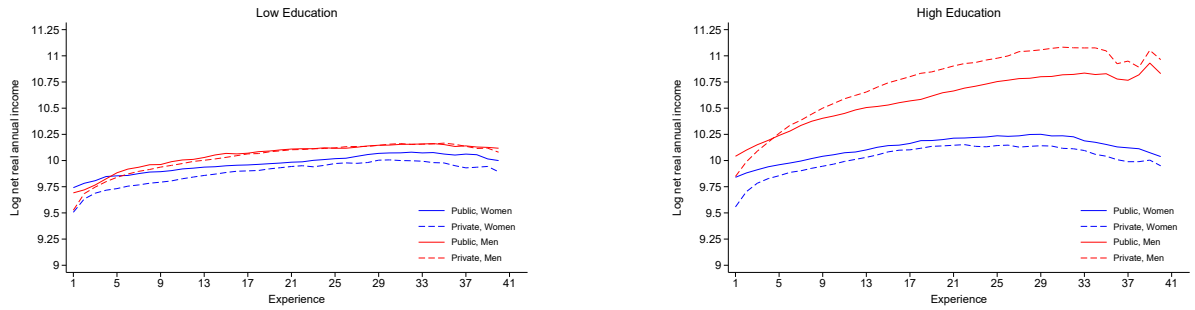
Note: The last column "Difference" is the difference in compensated hours between and women for each employment state and the *s denote the significance of t-tests (with uneqal variances) for these differences.

Women's labor supply patterns reveal a striking paradox across public and private employment sectors. Table 1.3 illustrates that while men consistently have more compensated hours on average in full-time roles, this trend reverses dramatically for part-time work. In the full-time private sector, men are compensated for 63.57 more hours annually than women. However, in part-time roles, women significantly outpace men in compensated hours, with the most pronounced difference in the public sector where women log 176.27 more hours annually than their male counterparts. This part-time paradox is further emphasized by the prevalence and persistence of part-time work among women: it accounts for 32% of women's employment compared to just 9% for men, and women are less likely to

transition out of these roles. These findings suggest that part-time work, particularly in the public sector, may be serving as a long-term strategy for women to balance work and other responsibilities, rather than a temporary state as it appears to be for men. Women's compensation within part-time employment is higher both in terms of annual and hourly wages in the public sector. A breakdown by education levels further reveals that the patterns in compensated hours persist across all categories (with higher educational attainment mitigating the gender gap in hours), with the public sector showing smaller differences in full-time compensated hours between men and women and larger differences in part-time compensated hours.

1.2.3 Differences in earnings are larger in the private sector, especially for educated workers

The analysis of wage differentials in France reveals persistent pay gaps between men and women that are more pronounced in the private sector, particularly for highly educated workers. Using OLS regressions on log wages, women earn 14.3% less in annual wages and 14.7% less in hourly wages compared to men, controlling for other factors (Table 1.4). However, this gap narrows in the public sector, as evidenced by the positive interaction between female and full-time public sector employment (5.4% for annual wages and 3.8% for hourly wages) extracted from an alternate specification[9]. Figure 1.2 which graphs mean wages by sector and sex alludes to these differences in annual pay across the sectors by gender. The trajectory of men in the private sector is, for the most part, indistinguishable from those in the public sector as there is a significant overlap. Conversely, for women, there is no overlap in the public and private sector earnings as the public sector trajectory lies squarely above that of the private sector.



(a) Low Education **(b)** High Education

Figure 1.2. Annual wage dispersion by sector, sex and sector, over experience

Part-time employment significantly impacts wages, with substantial penalties observed. Part-time private sector workers earn 71.7% less in annual wages and 15.7% less in hourly wages relative to full-time private sector workers. However, public part-time employment appears to offer better wage outcomes than private part-time employment. The panel fixed effects regression results presented in Table 1.4 show that while both types of part-time work are associated with lower annual wages compared to full-time private sector work, the penalty is smaller for public part-time work (-12.4% for women)

[9]Refer to table A1.4 in the appendix.

Table 1.4. Key Regression Results (Log Wages)

	OLS		Panel Fixed effects		Panel Fixed Effects by Gender			
Variable	Annual Wage	Hourly Wage	Annual Wage	Hourly Wage	Men Annual	Men Hourly	Women Annual	Women Hourly
Female	-0.143***	-0.147***						
	(0.001)	(0.001)						
Full-time Public	-0.025***	-0.047***	0.095***	0.015***	-0.005	-0.014***	0.156***	0.023***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.004)	(0.002)	(0.004)	(0.002)
Part-time Private	-0.717***	-0.157***	-0.378***	0.060***	-0.377***	0.099***	-0.373***	0.035***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
Part-time Public	-0.546***	-0.104***	-0.217***	0.061***	-0.469***	0.046***	-0.124***	0.064***
	(0.002)	(0.001)	(0.003)	(0.001)	(0.005)	(0.003)	(0.004)	(0.002)
Medium Education	0.300***	0.269***						
	(0.001)	(0.001)						
High Education	0.545***	0.531***						
	(0.001)	(0.001)						
Experience (per decade)	0.459***	0.308***	-0.595***	0.407***	0.637***	0.460***	0.546***	0.355***
	(0.002)	(0.001)	(0.003)	(0.001)	(0.003)	(0.002)	(0.004)	(0.002)
Experience Squared	-0.082***	-0.049***	-0.091***	-0.051***	-0.107***	-0.062***	-0.074***	-0.041***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Observations	1,441,964	1,437,538	1,441,964	1,437,538	725,048	722,229	716,916	715,309
Number of individuals	248,084	247,586	248,084	247,586	123,663	123,321	124,421	124,265
Year fixed effects	✓	✓	x	x	x	x	x	x
Rho			0.938	0.925	0.870	0.901	0.828	0.898

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

compared to private part-time work (-37.3% for women). The high rho indicates that most of the variance is due to individual fixed effects rather than idiosyncratic errors. Moreover, public part-time work is associated with modestly higher hourly wages (6.4% increase for women) compared to full-time private sector work, while private part-time work shows a smaller hourly wage premium (3.5% for women) - the slightly higher wage rates for part-time work is possibly a compensating differential.

Education plays a crucial role in wage determination, with higher education associated with substantial wage premia. Medium education increases wages by about 30%, while high education increases wages by approximately 54% compared to low education. However, the returns to education differ between sectors and genders. Experience contributes significantly to wage differentials, with wages increasing at a decreasing rate. The coefficient for experience (per decade) is 45.9% for annual wages and 30.8% for hourly wages, with negative squared terms indicating diminishing returns. While the public sector appears to offer a more equitable environment, especially for women and those with more experience, substantial gender pay gaps persist across both sectors, particularly for highly educated workers in the private sector. The advantage of public sector employment is especially pronounced for part-time workers, offering better wage outcomes compared to private sector part-time employment.

1.3 Model

The methodology in this study adapts the framework developed in a seminal paper on the public pay gap in Britain Postel-Vinay and Turon (2007), later extended in Dickson et al. (2014). The components directly borrowed from those papers are detailed in Appendix A1.3. For clarity and readability, the main likelihood equations (1.3.1 and 1.3.2) and the lifetime earnings equation (1.5) are retained in the main text, even though they closely mirror those in the earlier work. Likewise, the model structure outlined in Section 1.3.1 follows a similar naming convention but differs in several key components.

The main methodological adaptation in this paper lies in the specification of the income process. Whereas the earlier studies used an AR2 framework with restrictive sample selection, log earnings here are modeled as a more flexible AR1 process. This approach avoids sample selection restrictions, captures the persistence of log wages, and allows for the influence of both stochastic noise and covariate-driven variation.

The resulting dynamic model jointly estimates income trajectories, employment dynamics, and sectoral choice, while incorporating unobserved heterogeneity. The details on the sequential EM used is detailed in section A1.4. Individuals are allowed to differ not only in their earnings levels and earnings mobility, but also in their propensity to enter non-employment or to work in the public sector, conditional on fixed characteristics such as labor market experience, education, and gender. These latent patterns give rise to distinct income and mobility classes within the data. Lifetime values of jobs are then constructed for each sector and analyzed comparatively across states, transition classes, earnings groups, and gender.

1.3.1 Structure

The data is structured as follows: For each individual $i \in (1, ..., N)$, we have a record of respective income flows, employment states, fixed characteristics, denoted by a vector $\boldsymbol{x}_i = (\boldsymbol{y}_i, \boldsymbol{S}_i, \boldsymbol{z}_i^v, \boldsymbol{z}_i^f)$ over a time period T_i. Each boldface variable here represents a vector.

- $\boldsymbol{y}_i = (y_{i1}, ..., y_{iT_i})$ is the sequence of individual i's log income (real net annual wage) at each time period t where $t \in (1, 2, , ..., T_i)$
- $\boldsymbol{S_i} = (S_{i1}, ..., S_{iT_i})$ short for state, records the sectoral state of the worker. The different states are defined as follows for a person i in the t^{th} year:
  - $S_{it} = 0$: not employed.
  - $S_{it} = 1$: employed in a full-time private sector job.
  - $S_{it} = 2$: employed in a full-time public sector job.
  - $S_{it} = 3$: employed in a part-time private sector job.
  - $S_{it} = 4$: employed in a part-time public sector job.
- $\boldsymbol{z}_i^v = (z_{i1}^v, ..., z_{iT_i}^v)$ is the sequence of time-varying individual characteristics. In this analysis labor market experience over a decade (defined as the cumulative duration of employment of individual i) and its square are considered. Note that conditional on the *state* at time t, z_i^v is deterministic.
- $\boldsymbol{z}_i^f$ is the set of individual fixed characteristics. It includes highest academic qualification, labor market experience when first observed in the sample, and gender. $\boldsymbol{z}_i^v$ is thus deterministic conditional on $\boldsymbol{z}_i^f$.

In addition to observed individual heterogeneity captured by $\boldsymbol{z}_i^v$ and $\boldsymbol{z}_i^f$, there might exist certain unobserved individual characteristics which may influence wages or movement or selection into the various labor market states. Thus we should supplement the data vector $\boldsymbol{x}_i$ by appending a set $\boldsymbol{k}_i$ of such (time-invariant) unobserved characteristics. The aim of the model is to simultaneously

estimate transitions between the aforementioned labor market states and income trajectories within and between employment sectors. To this end, individual contributions to the *complete* likelihood—i.e. the likelihood of $(\boldsymbol{x}_i, \boldsymbol{k}_i)$, including unobserved variables is described as:

$$\mathcal{L}_i(\boldsymbol{x}_i, \boldsymbol{k}_i) = \ell_i(\boldsymbol{y}_i|\boldsymbol{S}_i, \boldsymbol{z}_i^f, \boldsymbol{k}_i) \cdot \ell_i(\boldsymbol{S}_i|\boldsymbol{z}_i^f, \boldsymbol{k}_i) \cdot \ell_i(\boldsymbol{k}_i|\boldsymbol{z}_i^f) \cdot \ell_i(\boldsymbol{z}_i^f) \tag{1.3.1}$$

Through each individual decomposition in equation (1.3.1), the aim is to capture the distribution of observed individual characteristics; the distribution of unobserved individual heterogeneity given observed characteristics; and the likelihood of individual earnings and labor market state trajectories given individual heterogeneity. Then, estimates are obtained from parameters by maximizing the sample log-likelihood

$$\sum_{i=1}^{N} \ln \left(\sum_{k_i^m=1}^{K^m} \sum_{k_i^y=1}^{K^y} \mathcal{L}_i(\boldsymbol{x}_i, \boldsymbol{k}_i) \right) \tag{1.3.2}$$

The next sub-sections delve deeper into the individual components of the individual likelihood equation (1.3.1) and elaborate on their implications and functional forms.

1.3.2 Unobserved heterogeneity

Starting from the third term of the likelihood function $\mathcal{L}_i(x_i, k_i)$ given by equation (1.3.1), two types of unobserved heterogeneity, $k_i = (k_i^y, k_i^m)$ is considered. This is motivated by the evidence presented in subsections 1.2.2 and 1.2.3 that men and women move differently in the labor market and thus earn different wages.

- k_i^m : **Transition class**. This accounts for heterogeneity of individuals in terms of propensity to be in the different non-/employment states, S_i. k_i^m conditions the parameters related to employment and sectoral histories.
- k_i^y : **Income class.** This accounts for heterogeneity in terms of income. k_i^y conditions the income parameters relating to individuals' income process.

Both these heterogeneity classes are time-invariant random effects for an individual. The transition class, k_i^m is conditional on the fixed individual characteristics and helps in modelling the selection of individuals into the various employment states, S_i. k_i^y on the other hand, being time invariant as well, helps to increase the persistence of income ranks of an individual. In essence, we are using a finite mixture approach here to model unobserved heterogeneity of individuals, where each individual can belong to one of K^m transition classes and K^y income classes. The total number of possible classes is thus $K = K^m \times K^y = 12$.

Details on estimation of unobserved heterogeneity, transition classes, and the number of classes chosen is presented in the appendix sections A1.3.1 and A1.3.2.

1.3.3 Income process and Income class: K^y

The first term of the individual likelihood equation (1.3.1) models the likelihood of the individual's income path over the years. Assume y_{it}, the log-income of individual i in year t to be the realization of a first-order Markov process. The likelihood of an individual's income trajectory can thus be expanded as:

$$\ell(\mathbf{y}) = \ell(y_1) \cdot \prod_{t=2}^{T} \ell(y_t|y_{t-1}) = \ell(y_1) \cdot \prod_{t=2}^{T} \frac{\ell(y_t, y_{t-1})}{\ell(y_{t-1})} \tag{1.3.3}$$

Now each of these yearly log-income streams, are assumed to be normal, and conditional on current labor market state and observed and unobserved heterogeneity.

$$y_{it}|S_{it}, z^v_{it}, z^f_i, k^y_i \sim \mathcal{N}(\mu_{it}, \sigma^2_{it})$$

$$\text{with } \mu_{it} = \mu(S_{it}, z^v_{it}, z^f_i, k^y_i) \text{ and } \sigma_{it} = \sigma(S_{it}, z^v_{it}, z^f_i, k^y_i)$$

We model the μ_{it} using an OLS regression of y_{it} over the current labor market state and observed and unobserved heterogeneity.

$$\mu(S_{it}, z^v_{it}, z^f_i, k^y_i) = \begin{pmatrix} S_{it} & z^v_{it} & z^f_i & k^y_i & S_{it} * k^y_i \end{pmatrix}' \cdot \mu \tag{1.3.4}$$

Here the $*$ represents the interaction term of two variables.

$$\sigma(S_{it}, z^v_{it}, k^y_i) = \sqrt{\exp\left[\begin{pmatrix} S_{it} & z^v_{it} & k^y_i & S_{it} * k^y_i \end{pmatrix}' \cdot \sigma \right]} \tag{1.3.5}$$

The functional form of $\sigma(\cdot)$ constrains the standard deviation of the log-income y_{it} to be positive. The effect of the fixed individual heterogeneity z^f_i is subsumed through its link to k^y_i. Using y_{it}, μ and σ, the normalized log-income is constructed as $\tilde{y}_{it} = \frac{y_{it} - \mu_{it}}{\sigma_{it}}$. Therefore, $\tilde{y}_{it} \sim \mathcal{N}(0, 1)$.

The pair $(\tilde{y}_{it}, \tilde{y}_{i,t-1})$ is a vector which follows a bivariate multinomial normal distribution. It has a covariance matrix $\underset{\sim}{\tau}^{(2)}_{it}$ which can be expanded as:

$$\underset{\sim}{\tau}^{(2)}_{it} = \begin{pmatrix} 1 & \tau_{i,t,t-1} \\ \tau_{i,t,t-1} & 1 \end{pmatrix}. \tag{1.3.6}$$

τ is individual-specific and is allowed to vary with observed and unobserved heterogeneity and with employment *state* at t and $t-1$. We therefore create a function, $\tau_1(\cdot)$,

$$\tau_{i,t,t-1} = \tau_1(S_{it}, S_{i,t-1}, z^v_{it}, z^f_i, k^y_t, k^m_i). \tag{1.3.7}$$

The functional form of $\tau_1(\cdot)$ is:

$$\tau_1(S_{it}, S_{i,t-1}, z^v_{it}, k^y_i) = -1 + 2 \cdot \Lambda\left[\begin{pmatrix} S_{it} & S_{i,t-1} & z^v_{it} & k^y_i & k^m_i & S_{it} * k^y_i & S_{i,t-1} * k^y_i \end{pmatrix} \cdot \zeta \right] \tag{1.3.8}$$

Here $\Lambda(x) = (1 + e^{-x})^{-1}$. The functional form of $\tau_1(\cdot)$ includes a transformation in the form of $-1 + 2 \cdot \Lambda(\cdot)$. This constrains the correlation coefficient between normalized incomes at year t and t-1 to lie between [-1,1]. Likelihood of an individual's income trajectory $\mathbf{y_i}$ defined earlier in equation (1.3.3) is:

$$\ell(\mathbf{y}) = \ell(y_1) \cdot \prod_{t=2}^{T} \frac{\ell(y_t, y_{t-1})}{\ell(y_{t-1})}$$

The numerator in this expression can be written as a joint density of a pair of normalized log-earnings, $\tilde{y}_{it} = \frac{y_{it} - \mu_{it}}{\sigma_{it}}$

$$\ell(y_t, y_{t-1}) = \frac{1}{\sigma_t \sigma_{t-1}} \cdot \varphi_2\left(\tilde{y}_{it}, \tilde{y}_{i,t-1}; \underline{\tau}_{it}^{(2)}\right) \tag{1.3.9}$$

where $\varphi_2(\cdot; \underline{\tau}^{(2)})$ is the bivariate normal pdf with mean 0 and covariance matrix $\underline{\tau}^{(2)}$[10]. Substituting this expression to equation (1.3.3) gives us:

$$\begin{aligned}
\ell(\mathbf{y}) &= \ell(y_1) \cdot \prod_{t=2}^{T} \frac{\ell(y_t, y_{t-1})}{\ell(y_{t-1})} \\
&= \frac{\varphi_1(\tilde{y}_1)}{\sigma_1} \cdot \left(\prod_{t=2}^{T} \frac{\varphi_2(\tilde{y}_t, \tilde{y}_{t-1}; \tau_t^{(2)})}{\varphi_1(\tilde{y}_{t-1})} \cdot \prod_{t=2}^{T} \frac{1}{\sigma_t} \right) \\
&= \left(\prod_{t=1}^{T} \frac{1}{\sigma_t} \right) \cdot \frac{\prod_{t=2}^{T} \varphi_2(\tilde{y}_t, \tilde{y}_{t-1}; \tau_t^{(2)})}{\prod_{t=2}^{T-1} \varphi_1(\tilde{y}_t)}
\end{aligned} \tag{1.3.11}$$

With these specifications in place, normalized incomes are thus assumed to follow an AR(1) process:

$$\tilde{y}_{it} = \rho_{it} \cdot \tilde{y}_{i,t-1} + \varepsilon_{it} \tag{1.3.12}$$

where the innovations ε_{it} are normal with mean 0 and serially uncorrelated.

1.3.4 Linking ρ to τ and σ^2

Equation 1.3.12 describes the AR(1) process for the normalized wage. We now resort to the standard relations between these parameters in a typical AR(1) process. We recall, in a first order autoregressive process, the autocovariance at lag 1,τ_1 is related to the σ and the autocorrelation coefficient ρ as:

$$\tau_1 = \frac{\sigma^2}{1 - \rho^2} \tag{1.3.13}$$

[10]Note that $\varphi_2(\cdot)$ is defined as:

$$\varphi_n(\boldsymbol{Y}; \Sigma) = \frac{1}{\sqrt{(2\pi)^2 \Delta_\Sigma}} \cdot exp(-\frac{1}{2} \cdot (\boldsymbol{Y} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{Y} - \boldsymbol{\mu})) \tag{1.3.10}$$

Simplifying and solving for ρ, we get

$$\rho = \frac{-\sigma^2 + \sqrt{\sigma^4 + 4\tau^2}}{2\tau}, \tag{1.3.14}$$

Now that we have ρ, we can recreate the income process based on the first wage and Equation 1.3.12. By substituting the equation for normalized log-wages into the AR(1) process, we can return to the original (unnormalized) process. Specifically, using $\tilde{y}_{it} = \frac{y_{it}-\mu_{it}}{\sigma_{it}}$:

$$y_{it} = \mu_{it} + \sigma_{it} \cdot \left(\rho_{it} \cdot \frac{y_{i,t-1} - \mu_{i,t-1}}{\sigma_{i,t-1}} + \varepsilon_{it} \right). \tag{1.3.15}$$

This equation describes how log-wages at time t depend on:

(1) The individual-specific mean μ_{it}, which reflects systematic factors such as experience, education, and sector,
(2) The deviation from the mean at the previous period, $(y_{i,t-1} - \mu_{i,t-1})$, weighted by the persistence parameter ρ,
(3) A stochastic noise term, $\sigma_{it} \cdot \varepsilon_{it}$, capturing unobserved factors.

This approach ensures that the model incorporates the underlying persistence of log-wages while accounting for the role of noise and covariate-driven variation.

1.4 Model Fit

1.4.1 Understanding the latent classes

The EM algorithm was run until convergence (distance of parameters $< 10^{-3}$) in order to arrive at the final estimates of the $\kappa^m, \chi, \chi_0, \mu, \sigma, \xi, \kappa^y$ coefficients. These final estimates of the coefficients are reported in the appendix table (A1.10) through (A1.14).

The estimation procedure identifies distinct latent transition and income classes, capturing the heterogeneity in employment dynamics and wage trajectories across the public and private sectors. The classification reveals four transition classes (related to employment sector mobility) and three income classes (characterizing wage persistence and mobility), jointly shaping career outcomes. We first look at the mobility estimations for our panel data. The individuals in our panel are sorted into 4 transition classes. 42% individuals are sorted into the Transition class 0, around 4% in class 1, 47% in class 2 and 7% in class 3. Note that the order is meaningless. The transition classes have the distribution as in Table (1.5). Note that women are overrepresented in transition classes with higher public sector participation, as expected.

Although the estimated latent transition classes differ markedly in their composition by sex, education, and age, the transition dynamics implied by the model are remarkably similar once these observed characteristics are conditioned on. In other words, the classes capture more the differences in *who*

Table 1.5. Share of individuals by Sex, Education, Transition Class and Income Class

Transition Class	Share of individuals	Female share	Education		
			Low share	Med share	High share
0	42%	65%	51%	16%	32%
1	4%	66%	71%	29%	0%
2	47%	33%	86%	14%	0%
3	7%	56%	68%	10%	22%
Income Class	**Share of individuals**	**Female share**	**Education**		
			Low share	Med share	High share
0	8%	40%	60%	13%	26%
1	32%	57%	72%	17%	11%
2	60%	47%	69%	15%	16%

participates in different segments of the labor market rather than differences in *how* individuals move between employment states over time.

The patterns of sectoral persistence and transition—such as the high stability of full-time employment in both the public and private sectors, and the relatively frequent transitions from part-time to non-employment–are close across transition classes. However, there are differing propensities in the latent classes: specifically, those in transition class 0, $k^m = 0$, have a slightly higher propensity to go into public full-time roles, those in $k^m = 1$ to public part-time roles, those in $k^m = 2$ to private full-time roles and finally those in $k^m = 3$ to private part-time roles. This is evidenced by the multinomial logit parameters in table A1.10.

Unobserved heterogeneity plays a fundamental role in explaining wage dynamics in our dataset. The latent income classification (k^y) uncovers distinct subpopulations characterized by different wage persistence and mobility regimes that go beyond observed demographics and labor market states. These income classes capture hidden variation in earnings trajectories that cannot be fully attributed to measured covariates like education, age, or sectoral affiliation.

The model's goodness-of-fit notably improves with the introduction of these latent income classes, reflecting important wage heterogeneity within transition classes (k^m). For instance, the highest income class ($k^y = 0$), comprising about 8% of individuals, is marked by stable high earnings and high education levels, while the largest income class ($k^y = 2$, 60%) exhibits greater wage volatility and lower overall earnings.

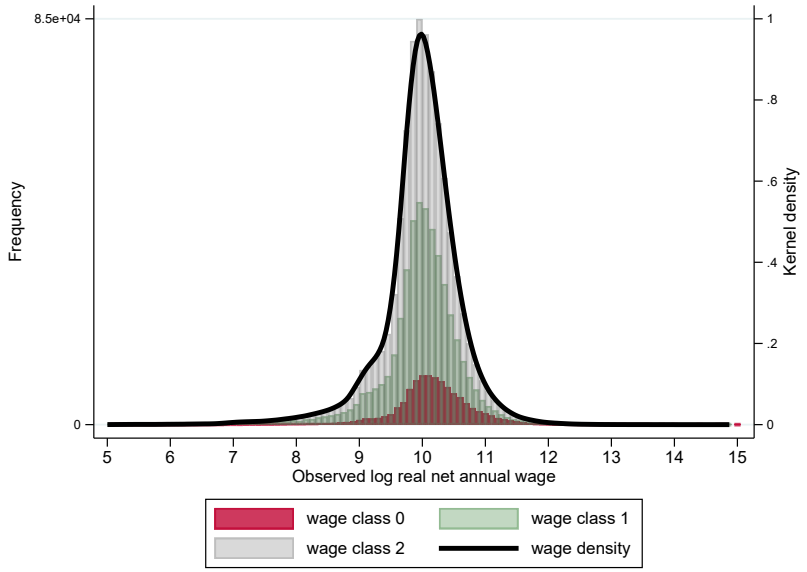


Figure 1.3. Observed latent wage class distribution

Figure 1.3 visually presents these distinct income distributions and highlights the stratification in wage stability and volatility across classes.

Taking a combined perspective on both income (k^y) and transition (k^m) latent classes, we observe layered unobserved heterogeneity: while the transition classes primarily capture compositional differences in employment states and demographics, the income classes reveal important heterogeneity in wage dynamics conditional on these employment patterns. This implies that unobserved heterogeneity in the model is multi-dimensional, affecting both employment transitions and wage mobility but through somewhat distinct latent structures.

This layered latent typology enriches our understanding of stratification and persistence mechanisms shaping the public-private wage gap and labor market careers. The detailed characterization of all twelve combined latent classes ($k^y \times k^m$) is compiled in the appendix (Table A1.7), providing comprehensive insight into the interaction between employment stability and wage mobility across demographic groups.

While this paper employs a similar latent class framework to Dickson et al. (2014), the findings reveal less pronounced unobserved heterogeneity in mobility classes. This difference can be attributed to the broader and more heterogeneous sample used here, which includes both men and women across five employment states (adding part-time work in both sectors). The inclusion of gender as an observed covariate absorbs substantial variation in sectoral attachment and employment stability that, in a male-only sample, would appear as unobserved mobility heterogeneity. Moreover, the richer state space—incorporating part-time employment—introduces more fluid and overlapping transition patterns, making it difficult for latent mobility classes to capture sharp behavioral distinctions. In contrast, the income classes reveal meaningful unobserved heterogeneity in wage dynamics, consistent

with prior findings, suggesting that while observed factors largely explain sectoral mobility, wage trajectories retain important latent structure even after controlling for demographics and employment states.

1.4.2 Prediction

As is customary, the statistical model's goodness-of-fit and predictive power were first tested in-sample presented in subsection 1.4.2.1. What has not been done in this particular literature thus far is the out-of-sample test with a data size three times as large as the training sample. This is presented in subsection 1.4.2.2.

1.4.2.1 In-sample

Let individuals be assigned initial states for the first year that they appear in the panel, based on probabilities generated by equation (A1.3.4) conditioned on individual fixed characteristics. The method of probabilistic assignment is used so as to respect the probability distributions and not directly assign the state with the highest probability [11]. The probabilities from equation A1.3.3 are used to assign the subsequent states, again employing probabilistic assignment. We can thus, test the goodness of fit and the predictive power of the model by checking year-on-year state transitions in the observed data and comparing it with the same transitions within the predicted data. We can observe quite similar patterns within these transitions as tabulated in Table A1.8 compared with the earlier descriptive table 1.2. The predictive strength is corroborated by the fact that the maximum distance between the observed and predicted transition probability matrices is 0.01 on aggregate.

Using the income process parameters, the log (real net annual) wages can be predicted. Once the simulation is done, we compare the simulated incomes to the observed incomes by comparing the observed and predicted wage densities as in figure 1.4. The predicted latent class composition is also tested: table 1.5 is almost exactly replicated. Moreover, the income/wage class composition predicted as in figure A1.3 is very close to the observed figure 1.3. A word of caution here, the model slightly overpredicts full-time wages and slightly underpredicts part-time wages as shown in figure A1.4. However, figure 1.4 (below) and figure A1.3 (in the appendix) provide strong evidence for the predictive power of this model. The actual observed and predicted income distributions are a near exact match, including the small kink towards the left of the observed income distribution.

[11] Probabilistic assignment is chosen over deterministic assignment to prioritize accuracy. Results are exactly replicable as the `set seed` command was used. Moreover, the correlation between values generated by probabilistic assignment and deterministic assignment is, on average 0.7.

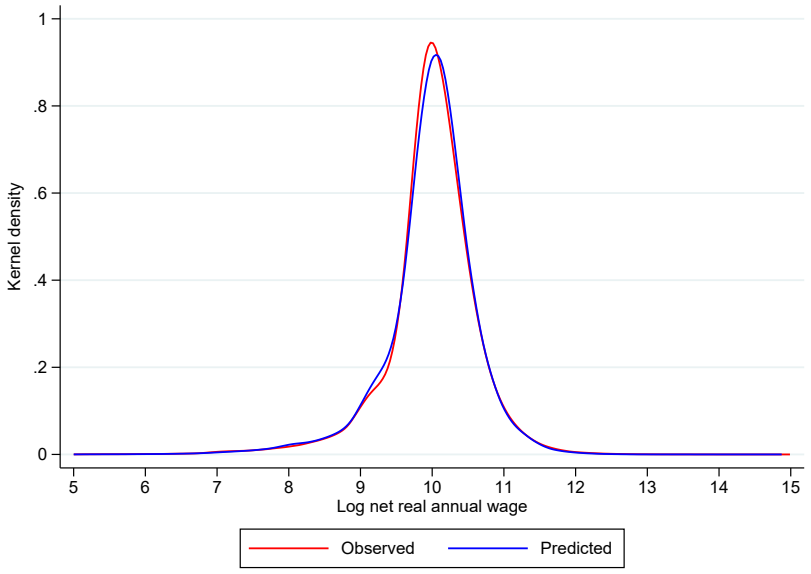


Figure 1.4. Observed and predicted wage densities

1.4.2.2 Out-of-sample

To test the model's robustness beyond the estimation sample, I applied the trained coefficients to a larger out-of-sample panel comprising 829,232 individuals and 6,194,037 observations (thrice the size of the in-sample panel used). In this prediction process, I retained only each individual's first observed employment spell and used the estimated coefficients to simulate their subsequent (non-)employment trajectories and associated wages. This approach allows for assessing how well the model generalizes to a broader population while accounting for both observed and unobserved heterogeneity in labor market outcomes.

All the state transition and wage prediction results presented in the in-sample section 1.4.2.1 were also replicated for this larger panel and are included in the appendix A1.5.2. The consistency of the predicted trends across samples reinforces the model's validity in capturing employment mobility, wages, and sectoral differences over time.

1.5 Results: Public and Gender Premia in Lifetime Earnings

Constructing Lifetime Earnings

Once the maximization problem is resolved and we have stable parameters of the income and job mobility parameters, simulations of employment and income trajectories for the individuals in the sample are carried out until retirement age. Assuming that, upon retiring (which is assumed to happen at age 60 denoted as T_R), a given individual enjoys a residual value of V_R, then the lifetime value as of

experience level t of an individual's simulated future income trajectory $y_{s\geq t}$ is written as:

$$V_t(y_{s\geq t}) = \sum_{s=t}^{T_R} \beta^{s-t} \cdot \exp(y_s) + \beta^{T_R-t} \cdot V_R \tag{1.5.1}$$

where $\beta \in (0,1)$ is the discount factor and $\exp(y_t)$ is the projected real net annual wage that the individual earns from (log) income y_t. We could also interpret $exp(y_t)$ as the risk-neutral utility from earnings. The log of the above V_t - the log of the discounted sum of lifetime earnings (hereon referred as "lifetime earnings") under different pathways are calculated and compared sectorally. The discount factor $\beta = 0.95$ per annum, as is standard in the literature. The value of retirement is defined as $V_R = \frac{1-\beta^{22}}{1-\beta} \times RR \times e^{y_{T_R-1}}$ in both the scenarios. Assuming that after retirement, individuals receive a constant flow of income equal to RR times their last income in activity and discount this flow over a period of 22 years (time between average retirement age and average life expectancy which is 82 years, in France). The particular values of $\beta = 0.95$ and $RR = 0.4$ that were taken obviously condition the results described. A lower β would clearly make lifetime values close to current income flows and would magnify the observed public-private pay gap (the private sector earnings being far more dispersed). An increase in RR on the other hand, would only magnify the impact of the income gap at the time of retirement. For the main results presented in section 1.5, the RR used is 0.40 using the precedent set by Dickson et al. (2014). For robustness, all the specifications to calculate the discounted sum of lifetime earnings were also estimated with an $RR = 0.7$ and separate rates for the public and private sectors: $RR_{pub} = 0.75$ and $RR_{pvt} = 0.71$. The results are largely consistent.

An implicit assumption in computing lifetime values, is that the prevailing economic environment is stationary and individuals anticipate aging and experiencing changes in their wage and job mobility given their current state and income level, but not the inherent model paramaters to be changing over their working life. Given the stability of the share of public sector employment and of the non-employment rate in France over the sample period, the sample time-period can be considered as an average state of the business cycle. The public premium is defined as the difference in log-lifetime values (referred to as "lifetime earnings" hereon) between the public and private sector. The 'whole sample, with selection' graph relates to predicted 'raw' differences, i.e. it plots the difference between quantiles of lifetime earnings among individuals effectively observed to hold public jobs in the initial period, and corresponding quantiles of lifetime earnings among workers observed to hold private jobs in the initial period. The whole sample graphs relate to predicted differences in lifetime earnings across sectors for all individuals in the sample, i.e. it compares income and lifetime values that each individual could potentially earn in either sector.

In this section, I present a series of insights that stem from comparing lifetime earnings[12] across sectors, by gender and education, with and without selection to offer a more holistic understanding of what we perceive as the public premium, first focusing on if there is one at all. Two broad sets of counterfactuals are considered: one, is to compare the "job-for-life" lifetime earnings wherein individuals (on aggregate, and by gender, by education, etc.) are assumed to hold either a public

[12]Recall that "lifetime earnings" refers to the log of the discounted sum of future earnings as defined in equation 1.5.1.

or private sector job in every period from the time they are first observed until age 60. Two, is to compare lifetime earnings "with mobility", meaning, people can transition into any of the 5 states (non-employment, full-time employment in private sector, in public sector, part-time employment in private and public sectors) until retirement based on the transition probabilities as per their observed (gender, education, experience) and unobserved factors (latent types). The first set of counterfactuals sheds light on what the earnings gap amount to over a lifetime *ceteris paribus* assuming that people are employed in every period, full-time in either sector. The second set ("with mobility") gives a more realistic and accurate picture as it estimates lifetime earnings "with mobility", taking transition patterns into account.

1.5.1 The Lifetime Public Sector Premium

A significant public sector premium exists in France, particularly for women, low-educated workers, and individuals over 45. This premium persists across most of the earnings distribution even after accounting for self-selection into sectors. An oft-encountered question here is if the public premium is just a result of selection. The short answer is no, as many studies confirm. But the selection effect indeed estimates higher and larger premia or penalties as the case may be as deconstructed herein. Figure A1.8 compares the counterfactual lifetime earnings from the "job-for-life" case with the counterfactual lifetime earnings wherein individuals, conditional on selection into either sector, transition freely across employment states. Of course, these transitions are informed by the patterns that are already captured from the model conditioned on observables and unobservables. Unsurprisingly, lifetime earnings are higher if individuals hold a full-time job in either sector in every period, but we know that is rare for everyone to experience. The losses in lifetime earnings is higher in private sector due to higher transitions into non-employment and worse part-time pay compared to the public sector.

Job-for-life case: *potential* lifetime earnings differences if everyone could hold a full-time job in every period. With selection there is a positive public premium until the 46th percentile on aggregate. But even without selection, there is one until the 30th percentile. To understand how much of the premium we see is just a matter of selection, we must first quantify if there is a public premium "with selection" and then, if there is one "without" and for whom, if at all.

- **"with selection"**: What happens if those observed to be in public sector full-time jobs were forced to switch to the private sector instead?
  Main result: Is a public premium observed? Yes. For whom? Women (until the 69th percentile), Low educated workers (until the 60th percentile) and older workers (over 45) (until the 76th percentile).
- **"without selection"**: What happens if everyone in the sample is employed full-time in the public sector vs. the private sector?
  Main result: Is a public premium observed? Yes. For whom? Women (until the 38th percentile) Low educated workers (until the 38th percentile) and older workers (over 45) (until the 49th percentile).

Figure A1.8 illustrates the public premium in lifetime earnings for job-for-life counterfactuals, both with and without selection. The graph demonstrates that a public premium persists even when selection is eliminated, though to a lesser extent and for a smaller portion of the earnings distribution. With selection, the public premium extends up to the 46th percentile of the aggregate earnings distribution, while without selection, it reaches the 30th percentile. For women, the premium without selection is observed up to the 38th percentile, compared to the 69th percentile with selection. Similarly, for low-educated workers, the premium without selection persists up to the 38th percentile (versus 60th with selection), and for older workers over 45, it extends to the 49th percentile (compared to 76th with selection).

These findings underscore that while selection plays a significant role in amplifying the observed public premium, it is not the sole factor. The persistence of the premium even without selection indicates structural differences in compensation between the public and private sectors, particularly for certain demographic groups and at lower earnings and education levels. This analysis reinforces the conclusion that the public premium is not merely a result of selection, but reflects underlying differences in sectoral compensation structures. One might be tempted looking at figure A1.8 to compare the bars across categories, for example, comparing the premia for men vs. for women to see potentially if women have a higher public premium than men, but that would be misleading. The bars in figure A1.8 correspond to counterfactual earnings *within* that category (the percentage difference of lifetime earnings of low educated women first observed to be in the public sector from the lifetime earnings of low educated women if forced to be in the public sector give us the "with selection" red bar). To answer the question about gender premia by sector, we must compare the lifetime earnings of men and women (section 1.5.2).

With mobility: *true* lifetime earnings differences show that losses in lifetime earnings higher in private sector. The estimated loss in earnings, calculated as the difference in (log) lifetime earnings of the job-for-life estimates and those with mobility, reveals that individuals initially observed in the private sector full-time experience greater income erosion due to higher transitions into non-employment and weaker part-time wage prospects. Converting these log differences into percentage terms, the lifetime earnings loss ranges from approximately 82% for the lowest quantiles to 5% for the highest quantiles in the private sector, whereas losses in the public sector are consistently lower (nearly half, throughout), ranging from 63% to 2%. This highlights the structural disadvantages faced by private-sector workers in earnings stability, emphasizing the role of sectoral labor mobility in shaping long-term economic outcomes. Of course, this effect is not gender-neutral either and the gaps are larger for women.

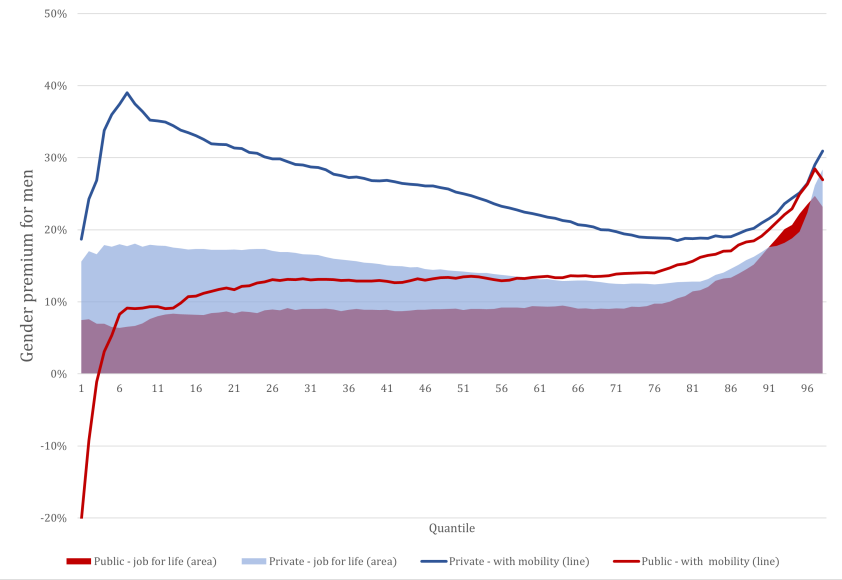


Figure 1.6. Gender premium in lifetime earnings

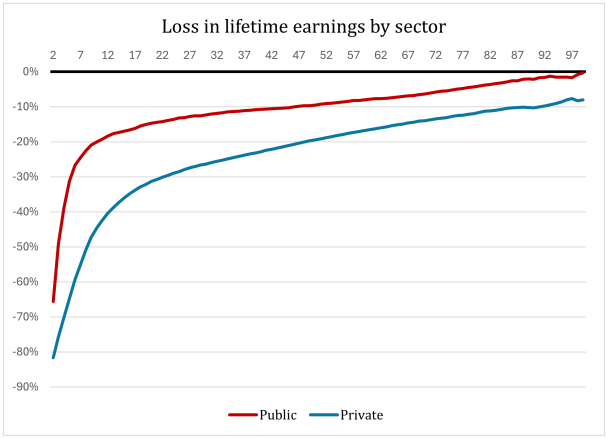


Figure 1.5. Loss in lifetime earnings with selection
Lifetime earnings percentage difference of job-for-life case and mobility case (the latter takes into account the probability of transitioning to non-employment or part-time work).

1.5.2 Men benefit from a gender premium in both sectors

Whether considering the job-for-life case in counterfactual lifetime earnings or the case with mobility, the takeaway remains that men benefit from a gender premium in both sectors. This comparison can be made in two ways: given selection into working full-time in either sector, the lifetime earnings quantiles of men and women can be compared sectorally. Alternatively, their lifetime earnings with mobility can be compared, again conditional on selection. Figure 1.6 graphically depicts both these cases. As expected, the case that takes mobility into account (lines in the figure) produces higher gender gaps relative to the job for life case (area in the figure), particularly in the private sector. The public sector gender premium is almost always lower compared to private. A deeper dive into these differences by gender and by education categories reveals that on average, higher education lowers this gender premium that men benefit from in both sectors, even though glass ceilings exist in both sectors.

A particularly novel insight emerges when analyzing the distribution of lifetime earnings quantiles between men and women in each sector. The findings show that the male gender premium is present across the entire distribution but is more pronounced at higher earnings quantiles, especially in the private sector. This indicates that as earnings increase, men benefit from larger relative advantages, whereas women's wages remain comparatively compressed, particularly in the public sector. Although much of the literature on gender pay gaps focuses on penalties faced by women, this study provides new quantitative evidence on the persistence of male wage advantages across the earnings distribution. Even in a labor market where public employment offers greater stability and wage regulation, men continue to systematically earn more than women at respective lifetime earnings quantiles, reinforcing cumulative gender inequalities over the career cycle.

1.5.3 Returns to education are lower in the public sector...

- **...for men and for women** Figure 1.7 illustrates that returns to education are consistently lower in the public sector compared to the private sector, for both men and women, across all lifetime earnings quantiles conditional on selection. The negative gap in lifetime earnings differences between public and private sector workers increases with education level, with highly educated workers experiencing the largest lifetime earnings penalty in the public sector.
  - Among low-educated workers, differences between public and private sector earnings remain moderate, with some evidence of a public sector premium at the lowest quantiles. This suggests that for lower-income workers, public sector employment still offers some earnings security benefits, particularly in early career stages.
  - However, for medium-educated workers, and even more so for highly educated workers, the public sector disadvantage becomes increasingly pronounced across the distribution, with lifetime earnings penalties widening at the upper quantiles. This suggests limited wage progression for highly skilled individuals in public sector jobs compared to their private sector counterparts.
  - The penalty for highly educated workers is particularly severe for men, reinforcing the idea that private sector careers provide greater returns to education through steeper earnings trajectories over a lifetime.

  These findings confirm that while public employment provides stability, it does so at the cost of significantly lower lifetime earnings, especially for those with higher educational attainment.
- **...but job stability and transitions to public part-time reduces these penalties, particularly for women** Despite the lower returns to education in the public sector, women, particularly those in the lower earnings quantiles, experience significant lifetime earnings advantages in the public sector when accounting for career stability and part-time transitions.
  - The "With Mobility" panels in Figure 9 show that women, particularly low-educated women, benefit significantly from public sector job stability. The public sector mitigates earnings volatility, allowing for a smoother and more predictable lifetime earnings trajectory.
  - The gap in lifetime earnings between public and private employment narrows significantly for women once sectoral transitions are considered, suggesting that career interruptions

and part-time work are better accommodated in the public sector. This is particularly relevant for women with medium and low education levels, who benefit from higher wage stability when transitioning to public part-time work.

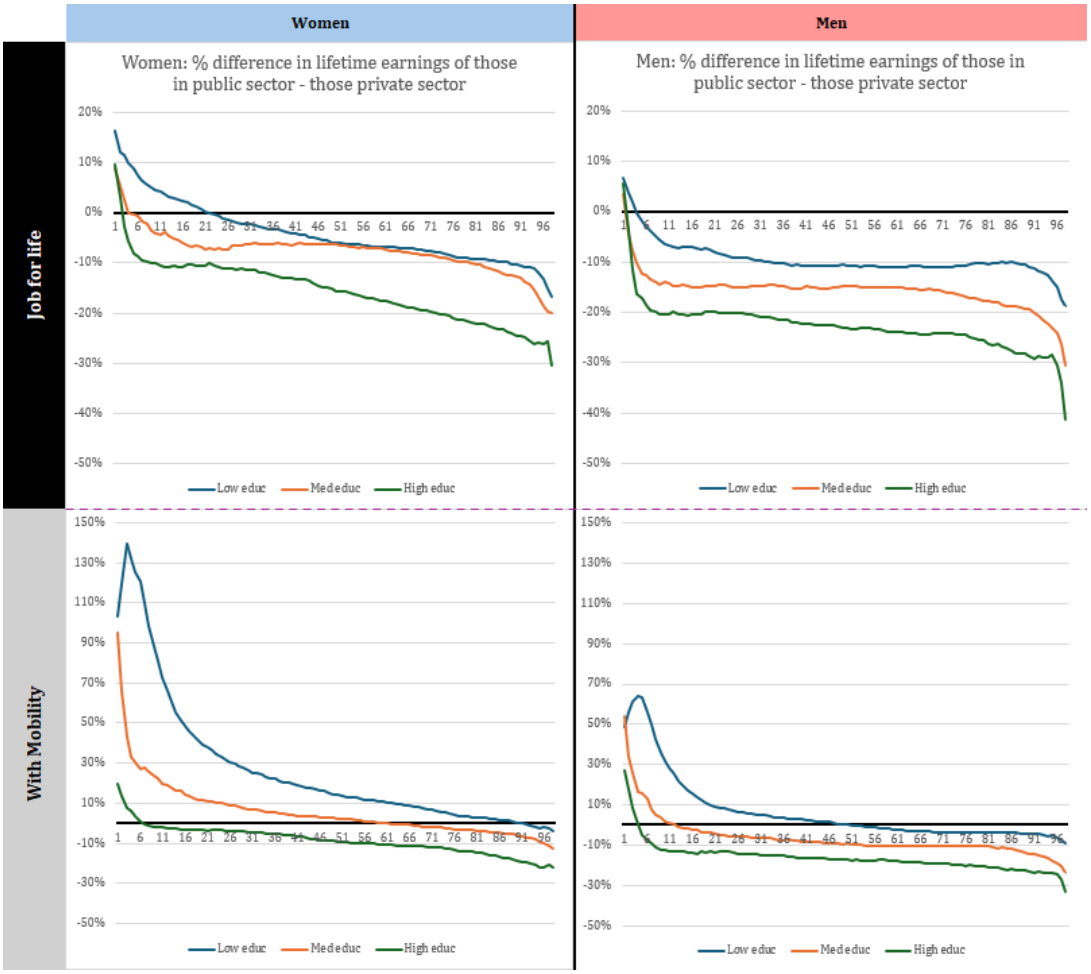


Figure 1.7. Percentage difference of lifetime earnings in the public and private sectors by education levels and sex

These findings highlight a key gendered dynamic: while the private sector provides higher returns to education, the public sector acts as a critical stabilizing force for women, particularly in lower and middle-income categories. Public sector employment appears to buffer the income shocks associated with career breaks and transitions to part-time work, making it a more viable long-term option for many women, despite lower formal returns to education.

1.6 Discussion

The findings of this study demonstrate that the model successfully captures observed income trajectories and sectoral mobility patterns, providing a robust framework for analyzing lifetime earnings differentials between public and private sector workers. Notably, the model replicates the substantial earnings dispersion between men and women, particularly in the private sector, where wage trajectories exhibit

greater volatility and gender disparities are more pronounced. However, while the model effectively accounts for sectoral differences in wage dynamics and transitions, institutional factors beyond direct wage earnings—such as pension schemes—remain an important avenue for future research. The French public and private sector pension systems differ significantly, with public sector workers typically benefiting from more stable retirement incomes. While these differences are not explicitly captured in this dataset, policy changes in pension structures could have significant implications for lifetime earnings differentials across sectors. Future research could extend this analysis by incorporating expected retirement benefits and pension wealth accumulation to provide a more comprehensive measure of lifetime earnings.

Another promising avenue for future work lies in the integration of fertility and household employment decisions into the modeling framework. Labor market outcomes are often shaped by household-level choices, particularly for women, whose career trajectories are more likely to be influenced by caregiving responsibilities. A dynamic approach that accounts for joint household decision-making and the impact of parental leave policies could offer deeper insights into the gendered effects of public versus private sector employment.

Finally, the findings suggest broader implications beyond the French labor market. The gendered lifetime public premium observed in this study is likely to be even more pronounced in developing economies, where public sector employment often provides one of the few stable career pathways. However, the extent of this effect will depend on institutional labor protections, the degree of informality in private sector employment, and access to social security systems. Future comparative research could test this hypothesis by analyzing cross-country differences in public-private earnings trajectories and gender disparities, particularly in economies where public employment is a critical employer for educated women.

By expanding on these dimensions—pension structures, household decision-making, and cross-country comparisons—future research can further refine our understanding of how sectoral employment choices shape lifetime earnings and economic inequalities across gender and educational backgrounds.

1.7 Conclusion

By using a dynamic model to capture unobserved heterogeneity and career transitions, I reconcile conflicting prior findings: while hourly wage analyses suggest penalties for educated workers, lifetime earnings reveal a public premium for women and low-skilled employees, driven by stability and part-time compensation. Yet, this premium comes at a cost. Highly educated workers trade a part of their lifetime earnings in the public sector, reflecting rigid wage progression that stifles returns to education.

These results carry implications beyond France. As a prototype of coordinated market economies, France's experience highlights the dual role of public sectors: mitigating gender inequality through stability while imposing ceilings on wage progression. The results confirm the existence of a public

premium in lifetime earnings, particularly for women, low-educated workers, and older employees. Even when accounting for selection effects, the public sector offers significant advantages to these groups in terms of income stability and reduced volatility. Policymakers in similar contexts (e.g., Scandinavia, Southern Europe) must weigh privatization's risks—eroding gender equity gains—against the need to reform rigid wage structures. For developing economies, where public sectors often dominate formal employment, the findings underscore the *potential* of public jobs to reduce inequality.

Future research should extend this framework to incorporate pension benefits and household dynamics, deepening our understanding of how institutions shape lifetime disparities. By bridging the gaps between literature on gender, education, and sectoral choice, this study offers a blueprint for analyzing equity-efficiency trade-offs in labor markets globally.

Bibliography

Olivier Bargain and Blaise Melly. Public sector pay gap in france: new evidence using panel data. Technical report, IZA Discussion Papers, 2008.

Olivier Bargain, Audrey Etienne, and Blaise Melly. Public sector wage gaps over the long-run: Evidence from panel administrative data. Discussion Paper 11924, IZA, 2018.

Magali Beffy and Thierry Kamionka. Public-private wage gaps: Is civil-servant human capital sector-specific? Working Paper 2010-55, CREST, 2010.

Marianne Bertrand, Claudia Goldin, and Lawrence F. Katz. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–255, July 2010. doi: None. URL https://ideas.repec.org/a/aea/aejapp/v2y2010i3p228-55.html.

Francine D. Blau and Lawrence M. Kahn. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865, 2017. doi: 10.1257/jel.20160995.

Matt Dickson, Fabien Postel-Vinay, and Hélène Turon. The lifetime earnings premium in the public sector: The view from europe. *Labour Economics*, 29:179–190, 2014. doi: 10.1016/j.labeco.2014.07.003.

Denis Fougère and Julien Pouget. Who wants to be a 'fonctionnaire'? the effects of individual wage differentials and unemployment probabilities on the queues for public sector jobs. Technical report, Mimeo, 2003.

Bertrand Garbinti, Jonathan Goupille-Lebret, and Thomas Piketty. Gender inequalities in lifetime earnings: Evidence from france. *American Economic Journal: Applied Economics*, 15(1):1–35, 2023. doi: 10.1257/app.20220265.

Pietro Garibaldi, Pedro Gomes, and Thepthida Sopraseuth. Public employment redux. *Journal of Government and Economics*, 1:100003, 2021. ISSN 2667-3193. doi: https://doi.org/10.1016/j.jge.2021.100003. URL https://www.sciencedirect.com/science/article/pii/S2667319321000033.

Claudia Goldin. A grand gender convergence: Its last chapter. *American Economic Review*, 104(4): 1091–1119, 2014. doi: 10.1257/aer.104.4.1091.

Pedro Gomes and Zoë Kuehn. You're the one that i want! understanding the over-representation of women in the public sector. *American Economic Journal: Macroeconomics*, 2025.

Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. What do data on millions of u.s. workers reveal about life-cycle earnings risk? *Econometrica*, 89(5):2303–2339, 2021. doi: 10.3982/ECTA14884.

James J. Heckman and Burton Singer. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320, 1984. doi: 10.2307/1911491.

Trond Petersen and Laurie A Morgan. Separate and unequal: Occupation-establishment sex segregation and the gender wage gap. *American Journal of Sociology*, 101(2):329–365, 1995.

Sophie Ponthieux and Dominique Meurs. Chapter 12 - gender inequality. In Anthony B. Atkinson and François Bourguignon, editors, *Handbook of Income Distribution*, volume 2, pages 981–1146. Elsevier, 2015. doi: 10.1016/B978-0-444-59428-0.00013-8.

Fabien Postel-Vinay and Hélène Turon. The public pay gap in britain: Small differences that (don't?) matter. *Economic Journal*, 117(523):F422–F442, 2007. doi: 10.1111/j.1468-0297.2007.02067.x.

CHAPTER 2

Job Mobility and Career Values

Evidence from Worker–Firm–Occupation Sorting

with Jean-Marc Robin

Abstract

This paper investigates the sources of wage dispersion and career earnings by jointly estimating sorting among workers, firms, and occupations, conditional on gender and age. We implement a flexible reduced-form CEM algorithm on French administrative data for 2015–2019. We observe job spells for worker i with gender f at firm j in occupation k, and recover latent worker types x and firm types y that constitute the unobserved heterogeneity. A job *is the pair (y, k); a worker's* labor-market state *is the combination (x, y, k), conditional on gender f and age. Using exact backward induction on the estimated Markov chain, we compute the net present value (NPV) of each state. We introduce three complementary decomposition methods to the sorting literature: Shapley variance decomposition, additive main-effects decomposition with pairwise sorting covariances, and Shapley-valued mutual information decomposition with synergy-redundancy analysis. Our main findings are four-fold: first, occupation is the single largest contributor to both wage and lifetime-value variation across all decomposition methods, and the dominant sorting channel runs between firms and occupations ($\rho = 0.201$), not between firms and workers. Second, the MI decomposition reveals that sorting among worker types, firm types, and occupations is approximately additive for wages (net synergy +1.5 pp) but deeply complementary for lifetime value: 20.2 percentage points of NPV entropy exist only in the interaction among factors and cannot be recovered from any single dimension. Third, gender operates primarily through non-additive complementarities—its own main effect is small, but 11 percentage points of wage variance and the bulk of the gender differences in wages and NPVs arise from cell-specific premia that depend on the particular worker type–firm type–occupation combination. Fourth, workers' destination choices respond to the capitalized current-wage component of value but not to the option-value component, indicating that observed mobility is driven by immediate wage signals rather than by the career-building value that the model reveals is present.*

When workers change jobs, do their transitions reflect forward-looking career optimization—or do they respond primarily to the wage signal in front of them? The distinction matters: if mobility is forward-looking, then the labor market sorts workers into jobs that maximize lifetime value; if mobility is myopic, then the allocation of workers to jobs is shaped more by immediate opportunities than by the career paths those opportunities imply. Answering this question requires two ingredients that standard sorting models do not jointly provide. First, the econometrician must observe—or recover—the full state that determines a worker's future earnings trajectory, including the worker's type, the firm's type, and the worker's occupation. Second, the econometrician must compute what each state is actually

worth over a career, not just at a point in time. This paper builds both ingredients and uses them to test whether observed job-to-job transitions are driven by current wages or by lifetime career value.

The dominant empirical framework for studying worker-firm sorting is the additive fixed-effects model of Abowd et al. (1999a). In that framework, log wages decompose into a worker effect α_i and a firm effect ψ_j, with assortative matching inferred from the correlation between the two. This approach has been applied extensively (e.g., Card et al., 2013; Song et al., 2019) and has produced important stylized facts about the contribution of firms to wage inequality. But it faces two well-known challenges. First, the additive structure imposes that the return to working at a high-ψ firm is the same for all workers, ruling out the complementarities that sorting theory predicts (Becker, 1973; Shimer and Smith, 2000; Eeckhout and Kircher, 2011). Second, the correlation between estimated $\hat{\alpha}_i$ and $\hat{\psi}_j$ is severely biased in finite samples (Andrews et al., 2008; Kline et al., 2020; Babet et al., 2023), making it difficult to sign the direction of sorting, let alone quantify it.

A growing literature addresses both problems by replacing the continuous fixed effects with discrete latent types. Bonhomme et al. (2019) introduce a distributional framework in which workers and firms are classified into a finite number of types, with cell-specific wage distributions that can accommodate arbitrary interactions between worker and firm heterogeneity. Bonhomme et al. (2022) extend this approach to settings where the underlying heterogeneity is continuous but the econometrician discretizes it for tractability. Lentz et al. (2023) push the framework, embedding discrete worker and firm types into a structural on-the-job search model with endogenous mobility. Their Classification EM algorithm—which iterates between classifying agents into types and re-estimating the model parameters—is the estimation engine we adopt in this paper.

All of these frameworks, however, operate in two dimensions: a worker type and a firm type. Occupation—the specific role a worker fills within a firm—is either absent, conditioned on, or absorbed into the two sides. This is a consequential omission. Occupations are not merely a feature of the worker or a feature of the firm: they reflect a distinct axis of heterogeneity with its own wage structure, its own transition dynamics, and its own contribution to lifetime value. A software engineer and a project manager at the same firm have different wages, different separation hazards, and different career trajectories. A two-way model that classifies both into the same firm type would potentially miss this variation entirely.

Several papers have studied occupations in the context of sorting and mobility, though from different angles than ours. Lise et al. (2016) develop an equilibrium search model with two-sided heterogeneity in which sorting arises endogenously from the interaction of match surplus and on-the-job search, establishing the structural framework that subsequent work extends to multiple dimensions. Lise and Postel-Vinay (2020) build on this foundation, allowing workers to accumulate multidimensional human capital—manual and cognitive skills—and sort across occupations that weight these skills differently. Their framework is rich and requires structural assumptions about production technology and the bargaining protocol. Lindenlaub (2017) develops the theory of multidimensional sorting

with continuous types, showing that the pattern of sorting depends on the interaction between the dimensionality of worker and job attributes. Lindenlaub and Postel-Vinay (2021) extend this to a search-and-matching environment with surplus-based matching. Azkarate-Askasua and Zerecero (2024) study occupational mobility and wage dynamics, documenting how occupation switches interact with employer changes to shape earnings growth. Borovičková and Shimer (2024) propose a complementary approach that decomposes wages into worker and firm components using a non-additive model, though without an explicit occupational dimension. The common thread is that occupations matter for sorting—but no paper in this literature has combined discrete latent types for both workers and firms with observed occupations in a unified classification framework, nor has any paper decomposed the relative contributions of these three dimensions to both wages and lifetime value.

This paper fills that gap. We observe job spells for worker i with gender f at firm j in occupation k, and recover latent worker types x and firm types y via a Classification EM algorithm. A *job* is the pair (y, k)—the firm type and occupation. A worker's *labor-market state* is the combination (x, y, k), conditional on gender f and age. Workers are free to move between any firm type and any occupation—including cross-track moves between production and administrative roles. Using the French DADS administrative panel (2015–2019, 18 million workers, 1.6 million firms, 27 million spells), we classify workers into $X = 3$ latent types and firms into $Y = 12$ latent types, while taking the $K = 8$ occupational categories as observed. Gender and age condition all model parameters but are not themselves objects of sorting—they are worker attributes that shape wages, hazards, and transition probabilities within each (x, y, k) cell. The CEM likelihood incorporates wages, separation hazards, job-to-job transitions, and occupation switches in a single unified framework.

Our estimation strategy follows the CEM architecture introduced by Lentz et al. (2023), but differs in an important respect. Where Lentz et al. (2023) embed the classification within a structural search framework—with an explicit meeting technology, random-utility acceptance probabilities, and on-the-job search—we estimate a flexible reduced-form specification that imposes no behavioral assumptions on how workers choose among offers or how matches form. Lentz et al. (2023) explicitly note that their framework is not an equilibrium model: wages are taken from the data rather than solved for endogenously, and the meeting technology is estimated rather than derived from primitives. Our approach goes further by also dropping the structural search assumptions—the latent types are identified purely from the joint distribution of wages, hazards, transitions, and occupation switches observed in the data, without imposing a value function or an acceptance rule. This reduced-form approach sacrifices the ability to decompose mobility into contact rates and acceptance probabilities, but it gains transparency: the estimated types and transition probabilities are sufficient statistics for the questions we ask, and they do not depend on assumptions about how workers evaluate offers.

After estimation, we compute the net present value (NPV) of lifetime earnings via backward induction on the estimated Markov chain. The matrix-based Bellman inversion delivers the exact expected present value for all 104 state cells, in closed form. We then decompose each worker's NPV into a

wage component—the capitalized value of staying in the current job forever—and an *option-value component*—the expected gain from future job mobility. This decomposition provides the test: if workers are forward-looking, destination choices should respond to both components; if workers are myopic, only the wage component should matter.

To attribute the relative contributions of worker type, firm type, occupation, and gender to wage and NPV variation, we introduce three complementary decomposition methods to the sorting literature. The first, Shapley variance decomposition, averages each factor's marginal contribution to between-cell variance across all possible orderings, yielding an exact, order-independent attribution that sums to the total (Shapley, 1953). The second, additive main-effects decomposition, extends the Abowd et al. (1999a) wage equation to four dimensions and separates total variance into pure-heterogeneity terms, pairwise sorting covariances, and a non-additive interaction residual. The third, Shapley-valued mutual information decomposition, replaces variance with Shannon entropy as the target, capturing the full distributional structure rather than just cell means, and revealing synergy and redundancy among factors. To our knowledge, this is the first application of all three methods—and in particular the MI synergy–redundancy analysis—to the sorting literature.

Four empirical results define the contribution. **First**, occupation is the single largest driver of both current wages and lifetime value across all decomposition methods, and the dominant sorting channel runs between firms and occupations ($\rho = 0.201$, 7.1% of wage variance), not between firms and workers ($\rho = -0.065$). This reframes a central puzzle in the AKM literature: the difficulty of signing the worker–firm sorting covariance may reflect not only limited-mobility bias but the fact that the primary assortative match operates along a dimension—the firm–occupation pairing—that two-way models do not represent.

Second, the MI decomposition reveals a qualitative shift in the structure of sorting between wages and lifetime value. For wages, the contributions of worker type, firm type, occupation, and gender are approximately additive (net synergy +1.5 pp): each factor's information about wages barely depends on whether the other factors are known. For NPV, all four factors are synergistic, with 20.2 percentage points of NPV entropy existing only in the interaction among factors and not recoverable from any factor in isolation. This pervasive complementarity reflects the forward-looking structure of lifetime value: NPV integrates over transition probabilities, separation hazards, and occupational switching rates that depend on the full (x, y, k) state jointly. The within-cell wage residual (36.3% of variance) captures forces—tenure, match quality, bargaining—that are large for current pay but average out over a career of transitions governed by the discrete classification. This asymmetry has a direct implication for model selection: two-way sorting models appear adequate when evaluated on wage fit, because the missing dimensions contribute additively; for lifetime value, omitting dimensions destroys synergistic information that cannot be recovered.

Third, gender operates primarily through non-additive complementarities rather than through its own main effect. The non-additive interaction term is 16.9% of wage variance, of which roughly 11

percentage points involve gender interacting with the (x, y, k) cell. Gender–occupation sorting is zero in the additive model ($\rho = -0.002$): the gender gap is not about women being in low-premium occupations but about cell-specific premia that a single gender coefficient cannot capture.

Fourth, workers' destination choices respond to the capitalized current-wage component of lifetime value but not to the option-value component. In structural log-share regressions with origin fixed effects, the wage component loads at $\hat{\beta} = 0.112$ (SE 0.031); the option component is indistinguishable from zero ($\hat{\beta} = 0.002$, SE 0.034). When facing a wage–option tradeoff, workers systematically take the wage: 36% of all job-to-job moves are *harvest* moves (wage up, option value down). Observed mobility responds to immediate wage signals but not to the career-building value that the model reveals is present.

The remainder of the chapter is organized as follows. Section 2.1 describes the data and the methodology. Section 2.2 presents the CEM model, the estimation algorithm, and the NPV backward induction. Section 2.3 introduces the variance and information decomposition framework. Section 2.5 reports the sorting and decomposition results, and Section 2.6 concludes.

2.1 Data and Methodology

2.1.1 Data Description

We use the DADS (Déclarations Annuelles de Données Sociales) Postes administrative dataset, which covers the universe of salaried workers in France for the years 2015–2019. Building the analysis sample from the raw DADS requires three stages: (i) constructing a consistent worker identifier across annual files and chaining workers over years, (ii) assembling yearly records into a single panel, and (iii) cleaning and collapsing the panel into worker-spells, and recoding occupations. The details of the data construction can be found in the appendix section A2.1. Gross wages are reported for each spell in the data along with the duration of employment (in days) for a given calendar year. We deflate the gross wages by the CPI published by INSEE with the year 2015 as base[1]. The primary measure of earnings used in this analysis is the log of daily wage which is constructed by dividing gross earnings by the number of days of employment. Occupations are observed as detailed 4-digit PCS codes and are aggregated into 8 categories [2] tabulated below with related summary statistics in table 2.1.

A spell is a contiguous period during which a worker remains in the same (firm, occupation) pair. A new spell begins whenever a worker changes employer, changes occupation, or transitions into or out of non-employment. Time between two (firm, occupation) pairs is treated as non-employment. We construct a worker-spell panel where each row records the worker's identity, firm, occupation, total spell duration, age at spell start, and gender. Worker identifiers are indexed by $i \in (1, ..., N)$, firm identifiers (`SIREN`) by $j \in (1, ..., J)$ and occupations by $k \in (1, ..., 8)$. After removing records with

[1] CPI deflator by INSEE: https://www.insee.fr/fr/statistiques/serie/001763852
[2] We thank Francis Kramarz for this classification of occupations.

missing or inconsistent identifiers and spells that cannot be unambiguously ordered, the estimation sample contains about 18 million workers, 1.6 million firms, and 27 million spells. With the above information, we can define stayers and movers: *stayers* are individuals who remain at the same firm j throughout the observed period, though they may change occupation k within that firm, and *movers* are those who transition to a new firm $j \to j'$ with the same or a different k.

Table 2.1. Summary Statistics by Occupational Category

k	Role	Unique workers	Mean daily wage	Mean duration	Share female
Production occupations					
1	Low-skilled manual worker	2,041,211	3.86	665	33
2	Skilled manual worker	4,348,542	4.24	1,018	17
3	Foremen and technicians	1,552,660	4.56	962	18
4	Engineers and technical management	1,370,424	5.00	925	23
Administrative occupations					
5	Unskilled administration, sales and service employees	2,629,432	3.85	742	67
6	Skilled administration, sales and service employees	4,856,927	4.13	929	71
7	Administrative lower management	3,430,665	4.37	993	70
8	Professionals and management	2,714,077	4.87	967	50

Note: This table presents descriptive statistics for eight occupational categories ($k = 1 \ldots 8$). "Unique workers" represents the total count of individuals, "Mean daily wage" refers to the average log of deflated gross daily wages, "Mean duration" is expressed in number of days and "Share female" is expressed as a percentage of unique workers in that cell. Note that the maximum possible spell length in days over the years studied is 1800 days.

About 79% of workers are stayers, i.e., they remain at the same firm throughout the data. Of these stayers, roughly two-thirds—67% of all workers—are observed in a single uninterrupted spell, never changing firm or occupation, while the remaining 12% change occupation at least once within the same firm, together adding to the 79% of workers who are stayers. The other 21% of workers are movers who change employer at least once during the five-year window. Across workers and spells, there are approximately 9.4 million transitions observed in total, i.e., a change of firm, of occupation, or both, including transitions into and out of non-employment. These transitions decompose into three broad categories: within-firm occupation changes (about 33% of all transitions), employer-to-employer moves excluding non-employment (about 26%), and transitions to or from non-employment[3] (about 42%). Of employer-to-employer moves, roughly three-quarters involve no change in occupation k. These same-occupation firm changes are broadly distributed, with $k = 7$ (administrative lower management), $k = 6$ (skilled administration, sales, and service employees), and $k = 8$ (professionals and management) each accounting for 17–21% of the total.

[3]Non-employment transitions are heavily concentrated by firm type defined in section 2.1.2. Firm types 8 and 12—the two dominant firm types by spell share that emerge from the CEM classification—account for about 66% of all separations into non-employment and 72% of all hires out of non-employment. This concentration underscores the role of these firm types in governing both the level and the composition of labor market flows.

Wage changes differ systematically across transition types. Within-firm occupation changes carry a mean log wage gain of +0.069, employer-to-employer moves in the same occupation yield +0.035, while joint firm-and-occupation changes yield only +0.019 with substantially greater dispersion (interquartile range from −0.20 to +0.28). Movers are also younger than within-firm occupation changers: the mean age at employer-to-employer transitions is 39 compared to 42 for within-firm changes. Workers who change both firm and occupation simultaneously are the youngest, with a mean age of 34. The share of women among employer changers (50%) is also higher than among within-firm occupation changers (44%), reflecting the strong gender segmentation across occupational categories documented in table 2.1.

Occupational mobility is also observed: 64% of all occupation-change spells are within the same firm. The remaining 36% is with a change of firm or transitions out of non-employment. The most common changes of occupation are the moves k=6 $\longrightarrow$ 7 (skilled administration, sales, and service employment to administrative lower management) and k=5 $\longrightarrow$ 6 (unskilled to skilled administration, sales, and service employment). Among production occupations, the most common change is k=1 $\longrightarrow$ 2 (low-skilled to skilled manual worker).

2.1.2 Methodology

2.1.2.1 Structure and Model Setup

We model a labor market populated by I workers and J firms observed over a panel of discrete spells. Each worker i is observed for T_i consecutive spells indexed $t = 1, \ldots, T_i$. A spell is a contiguous period during which the worker remains at the same employer in the same occupation; a new spell begins whenever the worker changes employer, changes occupation, or transitions into or out of non-employment.

Workers are characterized by three layers of heterogeneity. First, each worker carries a *time-invariant latent type* $x_i \in \{1, \ldots, X\}$ that is unobserved to the econometrician but fixed over the worker's career. This type captures persistent differences in wage levels, separation behavior, and mobility patterns that are not explained by observables. Second, each worker has an observed, time-invariant gender $fem_i \in \{0, 1\}$. Third, each worker's age enters continuously through AGE_{it} and discretely through an age bin $age_{it} \in \{1, \ldots, A\}$ constructed as population quintiles.

Firms are characterized by a *time-invariant latent type* $y_j \in \{1, \ldots, Y\}$, also unobserved. All workers employed at firm j at any point in time share the same firm type y_j. Non-employment is coded as $y = 0$ and is treated as an absorbing outside option with its own transition structure. Occupations are directly observed. Each spell carries an occupation code $k_{it} \in \{1, \ldots, K\}$ corresponding to 8 categories described in table 2.1. The state of worker i in spell t is therefore the tuple $(x_i, fem_i, age_{it}, y_{it}, k_{it})$, where $y_{it} = y_{j_{it}}$ is determined by the worker's current employer j_{it}.

2.1.2.2 Likelihood

The complete-data likelihood for worker i, conditional on the latent types $(x_i, y_{1:J})$, decomposes into six components: an initial condition, a wage process, a survival process, a separation event, a destination process, and an occupation transition process.

$$
\begin{aligned}
L_i(x_i, y_{1:J}) = {} & \underbrace{\pi(x_i \mid fem_i)}_{\text{type prevalence}} \cdot \underbrace{P_1(y_{i1}, k_{i1} \mid x_i, fem_i, age_{i1})}_{\text{initial match}} \cdot \underbrace{q(j_{i1} \mid y_{i1})}_{\text{initial firm draw}} \\
& \times \prod_{t=1}^{T_i} \underbrace{f(w_{it} \mid x_i, y_{it}, k_{it}, fem_i, age_{it})}_{\text{wage density}} \\
& \times \prod_{t=1}^{T_i} \underbrace{\exp\!\left[-P_\emptyset(x_i, y_{it}, k_{it}, fem_i, age_{it})\; D_{it}\right]}_{\text{survival through spell}} \times \prod_{t=1}^{T_i-1} \underbrace{P_\emptyset(x_i, y_{it}, k_{it}, fem_i, age_{it})}_{\text{separation event}} \\
& \times \prod_{t=1}^{T_i-1} \Bigg[\underbrace{P_M(y_{i,t+1} \mid x_i, y_{it}, k_{it}, fem_i, age_{it})}_{\text{destination type}} \cdot \underbrace{q(j_{i,t+1} \mid y_{i,t+1})}_{\text{destination firm draw}} \Bigg]^{M_{it}} \\
& \qquad \times \underbrace{P_{\neg M}(y \mid x_i, y_{it}, k_{it}, fem_i, age_{it})^{1-M_{it}}}_{\text{stayer probability}} \\
& \times \prod_{t=1}^{T_i-1} \underbrace{P_{C\neg M}(k_{i,t+1} \mid x_i, y_{it}, k_{it}, fem_i, age_{it})^{1-M_{it}}}_{\text{occupation transition (stayer)}} \cdot \underbrace{P_{CM}(k_{i,t+1} \mid x_i, y_{i,t+1}, k_{it}, fem_i, age_{it})^{M_{it}}}_{\text{occupation transition (mover)}}
\end{aligned}
\tag{2.1.1}
$$

Each component is described in turn below.

Type prevalence: $\pi(x \mid fem)$**.** The probability that worker i is of latent type x, conditional on gender. This allows the distribution of latent types to differ between men and women; for instance, if one type is characterized by frequent job changes, it may be more prevalent among men.

Initial match: $P_1(y, k \mid x, fem, age)$. The probability that a worker of type $(x, fem, agebin)$ is first observed in a match with a type-y firm in occupation k.

Firm draw: $q(j \mid y) = 1/N_y$. Conditional on the firm type y, the specific firm j is drawn uniformly from among the N_y firms of that type. This term appears both at the initial match ($t = 1$) and at every job-to-job transition. It plays an important role in estimation: concentrating too many firms into a single type increases N_y and reduces $q(j \mid y) = 1/N_y$, creating a penalty in the likelihood that pushes against degenerate classifications. For non-employment, $q(0 \mid 0) = 1$.

Wage density: $f(w \mid x, y, k, fem, age)$**.** Wages are modeled as log-normal with cell-specific means and common age effects:

$$w_{it} = \mu_{x_i, y_{it}, k_{it}, fem_i} + \beta_1 \cdot age_{it} + \beta_2 \cdot age_{it}^2 + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}(0, \sigma^2) \tag{2.1.2}$$

where $\mu_{x,y,k,fem}$ is the cell-specific intercept (estimated as the within-cell mean of age-adjusted log wages), β_1 and β_2 are common age–wage slopes estimated by within-cell demeaned OLS, and σ is the common residual standard deviation. During non-employment ($y = 0$), wages are missing and the wage term contributes 1 to the likelihood.

The cell means $\mu_{x,y,k,fem}$ form a four-dimensional array of $X \times (Y + 1) \times K \times fem$ parameters. These means absorb all between-cell wage variation attributable to the interaction of worker type, firm type, occupation, and gender. The residual σ captures within-cell dispersion due to unobserved individual heterogeneity, tenure, match quality, and measurement error.

Hazard (separation rate): $P_\emptyset(x, y, k, fem, age)$**.** The daily probability that the current spell ends follows a proportional hazard specification:

$$P_\emptyset(x, y, k, fem, age) = h_{x,y,k,fem} \cdot \exp\left(\gamma_1(age - \overline{age}) + \gamma_2(age - \overline{age})^2\right) \tag{2.1.3}$$

where $h_{x,y,k,fem}$ is the cell-specific baseline hazard and γ_1, γ_2 are common age–hazard slopes estimated by Newton–Raphson. The continuous-time approximation gives survival probability $\exp[-P_\emptyset \cdot D_{it}]$ over spell duration D_{it} days. The last spell of each worker ($t = T_i$) is right-censored: it contributes the survival term but not the separation event, since we do not observe when (or whether) it ends.

Destination type: $P_M(y' \mid x, y, k, fem, age)$. Conditional on separation ($F_{it} = 1$), the worker's destination is drawn from a multinomial over $Y + 2$ outcomes: remain at the current firm with an occupation change ("stayer," probability $P_{\neg M}$), transition to non-employment ($y' = 0$), or move to a firm of type $y' \in \{1, \ldots, Y\}$. In the implementation, stayer and mover outcomes are unified into a single multinomial P_M indexed by a destination code. When the worker moves ($M_{it} = 1$), the specific destination firm j' is drawn uniformly from among all type-y' firms via $q(j' \mid y')$.

Occupation transitions. Conditional on separation, the worker's next occupation $k' \in \{1, \ldots, K\}$ is drawn from one of two distributions depending on whether the worker changed employer:

- *Stayers* ($M_{it} = 0$): $P_{C\neg M}(k' \mid x, y, k, fem, age)$ conditions on the *current* firm type y. The outcome space is $k' \neq k$ (the worker changed occupation but stayed at the same firm).
- *Movers* ($M_{it} = 1$): $P_{CM}(k' \mid x, y', k, fem, age)$ conditions on the *destination* firm type y'. The outcome space is all K occupations including $k' = k$.

This distinction captures the fact that within-firm occupation changes and between-firm occupation changes may follow different patterns—for instance, within-firm promotions (stayer, $k \to k + 1$) versus cross-firm lateral moves (mover, $k \to k$).

A recap of dimensions and model parameters are listed in the appendix A2.2.1.

2.2 Estimation

2.2.1 Classification Expectation-Maximization

We estimate the model by Classification-EM (CEM), a variant of EM in which the E-step assigns each worker to a latent type and the C-step reclassifies firms to latent types using the assigned worker types. The latent objects are the worker types $x_{1:I}$ and the firm types $y_{1:J}$. Occupations, gender, age, wages, spell durations, and mobility outcomes are all observed. Each iteration proceeds in three steps. Given current type assignments $(x_{1:I}, y_{1:J})$:

M-step (parameters updation). All model parameters are updated by maximum likelihood given current type assignments. This includes the wage means and age slopes, the residual standard deviation, the separation hazards, the worker-type prior, the initial matching distribution, the firm sampling probability, the destination-type distribution conditional on separation, and the occupation transition distributions for stayers and movers. The wage and hazard age slopes are estimated by within-cell pooled OLS and Newton–Raphson on the Poisson likelihood respectively. All multinomial parameters are estimated by cell-specific frequency counts with additive (Laplace) smoothing.

E-step (worker classification). Given parameters Θ and current firm types $y_{1:J}$, each worker i is reassigned to the type x that maximizes their total log-likelihood:

$$x_i \leftarrow \arg\max_{x' \in \{1,\dots,X\}} \sum_{t=1}^{T_i} \ell_{it}(x', y_{j_{it}}) \tag{2.2.1}$$

where ℓ_{it} denotes the per-spell log-likelihood contribution from equation (2.1.1). The assignment for worker i is applied to all of i's spells simultaneously (the worker type is time-invariant).

C-step (firm classification). Given parameters Θ and current worker types $x_{1:I}$, each firm j is reassigned to the type y that maximizes the total log-likelihood of all spells at firm j:

$$y_j \leftarrow \arg\max_{y' \in \{1,\dots,Y\}} \sum_{\{(i,t):\, j_{it}=j\}} \ell_{it}(x_i, y') \tag{2.2.2}$$

The assignment for firm j is applied to all workers currently or previously employed at j (the firm type is time-invariant). Full closed-form expressions for each M-step block, together with the exact numerical procedures used for the iterative wage and hazard sub-routines, are provided in Appendix A2.2.2.

2.2.2 Implementation

Applying CEM to matched employer–employee data with 18 million workers and 1.6 million firms presents specific computational challenges related to initialization, convergence speed, and the extreme heterogeneity in firm sizes. We address each in turn. Throughout, our implementation choices favour closed-form, vectorized M-step updates over parametric estimators that would require nested numerical optimization at every CEM iteration. At our sample size, fitting a multinomial logit

(or analogous parametric model) over the full set of (*x, y, k, fem, age*) cells with K-way destination outcomes, repeated several hundred times across the cooling schedule, would be computationally infeasible within reasonable wall-clock budgets. Closed-form updates reduce each M-step block to a single accumulator pass, exploiting the columnwise vectorization of Mata. This is feasible because most cells in our data are dense—the average origin cell contains roughly 1,900 spells—and parametric pooling would offer minimal variance reduction relative to the bias risk it would introduce. Sparse cells are stabilized through Laplace smoothing.

2.2.2.1 Initialization

Random initialization of firm types leads to poor starting conditions: all Y types begin with similar wage distributions (each receives a random $1/Y$ slice of firms), so the first M-step estimates nearly identical cell means. The subsequent C-step has little signal to differentiate types, and the algorithm requires hundreds of iterations before a meaningful separation emerges.

We instead initialize firm types using a *size-stratified wage ranking*. Firms are separated into "large" (200 or more observed worker-spells) and "small" groups. The cutoff for "large" firms is set at 200 which includes 12,920 firms covering 48.8% (~50%) of all workers. The large firms are distributed to the Y classes in a serpentine manner. The largest firm goes to $Y = 12$, the second largest goes to $Y = 11, \ldots$ the 12^{th} largest goes to $Y = 1$, the 13^{th} largest goes to $Y = 1$, the 14^{th} largest goes to $Y = 2$, and so on. Within the "small" firms group, firms are ranked by their mean log wage and assigned to Y quantiles, so that each type receives approximately $1/Y$ of the small firms. This ensures that (i) the first M-step estimates well-separated cell means spanning the full wage distribution, and (ii) every type contains large firms that anchor its parameter estimates precisely. Worker types are initialized randomly, since worker types are identified primarily through mobility patterns rather than wage levels.

2.2.2.2 Simulated Annealing

The CEM objective is non-convex with many local optima. Pure arg max classification from a single initialization typically converges to a local optimum that depends on the starting point. To explore the likelihood surface more broadly, we replace the deterministic arg max in the E-step and C-step with temperature-scaled softmax sampling:

$$\Pr(x_i = x') \propto \exp\left(\frac{1}{\tau}\sum_{t=1}^{T_i} \ell_{it}(x', y_{j_{it}})\right), \qquad \Pr(y_j = y') \propto \exp\left(\frac{1}{\tau}\cdot \bar{\ell}_j(y')\right) \tag{2.2.3}$$

where $\tau > 0$ is the temperature and $\bar{\ell}_j(y')$ is the per-worker normalized log-likelihood defined below. At high temperature, assignments are nearly uniform; as $\tau \to 0$, the softmax converges to arg max. We use a geometric cooling schedule that decreases τ from a high initial value to a small terminal value, after which the algorithm switches to deterministic arg max for the remaining iterations. The geometric schedule spends more iterations at low temperatures, where small log-likelihood differences are most informative for fine-grained reassignments, while still permitting aggressive exploration in

the early phase. Specific values for the schedule parameters and the threshold for switching to arg max are reported in Appendix A2.2.2.

2.2.2.3 Per-Worker Normalization in the C-Step

A key challenge in firm classification arises from the extreme right skew of the firm size distribution. Under Zipf's law, a small number of mega-firms employ a disproportionate share of workers. A firm with $n_j = 100{,}000$ worker-spells accumulates a total log-likelihood on the order of $-500{,}000$, while a firm with $n_j = 3$ spells has a total log-likelihood around -15. The log-likelihood *gap* between the best and second-best type scales linearly with n_j: the mega-firm may have a gap of 5,000 while the small firm's gap is 1.5.

When this gap enters the softmax in equation (2.2.3), the effective temperature experienced by each firm is τ/Δ_j, where Δ_j is the firm's gap. For the mega-firm at $\tau = 0.5$: $\exp(5{,}000/0.5) = \exp(10{,}000) \approx \infty$—the softmax is deterministic regardless of τ. The mega-firm locks into its initially-assigned type from iteration 1, when parameters are still poorly estimated, and never explores alternatives. Meanwhile, small firms genuinely benefit from the stochastic exploration. In effect, the temperature schedule controls only small firms; large firms operate at effective $\tau = 0$ throughout.

To equalize the effect of temperature across firm sizes, we normalize the log-likelihood to a per-worker-spell average before applying the softmax:

$$\bar{\ell}_j(y') = \frac{1}{n_j} \sum_{\{(i,t):\, j_{it}=j\}} \ell_{it}(x_i, y') \tag{2.2.4}$$

The C-step then samples (or maximizes) based on $\bar{\ell}_j$ rather than the total. Since arg max is invariant to division by a positive constant, this normalization does not affect the arg max assignment: the type that maximizes the total also maximizes the average. The normalization only affects the stochastic phase, where it ensures that all firms—regardless of size—experience the same effective temperature and thus the same degree of exploration.

With normalization, the mega-firm's per-worker gap is $5{,}000/100{,}000 = 0.05$, and the softmax ratio becomes $\exp(0.05/0.5) = \exp(0.1) \approx 1.1$—genuine stochastic exploration. The small firm's per-worker gap is $1.5/3 = 0.5$, giving $\exp(0.5/0.5) = e \approx 2.7$. Both firms now have a meaningful probability of trying alternative types during the cooling phase.

2.2.2.4 Convergence

Convergence is monitored only during the deterministic (arg max) phase, since log-likelihood naturally fluctuates during stochastic annealing as workers and firms are reassigned probabilistically. We declare convergence when the absolute change in log-likelihood is small for several consecutive iterations. Given a total log-likelihood on the order of -450 million, the threshold corresponds to fewer than a

handful of effective worker reassignments per iteration, indicating that the classification has stabilized. Exact threshold and patience values are specified in Appendix A2.2.2.

2.2.3 Model Selection

We select the number of latent types (X, Y) by grid search over $X \in \{2, 3, 4\}$ and $Y \in \{5, 6, \ldots, 13\}$. For each (X, Y) pair, we run the full CEM algorithm and record the maximized log-likelihood. The best-fitting specification is $(X, Y) = (3, 12)$. The distribution of log-likelihoods across the grid of (X, Y) is shown in Figure 2.1.

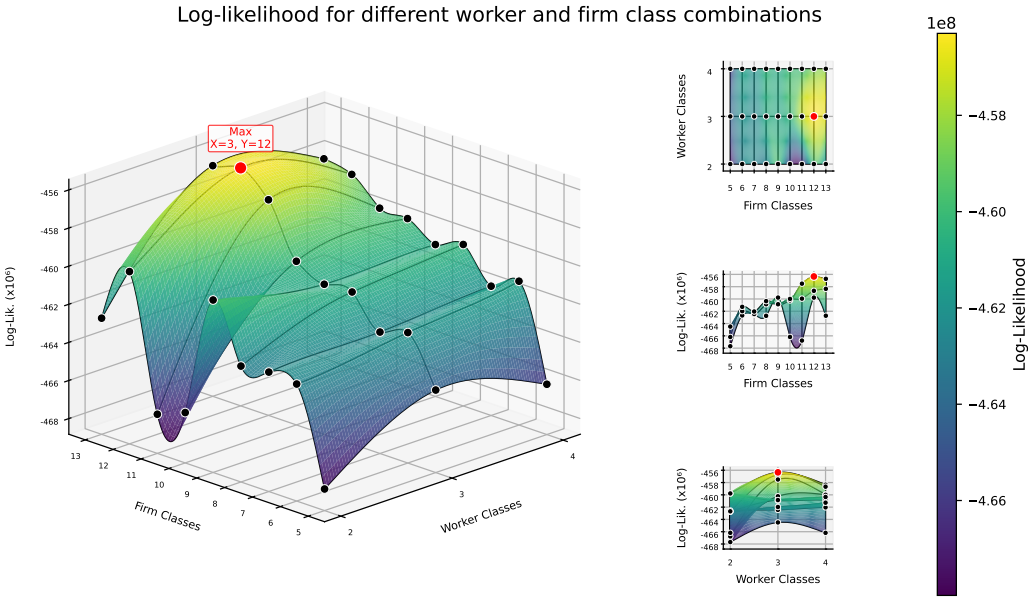


Figure 2.1. Log-likelihoods across the grid of (X, Y) pairs. The sub graphs look at the projections of the likelihood surface along the *xy* plane, *yz* plane, and *xz* plane.

2.2.4 Net Present Value Construction

A central object of interest is the net present value (NPV) of lifetime earnings associated with each labor market state. Rather than simulating individual career trajectories forward via Monte Carlo draws—which is computationally prohibitive given the size of the state space—we compute NPVs in closed form by exploiting the Markov structure of the estimated model. This matrix-based backward induction constitutes a methodological contribution: it delivers exact expected present values for every cell in the state space without sampling noise and at a fraction of the computational cost of simulation-based approaches.

State space. For a worker of type x and gender *fem*, the relevant state at any point in the career is the current job (y, k), where $y \in \{0, 1, \ldots, Y\}$ indexes firm type (with $y = 0$ denoting non-employment) and $k \in \{1, \ldots, K\}$ indexes occupation. The total number of states is $S = YK + K$: the first YK

states correspond to employment in each (firm type, occupation) pair, and the remaining K states correspond to non-employment with most recent occupation k.

Daily transition matrix. For each (x, fem, age) cell, we assemble an $S \times S$ daily transition matrix $\mathbf{Q}_{x,fem,age}$. From an employed state $s = (y, k)$, the worker separates with daily hazard $h_{x,y,k,fem,age}$ given by equation (2.1.3). Conditional on separation, the worker either stays at the same firm type but changes occupation (with probability $P_{\neg M} \cdot P_{C\neg M}$), moves to non-employment (with probability $P_M(\cdot \mid y' = 0)$), or transitions to a new firm type y' and occupation k' (with probability $P_M(\cdot \mid y') \cdot P_{CM}(k' \mid \cdot)$). With probability $1 - h_{x,y,k,fem,age}$, the worker remains in the same state. From a non-employment state $s = (0, k)$, the worker exits to employment at firm type y' in occupation k' at rate $h_{x,0,k,fem,age} \cdot P_M(y' \mid \cdot) \cdot P_{CM}(k' \mid \cdot)$, or remains non-employed. All transition probabilities are taken directly from the CEM parameter estimates. The rows of $\mathbf{Q}_{x,fem,age}$ sum to one by construction.

Backward induction. Let $\delta = e^{-r/360}$ denote the daily discount factor corresponding to an annual discount rate r. Let $\mathbf{w}_{x,fem,age}$ be the $S \times 1$ vector of daily wages in each state (with entries equal to zero for non-employment states), where the wage in state (y, k) is $\exp(\mu_{x,y,k,fem} + \beta_1 \cdot \overline{age}_A + \beta_2 \cdot \overline{age}_A^2)$ evaluated at the median age $\overline{age}_A$ of bin A. Let $p_A = 1/D_A$ denote the daily probability of ageing out of bin A, where D_A is the number of days spent in the bin (for the oldest bin, D_A equals the expected remaining working life until retirement).

The NPV vector $\mathbf{V}_{x,fem,age} \in \mathbb{R}^S$ satisfies the Bellman equation

$$\mathbf{V}_{x,fem,age} = \mathbf{w}_{x,fem,age} + \delta \left[(1 - p_{age})\, \mathbf{Q}_{x,fem,age}\, \mathbf{V}_{x,fem,age} + p_{age}\, \mathbf{Q}_{x,fem,age}\, \mathbf{V}_{x,fem,age+1} \right], \tag{2.2.5}$$

with the terminal condition $\mathbf{V}_{x,fem,5+1} = \mathbf{0}$ (retirement). Rearranging yields the closed-form solution

$$\mathbf{V}_{x,fem,age} = \left[\mathbf{I} - \delta(1 - p_{age})\, \mathbf{Q}_{x,fem,age} \right]^{-1} \left(\mathbf{w}_{x,fem,age} + \delta\, p_{age}\, \mathbf{Q}_{x,fem,age}\, \mathbf{V}_{x,fem,age+1} \right), \tag{2.2.6}$$

which we evaluate by backward recursion from $age = 5$ to $age = 1$. At each step, the computation reduces to solving a single $S \times S$ linear system. With $S = YK + K = 104$ states, each solve is instantaneous, and the full NPV table for all (x, fem, age) cells is computed in seconds.

Equivalence to simulated trajectories. The closed-form solution (2.2.6) is algebraically identical to the expected value one would obtain by drawing infinitely many forward trajectories from the estimated Markov chain. To see why, note that expanding the Bellman equation (2.2.5) forward yields

$$V_s = \mathbb{E}_s \left[\sum_{\tau=0}^{\infty} \delta^\tau\, w_{s_\tau} \right], \tag{2.2.7}$$

where the expectation is taken over all possible career paths $\{s_0 = s, s_1, s_2, \ldots\}$ generated by the transition matrix $\mathbf{Q}$ (with aging and retirement embedded). A simulation-based approach would approximate this expectation by drawing R independent trajectories, each realizing a random sequence of separations, firm-type draws, and occupation switches according to the estimated probabilities, and averaging the

discounted wage streams:

$$\hat{V}_s^{(R)} = \frac{1}{R} \sum_{r=1}^{R} \sum_{\tau=0}^{T_r} \delta^\tau \, w_{s_\tau^{(r)}}.$$

As $R \to \infty$, $\hat{V}_s^{(R)} \to V_s$. The matrix formula (2.2.6) computes this infinite-draw limit exactly in a single operation: the matrix inverse $[\mathbf{I} - \delta(1-p_a)\mathbf{Q}]^{-1} = \sum_{n=0}^{\infty} [\delta(1-p_a)\mathbf{Q}]^n$ sums over all possible multi-step paths and their probabilities simultaneously, weighting each path by its discounted wage flow. The result is the *exact* expected present value, computed for all S states at once via a single matrix inversion per age bin.

Decomposition. For each state s, the NPV V_s can be decomposed into a *wage component*—the discounted value of earnings in the current job absent any future transitions, $V_s^{\text{wage}} = w_s/[1 - \delta(1-h_s)]$—and an *option value* $V_s^{\text{opt}} = V_s - V_s^{\text{wage}}$, which captures the expected gains from future job mobility. This decomposition allows us to test whether observed transition flows are driven by immediate wage gains, by the value of future career options, or by both.

2.3 Measuring Sorting: Variance and Information Decomposition

We decompose the cross-sectional variation in wages and lifetime value using three complementary methods, each applied to the same four factors: worker type x, firm type y, occupation k, and gender f. The first method (Shapley variance decomposition) measures each factor's fair share of explained variance in cell means. The second (additive variance decomposition) separates that explained variance into pure heterogeneity, pairwise sorting covariances, and non-additive complementarities. The third (mutual information decomposition) measures how much each factor reduces uncertainty about the full wage or NPV distribution, not just its mean.

2.3.1 Shapley Variance Decomposition

For a target variable w and a subset of factors $S \subseteq \{x, y, k, f\}$, define the coalition value

$$V(S) = \text{Var}\big(\mathbb{E}[w \mid S]\big), \tag{2.3.1}$$

the between-cell variance when the population is partitioned by the factors in S. The Shapley value of factor j is

$$\phi_j^{\text{Var}} = \sum_{S \subseteq \mathcal{N} \setminus \{j\}} \frac{|S|!\,(N - |S| - 1)!}{N!} \Big[V(S \cup \{j\}) - V(S) \Big], \tag{2.3.2}$$

where $\mathcal{N} = \{x, y, k, f\}$, $N = 4$, and $V(\emptyset) = 0$. By the efficiency axiom, $\sum_j \phi_j^{\text{Var}} = V(\mathcal{N})$, the total between-cell variance. The within-cell residual $\text{Var}(w) - V(\mathcal{N})$ captures tenure, match quality, individual productivity, and measurement error.

The Shapley value answers: "What is each factor's fair share of explained variance?" But it bundles together each factor's own wage dispersion, its sorting covariances with other factors, and its share of the non-additive interaction. To separate these, we turn to the additive decomposition.

2.3.2 Additive Variance Decomposition

We regress w on main-effect dummies for all four factors:

$$w_{it} = \mu + \hat{\alpha}_{x_i} + \hat{\psi}_{y_{J(i,t)}} + \hat{\gamma}_{k_{it}} + \hat{\delta}_{f_i} + \varepsilon_{it}. \tag{2.3.3}$$

This is the four-dimensional analog of the Abowd et al. (1999a) wage equation, extended to include occupation and gender as separate additive components. After demeaning each fitted component, total variance decomposes exactly into twelve terms:

$$\mathrm{Var}(w) = \underbrace{\sum_j \mathrm{Var}(\hat{\beta}_j)}_{\text{4 heterogeneity}} + \underbrace{\sum_{j<m} 2\,\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_m)}_{\text{6 sorting}} + \underbrace{\mathrm{Var}(\mu_{xykf} - \hat{w}^{\mathrm{add}})}_{\text{interaction}} + \underbrace{\mathrm{Var}(w - \mu_{xykf})}_{\text{within-cell}}, \tag{2.3.4}$$

where $\mu_{xykf} = \mathbb{E}[w \mid x, y, k, f]$ is the full cell mean and $\hat{w}^{\mathrm{add}} = \mu + \hat{\alpha} + \hat{\psi} + \hat{\gamma} + \hat{\delta}$ is the additive prediction. Each covariance term is the direct analog of the AKM sorting covariance $2\,\mathrm{Cov}(\hat{\alpha}_i, \hat{\psi}_j)$, extended to all six factor pairs. The interaction term captures wage premia specific to particular (x, y, k, f) cells that cannot be decomposed into separate main effects—the non-additive complementarities. This term is zero if and only if wages are exactly additive in the four factors.

The two variance decompositions are related as follows. Each Shapley value approximately equals the factor's own variance plus half of each pairwise covariance it participates in, plus a share of the interaction term. But the Shapley uses the non-additive coalition values $V(S)$, not the additive regression, so the allocation of the interaction term is not exactly equal-split across factors—factors whose marginal contributions are amplified by the interaction (synergistic factors) receive a disproportionate share.

2.3.3 Mutual Information Decomposition

The variance decomposition measures how well the classification separates cell *means*. A classification can separate means well while leaving broad, overlapping conditional distributions that make individual wages hard to predict. Mutual information captures the full distributional structure.

Given a target variable w discretized into quantile bins and a set of categorical predictors S, the Shannon entropy of w is

$$H(w) = -\sum_b p(w{=}b)\log_2 p(w{=}b), \tag{2.3.5}$$

measured in bits. The mutual information between w and S is

$$I(S; w) = H(w) - H(w \mid S) = \sum_{s,b} p(s, b)\log_2 \frac{p(s, b)}{p(s)\,p(b)}. \tag{2.3.6}$$

$I(S; w)$ measures how much knowing S reduces residual uncertainty about w. Unlike the R^2 of a linear regression, mutual information captures arbitrary statistical dependencies—including non-linear and

non-monotone ones—without imposing functional form. It is invariant to invertible transformations: $I(S; w) = I(S; g(w))$ for any monotone g.

We compute Shapley values in the MI setting by replacing the coalition value $V(S)$ in equation (2.3.2) with $I(S; w)$:

$$\phi_j^{\mathrm{MI}} = \sum_{S \subseteq \mathcal{N} \setminus \{j\}} \frac{|S|!\,(N - |S| - 1)!}{N!} \Big[I(S \cup \{j\}; w) - I(S; w) \Big], \tag{2.3.7}$$

with $\mathcal{N} = \{x, y, k, f\}$, $N = 4$, and $I(\emptyset; w) = 0$. Efficiency gives $\sum_j \phi_j^{\mathrm{MI}} = I(x, y, k, f; w)$.

2.3.3.1 Four-factor Shapley weights

With $N = 4$ factors, the Shapley formula for each factor averages over eight marginal contributions—one for each of the $2^3 = 8$ coalitions not containing factor j—with weights determined by the coalition size:

- $|S| = 0$ (entering first): weight = $6/24 = 1/4$.
- $|S| = 1$ (entering second): weight = $2/24 = 1/12$, for each of three single-factor coalitions.
- $|S| = 2$ (entering third): weight = $2/24 = 1/12$, for each of three two-factor coalitions.
- $|S| = 3$ (entering last): weight = $6/24 = 1/4$.

The Shapley value for worker type x, for example, is

$$\begin{aligned} \phi_x = \frac{1}{24} \Big[& \; 6\, I(x; w) \\ & + \; 2\, I(x; w \mid y) + 2\, I(x; w \mid k) + 2\, I(x; w \mid f) \\ & + \; 2\, I(x; w \mid y, k) + 2\, I(x; w \mid y, f) + 2\, I(x; w \mid k, f) \\ & + \; 6\, I(x; w \mid y, k, f) \Big], \end{aligned} \tag{2.3.8}$$

where each conditional MI is computed by the chain rule, e.g., $I(x; w \mid y) = I(x, y; w) - I(y; w)$. Analogous expressions hold for ϕ_y, ϕ_k, and ϕ_f by relabeling.

2.3.3.2 Synergy and redundancy

Beyond the Shapley attribution, the full set of coalition values reveals whether factors carry *redundant* or *synergistic* information about the target. A factor whose marginal contribution *rises* as more factors enter the coalition is synergistic: the other factors make it more informative, indicating complementarity in how the dimensions interact. A factor whose marginal contribution *falls* is redundant: the other factors already capture part of its information. Comparing the unconditional MI $I(j; w)$ (entering first) with the full conditional $I(j; w \mid \mathcal{N} \setminus \{j\})$ (entering last) provides the sharpest test, but reporting the full profile of marginal contributions across all coalition sizes—as we do in the results tables—makes the pattern transparent.

At the aggregate level, the comparison between the sum of unconditional MIs $\sum_j I(j; w)$ and the joint MI $I(\mathcal{N}; w)$ summarizes the net balance of redundancy and synergy. If $\sum_j I(j; w) > I(\mathcal{N}; w)$, the factors collectively carry redundant information (overlapping signals). If $\sum_j I(j; w) < I(\mathcal{N}; w)$, the factors are net synergistic: their joint information exceeds the sum of their parts, and the gap measures information about the target that exists only in the interaction among factors and cannot be recovered from any factor in isolation.

2.3.4 Application

We apply all three decompositions—Shapley variance, additive variance, and Shapley MI—to two targets. For wages, we work with age-adjusted log wages $\tilde{w} = w - \beta_1 \text{AGE} - \beta_2 \text{AGE}^2$, where the common age polynomial is estimated alongside the cell means; for the MI decomposition, we discretize $\tilde{w}$ into twenty population quantiles. For lifetime value, we work with within-age-bin demeaned NPV $\text{NPV}_{\text{adj}} = \text{NPV} - \overline{\text{NPV}}_{\text{age}}$, which removes the mechanical decline in NPV from shortening career horizons; for the MI decomposition, we discretize NPV_{adj} into twenty within-age quantiles. The decomposition operates over four factors: worker type $x \in \{1, 2, 3\}$, firm type $y \in \{1, \ldots, 12\}$, occupation $k \in \{1, \ldots, 8\}$, and gender $f \in \{0, 1\}$. The first three are sorting dimensions—workers choose or are allocated to combinations of (x, y, k)—while gender is a conditioning attribute that shapes wages, hazards, and transitions within each (x, y, k) cell. All four enter the decomposition symmetrically to ensure complete attribution.

The combination of three decomposition methods on two targets, each with four factors, provides a comprehensive picture. The variance Shapley gives each factor's fair share of explained variance. The additive decomposition reveals how much of that is pure heterogeneity, how much is pairwise sorting, and how much requires non-additive complementarities. The MI Shapley measures distributional informativeness and exposes synergy and redundancy among factors. Comparing wages and NPV reveals how the classification's explanatory power changes when the target shifts from current pay to lifetime career value.

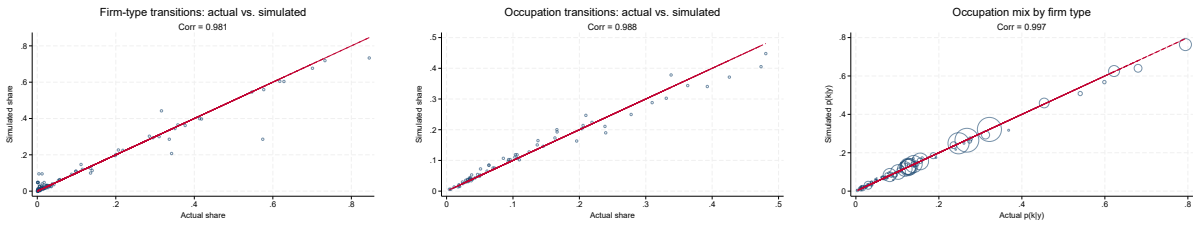
2.4 Model Fit

The CEM estimates a parametric model with lognormal wages, exponential separation hazards, occupation transition matrices, and non-employment exit rates, each indexed by the latent types (x, y, k). We evaluate whether the estimated model reproduces the key features of the data. The validation proceeds in three stages: we first assess whether the model's transition structure—the core of the identification argument—is replicated in simulation; we then evaluate the within-cell wage distributions; and finally we examine classification sharpness. Additional diagnostics including separation hazard fit, career-level comparisons, duration fit, and age-specific separation rates are reported in Appendix A2.3.3.

Simulation methodology: We simulate 50,000 career trajectories from the estimated model parameters. Each simulated worker is initialized by drawing a worker type x from the estimated marginal $\pi(x)$, an initial firm type y and occupation k from the estimated initial match distribution $P_1(y, k \mid x)$, and an initial age from the empirical age distribution. The worker then evolves through the estimated mobility structure: each employment spell has a duration drawn from the exponential distribution with separation hazard rate $h(x, y, k, \text{age})$; upon separation, the destination firm type is drawn from the estimated firm-type distribution $\lambda(y' \mid x, y, k)$ with a new occupation drawn from $P^{CM}(k' \mid x, y', k)$; conditional on remaining at the same firm, the occupation may change according to $P^{C\neg M}(k' \mid x, y, k)$. Workers who exit to non-employment draw a non-employment duration and re-enter employment through the estimated re-employment rates $\psi(y, k \mid x)$. All wages are drawn from the cell-specific lognormal distributions $\mathcal{N}(\mu_{x,y,k}, \sigma^2)$. The simulation uses the same 5-year panel structure as the data.

2.4.1 Transition Structure

The model's ability to reproduce worker mobility patterns is the central validation, since the CEM classification is identified jointly from wages and transitions. We compare simulated and actual transition shares at multiple levels of aggregation (Figure 2.2).



(a) Firm-type transitions **(b)** Occupation transitions **(c)** Occupation mix $p(k \mid y)$

Figure 2.2. Simulated vs. actual transition shares. Each point represents an origin-destination pair (panels a–b) or a firm-type/occupation cell (panel c). The 45-degree line indicates perfect fit.

At the firm-type level, the correlation between actual and simulated destination shares is $\rho = 0.98$ across 143 origin-destination (y, y') pairs (Figure 2.2a). The model accurately captures which firm types exchange workers and at what rates. For occupation transitions, the correlation is $\rho = 0.99$ across 64 (k, k') pairs (Figure 2.2b), reproducing the concentration of transition flows along the dominant origin-destination pairs. When conditioning on both origin firm type and occupation, the fit remains strong: the correlation is $\rho = 0.96$ for $(y, k) \to y'$ flows (593 cells) and $\rho = 0.90$ for $(y, k) \to k'$ flows (536 cells). The lower correlation for occupation destinations conditional on (y, k) reflects that within-firm-type occupation switching patterns involve more interaction terms than the model's parameterization can fully capture; nevertheless, the overall structure is well-reproduced.

Two further tests confirm the model's fidelity. First, the within-firm-type occupation composition $p(k \mid y)$—an equilibrium object implied by the estimated transition matrices but not a targeted

moment in the likelihood—has a correlation of $\rho = 0.997$ between data and simulation (Figure 2.2c). The model reproduces which occupations concentrate in which firm types almost exactly. Second, the distribution of workers transitioning from non-employment to employment across (y, k) destination pairs has a correlation of $\rho = 0.996$ (see Appemdic A2.3.3.9), confirming that the estimated re-employment rates ψ accurately capture which firm-type/occupation combinations absorb workers exiting non-employment.

2.4.2 Within-Cell Wage Fit

The model assumes log-wages within each $(x, y, k,$ female) cell follow $\mathcal{N}(\mu_{x,y,k}, \sigma^2)$ with a common standard deviation σ across all cells. This is a strong restriction: it imposes that the entire cross-sectional wage distribution is generated by variation in cell means $\mu_{x,y,k}$, with residual dispersion constant. Figure 2.3 evaluates this assumption directly by plotting the empirical distribution of standardized residuals $(w - \mu_{x,y,k})/\sigma$ against the theoretical $\mathcal{N}(0, 1)$ for the four largest (x, y, k) cells, which together account for over 7.5 million spells. In all four cells, the empirical density closely tracks the normal benchmark through the center and into the tails, with no systematic skewness or excess kurtosis.

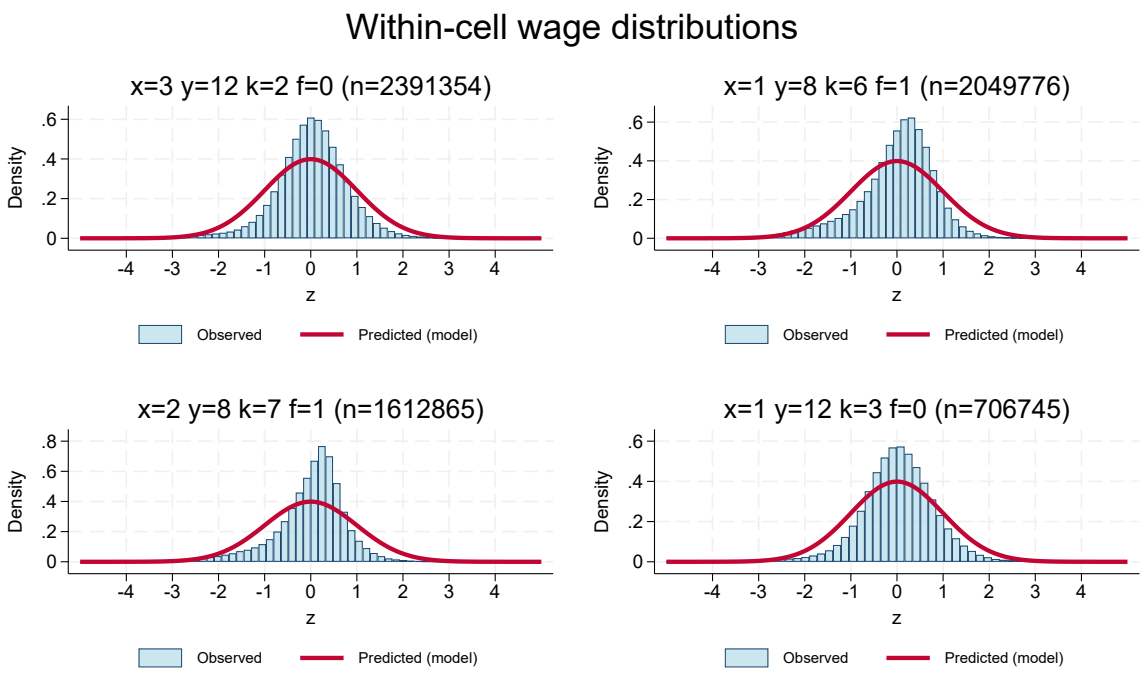


Figure 2.3. Standardized within-cell wage distributions for the four largest (x, y, k) cells, pooling across gender. Each panel plots the empirical density of standardized residuals $(w - \mu_{x,y,k})/\sigma$ against the model's predicted $\mathcal{N}(0, 1)$. Cell sizes range from 950,458 to 2,454,572 spells.

Across all 576 populated cells, the median ratio of empirical cell standard deviation to the model's σ is 1.01—indicating that the common-variance assumption is centered correctly—and 93% of cells have ratios within $[0.5, 2.0]$. The mean ratio is 1.12, reflecting a modest right tail: cells with below-average wages tend to have somewhat higher within-cell dispersion. This pattern—higher residual dispersion

in lower-wage cells—is a common feature of administrative earnings data and likely reflects unmodeled heterogeneity in hours or contract types that concentrates in low-wage matches. Importantly, the common-σ assumption affects only the relative weighting of observations in the likelihood, not the identification of cell means $\mu_{x,y,k}$ or transition parameters, which are the objects entering the sorting decomposition. This is because the cell means are identified by the first moment of the within-cell wage distribution, which is invariant to the scale of residual dispersion, and the transition parameters are identified entirely from mobility patterns. A cell with a true standard deviation 20% above σ contributes a slightly wider posterior over worker types, but the impact on aggregate classification and sorting measures is negligible given that 93% of cells fall within the [0.5, 2.0] range. [4]

2.4.3 Classification Sharpness

Firm classification uses the full likelihood—incorporating wages, separation hazards, and destination distributions—rather than wages alone. Classification accuracy increases sharply with firm size: from 45% for the smallest firms (1–2 spells) to 71% for firms with approximately 20 spells and 92% for the largest firms (over 3,000 spells; Figure 2.4). The full-likelihood classification substantially outperforms a wage-only classification (which achieves 25–63% across the same size bins), confirming that mobility patterns carry significant identifying information for firm types beyond what wages provide.[5]

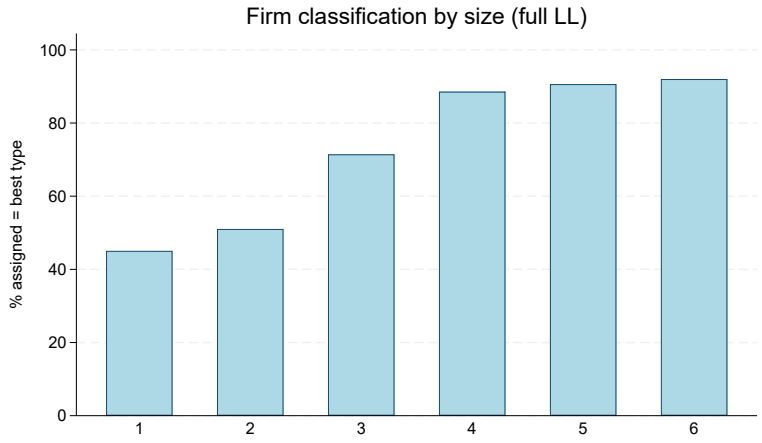


Figure 2.4. Firm classification accuracy by firm size. Accuracy is defined as the share of firms whose assigned type under the full likelihood (wages, hazards, and destinations) coincides with their best-fit type. Bins are defined by the number of observed spells at each firm.

Worker classification relies on posterior probabilities $p_i(x)$ computed from each worker's full trajectory. With $X = 3$ latent types over a 5-year panel, classification sharpness is moderate: 53% of workers have an implied maximum posterior around 0.79, while 33% have near-uniform posteriors (maximum ≈ 0.58), concentrated among single-spell workers. Importantly, the EM algorithm integrates over the

[4]Appendix A2.3.3.6 reports the full distribution of cell standard deviation ratios, a scatter of cell dispersion against cell mean wage, and within-cell wage distributions separately by gender, confirming that the gender differences in wages operate as within-cell mean shifts rather than shape differences.

[5]See Appendix table A2.3.3.8 for the full table of classification accuracy by size bin with mean log-likelihood gaps.

full posterior distribution at each iteration, so aggregate parameters and sorting measures are robust to individual-level classification uncertainty.[6]

2.5 Results

The estimated model recovers three latent worker types and twelve latent firm types given eight observed occupations, five age bins, and two genders. The sample comprises approximately 18 million workers, 1.6 million firms, and 27 million worker-spells. We organize the presentation in four parts. Section 2.5.1 describes the economic content of the estimated types. Section 2.5.2 decomposes wage and NPV variation into contributions from worker type, firm type, occupation, and gender. Section 2.5.2.3 studies how the association between these dimensions evolves over the life cycle and across genders. Section 2.5.3 examines observed transition patterns and tests whether workers choose destinations based on current wages, lifetime value, or the forward-looking option component of value.

2.5.1 Estimated Types

2.5.1.1 Worker types

Worker types are recovered by clustering on the full complement of observables: wages, spell durations, destination choices at separation, and occupation transitions. Table 2.2 reports the composition of each type.

Table 2.2. Worker Type Profiles

Type	Share	Mean log w	SD log w	Mean spell (days)	% female	Mean age bin	Dominant occupations
$x = 1$	30.3%	4.21	0.72	943	56.8%	3.97	$k = 6, 1, 3$
$x = 2$	39.8%	4.38	0.55	896	59.9%	4.05	$k = 7, 6, 5$
$x = 3$	29.9%	4.29	0.65	924	24.3%	4.07	$k = 2, 5, 1$

Worker types differ sharply in gender composition and occupation distribution, but much less sharply in mean wages (range 4.21 to 4.38). Types $x = 1$ and $x = 2$ are female-dominated (57% and 60% female respectively) and concentrate in administrative and service occupations. Type $x = 3$ is male-dominated (24% female) and concentrates in skilled manual production occupations ($k = 2$: 48% of $x = 3$ workers). The types are therefore not a simple skill or productivity hierarchy; they capture joint variation in occupation exposure and mobility behavior. This is consistent with the estimation using the full likelihood—including transitions and hazards—rather than wages alone.

We also compute the net present value of lifetime earnings for each cell in the state space defined by worker type x, firm type y, and occupation k. Figure 2.5 visualizes these NPVs in a heatmap, averaged across genders and age bins. The figure illustrates the complex interplay between worker type, firm type, and occupation in shaping lifetime value. For example, while type $x = 2$ has the highest mean

[6] See Appendix A2.3.3.7 for the full distribution of posterior entropy with approximate maximum posteriors by entropy bin.

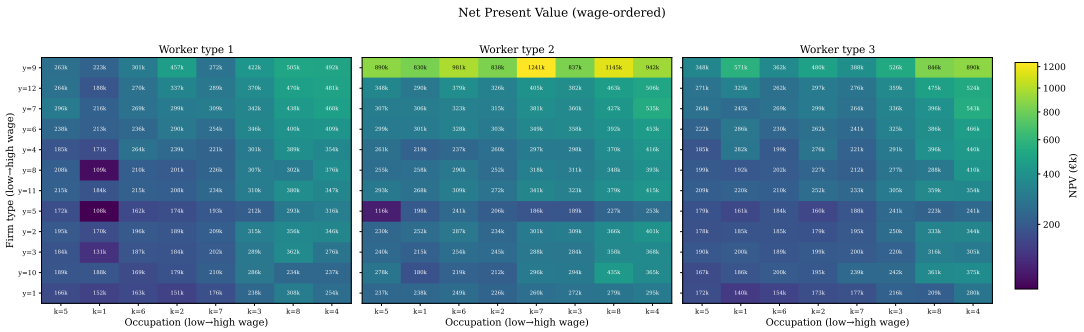


Figure 2.5. Net present value (NPV) of lifetime earnings by worker type x and firm type y and occupation k. Each cell reports the mean NPV for workers of type x employed at firms of type y in occupation k, averaged across genders and age bins.

wage, its NPV is not uniformly higher across all firm-occupation pairs; it depends on the specific match and the career options it affords. Similarly, certain occupations may have low current wages but high NPVs due to favorable mobility patterns or low separation hazards. The heatmap reveals these nuanced patterns that would be obscured in a one-dimensional wage ranking.

2.5.1.2 Firm types

Firm types are estimated from the joint distribution of wages, hazard rates, and mobility patterns. Table 2.3 reports key statistics, ordered by mean wage.

Table 2.3. Firm Type Profiles (ordered by mean log wage)

y	% spells	# firms	Mean size	Mean log w	SD log w	Sep. rate	% female	Mean age bin
1	1.1	81,730	3.8	2.86	1.06	0.64	63.0	3.96
10	0.4	40,838	2.6	3.31	0.63	0.71	46.5	3.00
3	3.7	106,763	9.6	3.64	0.60	0.63	49.3	3.95
2	3.4	117,452	8.2	3.65	0.60	0.80	63.3	3.76
5	0.2	34,840	1.6	3.73	0.39	0.16	38.8	4.08
11	3.9	112,546	9.8	4.06	0.50	0.89	28.4	3.33
8	34.0	497,956	19.1	4.17	0.51	0.62	68.2	4.01
4	0.3	33,835	2.3	4.26	0.41	0.67	38.6	3.85
6	3.7	86,259	11.9	4.36	0.49	0.92	50.6	4.16
7	1.7	55,026	8.6	4.48	0.38	0.72	32.1	4.02
12	40.0	441,320	25.4	4.58	0.54	0.40	31.2	4.08
9	0.6	22,845	7.8	5.40	0.85	0.58	42.7	3.93

Two firm types account for the majority of employment: $y = 8$ covers 34% of spells (mean log wage 4.17, separation rate 0.62), and $y = 12$ covers 40% of spells (mean log wage 4.58, separation rate 0.40). Type 12 is the canonical "good job"—high wages and low turnover—while type 8 has moderate wages and moderate turnover. Ten smaller firm types span the rest of the wage and hazard distribution. Wage means range from 2.86 (type 1) to 5.40 (type 9), a factor of over 12 in wage levels. Separation rates range from 0.16 (type 5: stable mid-wage firms) to 0.92 (type 6: extreme-churn high-wage firms).

Types 2 and 3 illustrate the multidimensional content of the classification. Both have nearly identical mean log wages (3.645 vs 3.644) but differ sharply in other dimensions: type 2 has 80% separation rate

and is 63% female (administrative and service firms), while type 3 has 63% separation rate and is 49% female (production firms). A wage-only classification would pool these two types; the multidimensional classification keeps them separate. Type 5 offers another instructive case: it has mid-range wages (3.73) but an extremely low separation rate (0.16), consistent with public-adjacent or heavily unionized firms where job security dominates.

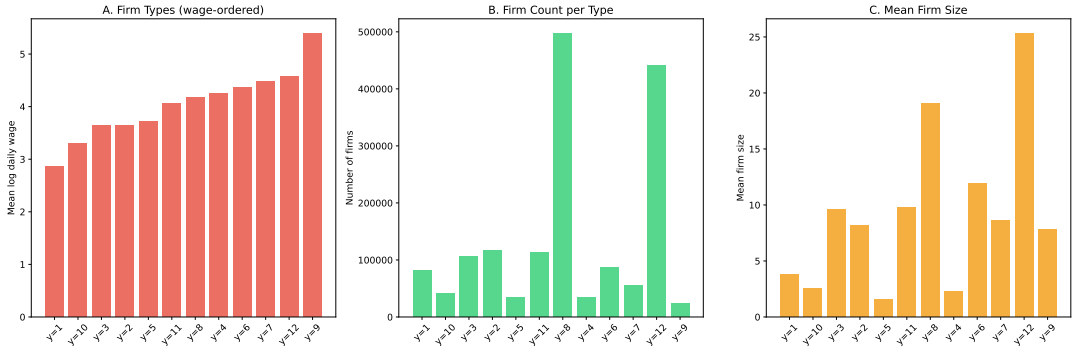


Figure 2.6. Firm type profiles. Panel A: mean daily wage (firm types ordered by wage). Panel B panels show number of firms in each firm type and Panel C plots the mean firm size by firm type.

Figure 2.6 provides a visual summary of this heterogeneity across the twelve estimated firm types. Panel A illustrates the explicit ordering of firm types by their mean log daily wage, displaying a clear progression from the lowest-paying firms (type $y = 1$) to the highest-paying firms (type $y = 9$). Beyond the wage structure, Panels B and C highlight the stark differences in firm types and labor market concentration. Panel B plots the total count of firms within each type, visually emphasizing the dominance of the two primary types, $y = 8$ and $y = 12$, which together make up the vast majority of the firm population. Finally, Panel C demonstrates that firm scale is highly heterogeneous. The high-retention firms of type $y = 12$ are significantly larger on average than all other types, driving their massive 40% share of overall employment. Conversely, both the highest-paying elite firms (type $y = 9$) and the lowest-paying firms (types $y = 1$ and $y = 10$) operate at much smaller average scales.

2.5.1.3 Occupations

Table 2.4 reports mean log wages, the female share, and the within-occupation composition of worker types for the eight occupations (type $x = 3$ is the omitted category). Occupations span a wage range comparable to firm types (3.85 to 5.00 log points), with Occupations 4 (engineers/technical management) and 8 (professionals/upper management) at the top, and Occupations 1 (low-skilled manual) and 5 (unskilled admin) at the bottom. Gender and worker-type composition vary sharply across occupations. Administrative and service occupations (5–7) are 67–71% female, whereas skilled manual (2) and technical management (4) are 17–23% female. Each occupation has a different propensity to be classified into the three worker types x.

Table 2.4. Occupation Profiles

k	Label	Mean log w	% female	% $x = 1$	% $x = 2$
1	Low-skilled manual	3.86	32.7	35.3	41.6
2	Skilled manual	4.24	17.2	7.0	18.4
3	Foremen/technicians	4.56	17.8	52.9	33.7
4	Engineers/tech. mgmt	5.00	22.5	12.2	60.5
5	Unskilled admin	3.85	66.8	25.7	50.2
6	Skilled admin	4.13	71.0	57.9	32.1
7	Admin lower mgmt	4.37	69.6	7.1	84.0
8	Professionals/upper mgmt	4.87	49.6	33.0	28.5

2.5.2 Variance and Information Decomposition

We decompose the cross-sectional variation in two outcomes—age-adjusted log wages $\tilde{w}$ and age-demeaned net present value NPV_{adj}—using three complementary methods, each applied to the same four factors: worker type x, firm type y, occupation k, and gender f. The first three are dimensions along which workers sort; gender is a conditioning attribute that enters the decomposition on equal footing to ensure complete attribution. The Shapley variance decomposition measures each factor's fair share of explained variance in cell means. The additive variance decomposition separates that explained variance into pure heterogeneity, pairwise sorting covariances, and non-additive complementarities. The mutual information (MI) decomposition measures how much each factor reduces uncertainty about the full distribution of the target, not just its mean, and reveals synergy and redundancy among factors.

2.5.2.1 Wage decomposition

The total variance of age-adjusted log wages is $\text{Var}(\tilde{w}) = 0.366$.

Shapley variance decomposition.

Table 2.5 reports the 4-factor Shapley decomposition.

Table 2.5. Wage Variance Shapley Decomposition

Component	Shapley value	% of $\text{Var}(\tilde{w})$
Occupation k	0.097	26.6%
Firm type y	0.084	23.0%
Worker type x	0.029	7.8%
Gender f	0.023	6.3%
Total between-cell	0.233	63.7%
Within-cell residual	0.133	36.3%

The (x, y, k, f) classification explains 63.7% of wage variance. Occupation is the single largest contributor (26.6%), followed by firm type (23.0%), worker type (7.8%), and gender (6.3%). The within-cell residual (36.3%) reflects unexplained wage variance which may include tenure, match quality, bargaining, and

measurement error which are not modeled here. The Shapley values are order-independent and sum exactly to the between-cell variance, but they bundle each factor's own wage dispersion, its sorting covariances, and its share of the non-additive interaction into a single number.

Additive variance decomposition.

To unbundle these, we regress $\tilde{w}$ on main-effect dummies for all four factors and decompose total variance into 12 terms that sum exactly to $\text{Var}(\tilde{w})$: 4 pure-heterogeneity variances, 6 pairwise sorting covariances, the variance of the non-additive interaction $\mu_{xykf} - \hat{\alpha} - \hat{\psi} - \hat{\gamma} - \hat{\delta}$, and the within-cell residual. Table 2.6 reports the results.

Table 2.6. Additive Wage Variance Decomposition

Component	Value	% of $\text{Var}(\tilde{w})$	ρ
Pure heterogeneity			
$\text{Var}(\hat{\gamma}_k)$ — occupation	0.065	17.8%	
$\text{Var}(\hat{\psi}_y)$ — firm type	0.064	17.4%	
$\text{Var}(\hat{\delta}_f)$ — gender	0.006	1.5%	
$\text{Var}(\hat{\alpha}_x)$ — worker type	0.005	1.4%	
Pairwise sorting			
$2\text{Cov}(\hat{\psi}, \hat{\gamma})$ — firm–occ.	0.026	7.1%	0.201
$2\text{Cov}(\hat{\psi}, \hat{\delta})$ — firm–gender	0.008	2.3%	0.221
$2\text{Cov}(\hat{\alpha}, \hat{\gamma})$ — worker–occ.	0.002	0.6%	0.056
$2\text{Cov}(\hat{\alpha}, \hat{\delta})$ — worker–gender	−0.003	−0.7%	−0.245
$2\text{Cov}(\hat{\alpha}, \hat{\psi})$ — worker–firm	−0.002	−0.7%	−0.065
$2\text{Cov}(\hat{\gamma}, \hat{\delta})$ — occ.–gender	−0.000	−0.0%	−0.002
Non-additive components			
Var(interaction)	0.062	16.9%	
Var(within-cell residual)	0.133	36.3%	
Var(additive)	0.171	46.7%	
Total between-cell	0.233	63.7%	

First, occupation and firm type have nearly identical own wage dispersion (17.8% vs. 17.4%), while worker type and gender are each small (~1.5%). The Shapley values for occupation and firm type exceed their pure-heterogeneity shares because each absorbs a proportional share of the pairwise covariances and the interaction term.

Second, the dominant sorting channel is firm–occupation: $\rho(\hat{\psi}, \hat{\gamma}) = 0.201$, accounting for 7.1% of total wage variance. High-wage firm types employ workers in high-wage occupations. This is the four-dimensional analog of the Abowd et al. (1999a) sorting covariance, but here the primary assortative match runs between firms and roles rather than between firms and workers. Worker–firm sorting is weakly negative ($\rho = -0.065$), consistent with the well-documented finding that worker–firm assortative matching is weak or absent in administrative data. This result reframes a central puzzle in the AKM literature. The difficulty of signing the worker–firm sorting covariance (Andrews et al., 2008; Kline et al., 2020; Babet et al., 2023) is typically attributed to limited-mobility bias. Our

estimates suggest a complementary explanation: the underlying correlation is genuinely near zero ($\rho = -0.065$), because the primary assortative match in this labor market runs between firms and occupations ($\rho = 0.201$), along a dimension that additive two-way models do not represent. The gender sorting terms reveal a composition effect: high-wage worker types are disproportionately male ($\rho(\hat{\alpha}, \hat{\delta}) = -0.245$), while gender–occupation sorting is essentially zero ($\rho = -0.002$). The near-zero gender–occupation correlation does not imply the absence of occupational segregation—it implies that gender segregation does not operate through the additive occupation premium $\hat{\gamma}_k$.

Third, the non-additive interaction is large: 16.9% of total wage variance. Of the 63.7% between-cell variance, 46.7% is additive and 16.9% requires cell-specific premia. Comparing to the three-factor decomposition (without gender), where the interaction was only 5.5%, reveals that 11.4 percentage points of complementarity involve gender interacting with the (x, y, k) cell. The wage differences by gender vary substantially across cells in ways that a single gender coefficient cannot capture. In our decomposition, 11.4 percentage points of wage variance arise from gender interacting non-additively with the (x, y, k) cell—a component that a single gender coefficient neither captures nor approximates.

To formalize this comparison, Appendix A2.4 estimates the Abowd et al. (1999b) two-way fixed-effect model on the same population, constructing a worker-year panel via dominant-employer assignment to maintain sample comparability. The AKM decomposition yields a negative sorting correlation ($\rho = -0.27$), more negative than typically reported in the French literature—a consequence of the data construction required for comparability with the CEM. Projecting the AKM effects onto the CEM classification reveals why the two frameworks diverge: seven occupation dummies explain 27% of the variance of AKM's 17 million worker fixed effects ($R^2 = 0.27$), while the CEM worker type adds less than 1% beyond occupation. What AKM attributes to persistent worker heterogeneity is predominantly occupation—the dimension along which the primary assortative match operates in our estimates. When AKM effects are aggregated to the CEM type level, the sorting correlation reverses sign to $\rho = +0.17$, consistent with the positive firm–occupation sorting documented above. The full comparison, including bias corrections, a three-way decomposition with occupation fixed effects, and a mover event study, is reported in Appendix A2.4.

Occupation as a distinct dimension.

To confirm that occupation is a genuine axis of wage variation rather than a proxy for worker or firm heterogeneity, Table 2.7 compares the additive decomposition across three specifications that hold the discrete-type resolution constant while varying the set of factors included.

Three results stand out. First, the worker-type share is essentially unchanged across specifications (1.4–2.2%), ruling out the interpretation that occupation is a proxy for worker heterogeneity. If occupation merely relabeled worker types, $\text{Var}(\hat{\alpha}_x)$ in the (x, y) model would absorb occupation's 17.8% and be correspondingly inflated. It is not.

Table 2.7. Additive Wage Decomposition: Effect of Adding Occupation

Component	(x, y)	(x, y, f)	(x, y, k, f)
Pure heterogeneity (% of Var($\tilde{w}$))			
Var($\hat{\alpha}_x$) — worker type	1.6	2.2	1.4
Var($\hat{\psi}_y$) — firm type	29.0	26.8	17.4
Var($\hat{\gamma}_k$) — occupation	—	—	17.8
Var($\hat{\partial}_f$) — gender	—	1.4	1.5
Structure (% of Var($\tilde{w}$))			
Var(additive)	29.9	31.0	46.7
Var(interaction)	1.7	6.5	16.9
Between-cell total	31.6	37.5	63.7
Within-cell	68.4	62.5	36.3

Notes. All values as % of Var($\tilde{w}$) = 0.366. The (x, y) specification is the discrete-type analog of Abowd et al. (1999a): a regression of wages on 3 worker-type dummies and 12 firm-type dummies. All three specifications use the same CEM-recovered latent types, varying only the included factors.

Second, the firm-type share drops by 11.6 percentage points—from 29.0% to 17.4%—when occupation is added. In the two-factor model, $\hat{\psi}_y$ captures both the firm-type wage premium and the occupational composition of each firm type. When occupation enters separately, the composition effect moves to Var($\hat{\gamma}_k$) and the firm–occupation sorting covariance (7.1%), leaving only the pure firm premium. This explains why two-way sorting models attribute a large share to firm heterogeneity: part of the "firm effect" is really a compositional effect of which occupations the firm houses.

Third, the within-cell residual drops from 68.4% to 36.3% when occupation is added—a fall of 32.1 percentage points. Most of occupation's contribution to wage variance was not misattributed to worker or firm effects in the two-factor model; it was simply missing, absorbed into the residual. The between-cell variance doubles from 31.6% to 63.7%, indicating that the two-factor classification leaves the majority of systematic wage variation unexplained. This result clarifies why Babet et al. (2023) find no significant contribution of occupation when added to a standard AKM specification with individual-level fixed effects: with 18 million worker parameters, $\hat{\alpha}_i$ has already absorbed occupation mechanically, leaving nothing for the occupation dummies to capture. At the CEM resolution of 3 worker types, this absorption is impossible, and occupation emerges as the single largest contributor to wage variation.

Mutual information decomposition.

Table 2.8 reports the 4-factor MI decomposition, showing each factor's unconditional MI $I(j; \tilde{w})$, its full conditional MI $I(j; \tilde{w} \mid \text{other 3})$, and its Shapley value.

The (x, y, k, f) classification reduces wage entropy by 17.2%—far less than the 63.7% of variance it explains. This gap is not a contradiction; it reflects the different metrics: variance decomposition measures separation of cell means, MI measures concentration of conditional distributions. The

Table 2.8. Wage Mutual Information Decomposition (% of $H(\tilde{w}) = 4.322$ bits)

Factor	Unconditional	Full conditional	Shapley ϕ_j
Occupation k	8.8%	7.2%	8.2%
Firm type y	5.8%	4.7%	5.2%
Worker type x	0.5%	3.3%	2.1%
Gender f	0.7%	2.3%	1.6%
$I(x, y, k, f; \tilde{w})$		17.2%	
$\sum_j I(j; \tilde{w})$		15.8%	
Net synergy		+1.5 pp	

within-cell wage dispersion creates broad, overlapping distributions that leave wage quantiles hard to predict even when the cell is known.

The synergy–redundancy pattern distinguishes the four factors sharply. Occupation is mildly redundant: its marginal contribution falls from 8.8% (entering alone) to 7.2% (entering last), indicating 1.6 pp of overlap with the other factors. This near-uniqueness is information-theoretic evidence that occupation constitutes a genuinely separate dimension of labor-market heterogeneity, not a proxy for worker ability or firm quality. If occupation merely reflected the skill component of worker type, its conditional MI would collapse once x is known; it does not. If it merely reflected firm-specific job ladders, its conditional MI would collapse once y is known; it does not. The additive decomposition confirms this: occupation's own variance (17.8%) is as large as firm type's (17.4%) and twelve times worker type's (1.4%). Firm type is similarly redundant (5.8% → 4.7%). Worker type is strongly synergistic: its contribution rises sixfold from 0.5% to 3.3%, meaning worker type becomes informative about wages only once the firm–occupation–gender context is known. Gender is likewise synergistic (0.7% → 2.3%), tripling in informativeness after conditioning—consistent with the additive decomposition's finding that gender's wage effects are predominantly non-additive.

2.5.2.2 NPV decomposition

We compute NPV by backward induction on the estimated Markov chain (Section 2.2.4). We decompose within-age-bin demeaned NPV using the same three methods. The total variance is $\text{Var}(\text{NPV}_{\text{adj}}) = 1.638$.

Shapley variance decomposition.

The contrast with wages is immediate. The classification explains 89.3% of NPV variance versus 63.7% for wages, and the within-cell residual shrinks from 36.3% to 10.7%. The ranking shifts: occupation dominates even more strongly (40.8% vs. 26.6%), firm type rises (29.5% vs. 23.0%), and gender nearly doubles its share (10.5% vs. 6.3%). Worker type remains the smallest contributor but rises from 7.8% to 8.4%. The classification captures far more of lifetime value than of current wages.

Additive variance decomposition.

Table 2.9. NPV Variance Shapley Decomposition

Component	Shapley value	% of Var(NPV)
Occupation k	0.668	40.8%
Firm type y	0.483	29.5%
Gender f	0.172	10.5%
Worker type x	0.138	8.4%
Total between-cell	1.462	89.3%
Within-cell residual	0.176	10.7%

Table 2.10. Additive NPV Variance Decomposition

Component	Value	% of Var(NPV)	ρ
Pure heterogeneity			
Var($\hat{\gamma}_k$) — occupation	0.573	35.0%	
Var($\hat{\psi}_y$) — firm type	0.318	19.4%	
Var($\hat{\delta}_f$) — gender	0.091	5.6%	
Var($\hat{\alpha}_x$) — worker type	0.023	1.4%	
Pairwise sorting			
2Cov($\hat{\psi}$, $\hat{\gamma}$) — firm–occ.	0.154	9.4%	0.181
2Cov($\hat{\psi}$, $\hat{\delta}$) — firm–gender	0.083	5.1%	0.245
2Cov($\hat{\alpha}$, $\hat{\gamma}$) — worker–occ.	0.018	1.1%	0.080
2Cov($\hat{\alpha}$, $\hat{\psi}$) — worker–firm	−0.018	−1.1%	−0.103
2Cov($\hat{\gamma}$, $\hat{\delta}$) — occ.–gender	−0.016	−1.0%	−0.035
2Cov($\hat{\alpha}$, $\hat{\delta}$) — worker–gender	−0.027	−1.7%	−0.294
Non-additive components			
Var(interaction)	0.261	15.9%	
Var(within-cell residual)	0.176	10.7%	
Var(additive)	1.201	73.3%	
Total between-cell	1.462	89.3%	

The NPV additive decomposition amplifies every pattern from the wage decomposition. Occupation's own variance share roughly doubles (35.0% vs. 17.8%) and firm type's rises mildly (19.4% vs. 17.4%). Gender's own variance nearly quadruples (5.6% vs. 1.5%), reflecting the compounding of gender-specific separation hazards and mobility rates over a career. Worker type remains small (1.4%).

The pairwise sorting structure is qualitatively similar to wages but larger in magnitude. Firm–occupation sorting rises from 7.1% to 9.4% ($\rho = 0.181$), and firm–gender sorting more than doubles from 2.3% to 5.1% ($\rho = 0.245$). Worker–gender sorting is the most negative term (−1.7%, $\rho = -0.294$), indicating that high-NPV worker types are even more disproportionately male when the outcome is lifetime value rather than current wages. Occupation–gender sorting, which was zero for wages, becomes slightly negative for NPV (−1.0%, $\rho = -0.035$): women are in marginally lower-$\hat{\gamma}$ occupations when measured by career value, even though they are not in lower-$\hat{\gamma}$ occupations when measured by current wage.

The interaction term is 15.9%—comparable in share to wages (16.9%) despite the much larger total variance. The additive model explains 73.3% of NPV variance, substantially more than the 46.7% for wages, indicating that the main effects are more powerful for lifetime value. The remaining 15.9% of between-cell variance still requires cell-specific premia, confirming that non-additive complementarities persist for career value.

Mutual information decomposition.

Table 2.11. NPV Mutual Information Decomposition (% of $H(\text{NPV}) = 4.280$ bits)

Factor	Unconditional	Full conditional	Shapley ϕ_j
Occupation k	24.3%	28.3%	27.7%
Firm type y	13.9%	25.9%	20.5%
Worker type x	7.8%	17.5%	14.2%
Gender f	4.8%	10.2%	8.7%
$I(x, y, k, f; \text{NPV})$		71.0%	
$\sum_j I(j; \text{NPV})$		50.8%	
Net synergy		+20.2 pp	

The (x, y, k, f) classification explains 71.0% of NPV entropy, compared to 17.2% for wages. Unlike the wage decomposition, where the gap between variance and MI shares was large (63.7% vs 17.2%), the NPV gap is small (89.3% vs 71.0%), reflecting the low within-cell residual (10.7%): the classification concentrates the conditional NPV distributions tightly, not just their means. However, the total explanatory power of the classification for NPV is partly mechanical—NPV is constructed from the (x, y, k, f)-indexed transition matrix, so within-cell NPV dispersion is small by construction within age bins. The more informative comparison is the *synergy structure*, which is not mechanical: +20.2 pp of net synergy for NPV versus +1.5 pp for wages. The sum of unconditional MIs (50.8%) falls 20.2 percentage points below the joint MI (71.0%). Over a fifth of NPV entropy is explained by information that exists only in the *interaction* among factors and cannot be recovered from any factor in isolation. The full NPV additive decomposition, with all 12 terms, is reported in Table A2.21 in the Appendix.

All four factors are synergistic. Every factor's full conditional MI exceeds its unconditional MI:

- Worker type: 7.8% → 17.5% (+9.7 pp) — more than doubles.
- Firm type: 13.9% → 25.9% (+12.0 pp) — nearly doubles.
- Occupation: 24.3% → 28.3% (+4.0 pp) — modest gain.
- Gender: 4.8% → 10.2% (+5.4 pp) — more than doubles.

For wages, occupation and firm type were *redundant*; for NPV, they are synergistic. The shift reflects the forward-looking structure of lifetime value: NPV integrates over transition probabilities, separation hazards, and occupational switching rates that are governed by the full (x, y, k, f) classification jointly. A worker type that is nearly uninformative about current wages on its own (0.5%) becomes highly informative about career value once the firm–occupation–gender context is known (17.5%).

Wage–NPV comparison.

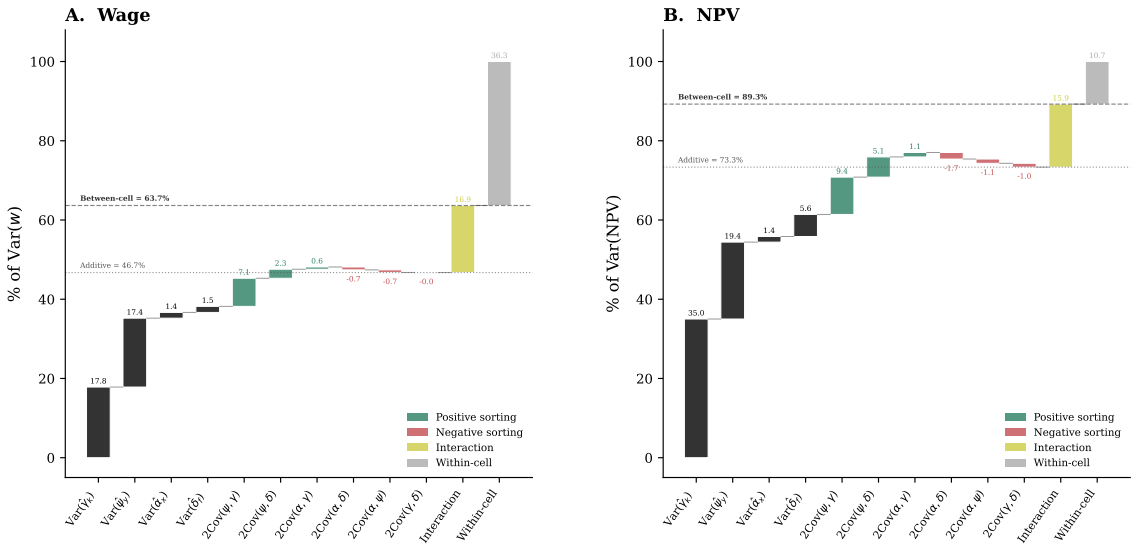


Figure 2.7. Additive variance decomposition of wages (Panel A) and NPV (Panel B). Each bar shows the contribution of a variance or covariance term to total variance (percentage points). Positive covariances are green, negative red, the interaction term yellow, and within-cell residual grey. Dotted and dashed reference lines mark the additive total and between-cell total respectively.

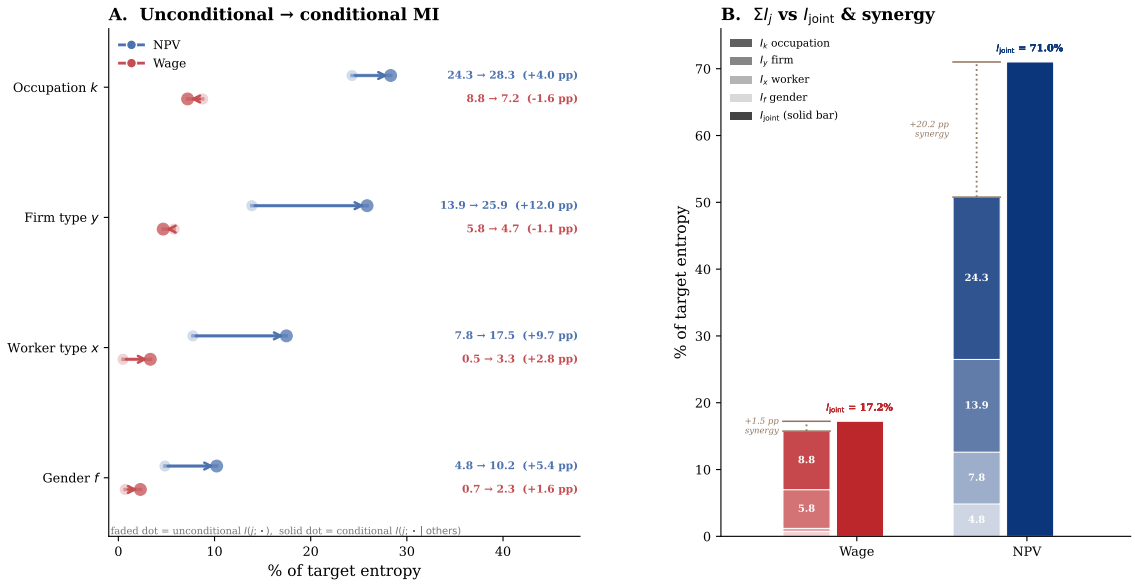


Figure 2.8. Mutual information decomposition: wages versus NPV. *Panel A:* unconditional MI (faded dot) versus full conditional MI (solid dot) for each factor; arrows indicate the synergy (+) or redundancy (−) shift. *Panel B:* sum of unconditional MIs (stacked bar) versus joint MI (solid bar); the gap is net synergy.

Figure 2.8 summarizes the central contrast. For wages (crimson), the stacked unconditional bar and the joint bar are nearly identical—the factors are approximately additive. For NPV (navy), the joint bar towers 20.2 pp above the sum of parts, with all four arrows pointing rightward (synergistic). The three decompositions converge: the same classification that explains 63.7% of wage variance but only

17.2% of wage entropy explains 89.3% of NPV variance and 71.0% of NPV entropy. The gap between variance and MI shares—large for wages, small for NPV—reflects the within-cell residual: large for wages (36.3%), small for NPV (10.7%). Table A2.22 in the Appendix consolidates all four factors across the three decomposition methods and both targets in a single summary.

This asymmetry has a direct implication for model selection in the sorting literature. Two-way sorting models that omit occupation appear adequate when evaluated on wage fit, because net synergy for wages is approximately zero—the missing dimension contributes additively, so its absence merely reduces explained variance without distorting the remaining attribution. For lifetime value, the same omission is far more consequential: 20.2 percentage points of NPV entropy exist only in the interaction among all four factors, and this synergistic information is destroyed, not merely redistributed, when a dimension is dropped.

2.5.2.3 Lifecycle dynamics

Table 2.12 reports pairwise mutual information between all six factor pairs for each age bin, measuring how strongly each pair is associated in the cross-section of employment spells at each career stage.

Table 2.12. Pairwise Sorting Mutual Information by Age Bin (bits)

Age bin	$I(x;y)$	$I(x;k)$	$I(x;f)$	$I(y;k)$	$I(y;f)$	$I(k;f)$
1	0.093	0.221	0.101	0.473	0.082	0.137
2	0.051	0.439	0.077	0.392	0.094	0.173
3	0.030	0.381	0.059	0.364	0.088	0.162
4	0.033	0.411	0.071	0.344	0.096	0.175
5	0.036	0.413	0.092	0.320	0.095	0.177

Three lifecycle patterns emerge among the structural factors. Worker–firm association $I(x;y)$ declines with age (0.093 → 0.036), indicating that the correspondence between worker type and firm type weakens as workers move across firms. Worker–occupation association $I(x;k)$ roughly doubles from 0.221 to 0.413, reflecting cumulative occupational specialization: workers increasingly sort into the occupations their type is best suited for. Firm–occupation association $I(y;k)$ declines steadily (0.473 → 0.320), meaning the occupational composition of firm types becomes more mixed with worker age, consistent with older workers having spent time at multiple firms.

The three pairs involving gender reveal distinct dynamics. Worker–gender association $I(x;f)$ starts moderate (0.101), dips in mid-career (0.059), and recovers to 0.092 by age bin 5. The mid-career trough is consistent with the most gender-balanced period of labor force participation; the late-career recovery reflects selective retention of women in type-specific roles. Firm–gender association $I(y;f)$ is remarkably stable (0.082–0.096), indicating that gender segregation across firm types neither intensifies nor dissipates over the career. Occupation–gender association $I(k;f)$ rises from 0.137 to 0.177, indicating that occupational segregation by gender *increases* over the lifecycle—men and women increasingly concentrate in different occupations as they age, consistent with gendered specialization deepening over the career.

Table 2.13 reports the 4-factor decomposition by age bin for both wages and NPV.

Table 2.13. Decomposition by Age Bin (% of subgroup variance or entropy)

	Variance Shapley						MI Shapley					
Age	ϕ_x	ϕ_y	ϕ_k	ϕ_f	Btw	Wth	ϕ_x	ϕ_y	ϕ_k	ϕ_f	I_j	Syn
Panel A: Wages												
1	12.6	19.8	13.7	6.0	52.0	48.0	2.5	4.7	4.7	1.1	13.0	+1.7
2	6.7	25.3	26.2	5.5	63.8	36.2	1.6	5.9	8.0	1.3	16.8	−0.2
3	7.3	24.7	28.9	6.2	67.1	32.9	2.2	6.0	9.1	1.7	18.9	+0.9
4	7.4	23.7	30.5	6.8	68.5	31.5	2.5	5.7	9.7	2.0	19.8	+1.2
5	8.5	22.4	29.6	7.4	67.9	32.1	3.0	5.2	9.3	2.2	19.6	+1.4
Panel B: NPV												
1	13.8	28.4	44.8	12.9	100	0	20.6	32.2	34.7	12.5	100	+10.7
2	10.9	29.9	47.5	11.7	100	0	20.8	27.7	38.6	12.9	100	+6.0
3	10.6	33.2	44.3	11.9	100	0	20.3	29.4	38.0	12.2	100	+15.7
4	9.6	35.3	44.0	11.1	100	0	20.1	29.6	38.6	11.6	100	+12.8
5	9.9	35.0	44.0	11.2	100	0	20.0	28.5	39.9	11.6	100	+13.7

Notes. Btw = total between-cell variance share; Wth = within-cell residual. $I_j = I(x, y, k, f; \cdot)$ is total MI. Syn $= I_j - \sum_j I(j; \cdot)$ is net synergy. NPV is deterministic in $(x, y, k, f,$ age), so between-cell = 100% and $I_j = H(\text{NPV})$ within each age bin.

The full decomposition for each age bin—including unconditional MI, full conditional MI, MI Shapley, and variance Shapley for all four factors is reported in Tables A2.15–A2.19 in the Appendix.

Wages over the lifecycle. For wages, the variance Shapley reveals a lifecycle pattern. Occupation's share more than doubles from 13.7% (age bin 1) to 30.5% (age bin 4), reflecting cumulative occupational specialization: as workers age, they sort into the occupations that best match their type, and the wage return to that sorting grows. Firm type is relatively stable (19.8–25.3%), worker type declines from 12.6% in the youngest bin to 6.7–8.5% thereafter, and gender's share rises modestly (6.0% → 7.4%), consistent with the deepening occupation–gender segregation documented in Table 2.12. Total between-cell variance rises from 52.0% to 68.5%, driven almost entirely by occupation.

The MI Shapley tells a similar story at much lower levels. Occupation's MI share doubles from 4.7% to 9.7% over the career, while firm type is hump-shaped and worker type is flat and small. Gender's MI share doubles (1.1% → 2.2%), indicating that gender becomes a sharper predictor of within-cell wage position as careers progress—a pattern consistent with the rising $I(k; f)$ in the pairwise sorting table, which suggests that gendered occupational specialization compounds over time. Net synergy is small at every age (−0.2 to +1.7 pp), confirming that the four dimensions of sorting remain approximately additive for wages throughout the career. This lifecycle pattern speaks to a central mechanism in the search-and-matching literature, where career wage growth is typically modeled as climbing a firm ladder through on-the-job search (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002). In our data, workers do not sort into progressively better firm types over their careers—$I(x; y)$ *declines* with age (Table 2.12). Instead, they sort into the occupations that match their type, and the wage return to that occupational specialization drives career wage growth.

NPV over the lifecycle. NPV within each age bin is a deterministic function of (x, y, k, f), computed by backward induction on the Markov chain. The between-cell variance share is therefore 100% and $I(x, y, k, f; \text{NPV}) = H(\text{NPV})$ at every age. The interesting question is how the Shapley values distribute this total among the four factors, and how the synergy structure evolves.

In the variance Shapley, occupation is the dominant factor at every age (44–48%), followed by firm type (28–35%), gender (11–13%), and worker type (10–14%). Firm type's share rises steadily from 28.4% to 35.0% over the career, while worker type declines from 13.8% to 9.9%. This pattern reflects career dynamics: for young workers, future occupational switches are frequent and worker type governs which paths are likely; for older workers, the remaining career horizon is shorter and the current firm-type premium accounts for a larger share of remaining value. Gender's variance share is stable at 11–13% across all ages, indicating that its contribution to lifetime-value dispersion neither intensifies nor dissipates over the career—even as occupation–gender segregation deepens (Table 2.12).

In the MI Shapley, a different pattern emerges. Worker type's MI share is remarkably stable at 20.0–20.8%—compared to its declining variance share (13.8% → 9.9%). This divergence arises because the MI Shapley captures distributional structure, not just mean differences: even when worker type explains less *variance* in NPV for older workers, it remains equally informative about *which part of the NPV distribution* a worker falls in. The same stability holds for gender (MI share 11.6–12.9% vs. variance share 11–13%) and for worker type's pairwise association with occupation ($I(x; k)$ stabilizes above 0.38 from age bin 2 onward). Net synergy among all four factors is large and positive at each age bin, with a pronounced mid-career peak: +6.0 pp in age bin 2, rising to +15.7 pp in age bin 3, then stabilizing at +12.8–+13.7 pp. The mid-career intensification is consistent with transition dynamics becoming more structured as workers accumulate sorting history: the career paths implied by each (x, y, k, f) combination diverge most sharply for experienced workers.

2.5.3 Destination Choice and Pecuniary Motives

2.5.3.1 Gross mobility patterns

Of the 6.36 million observed transitions in the estimation sample, 81.3% change firm type and 73.1% change occupation; 14.9% stay in the same (firm-type, occupation) cell. This dispersion of transitions across cells is the key feature that enables sorting analysis.

Among job-to-job movers who change employer (excluding moves to non-employment), Table 2.14 summarizes the fraction of moves that remain within the same firm type.

Within-firm-type mobility increases with age (14% at age bin 1 rising to 22% at age bin 3 and plateauing thereafter). Women are about three percentage points more likely than men to move within firm type (20% vs 17%), consistent with their higher representation in industries with internal labor markets. Across worker types, $x = 3$ (male, manual-occupation) workers have the highest across-type mobility (85%), while $x = 2$ (female, administrative) workers have the lowest (79%).

Table 2.14. Within-y Mobility Among J2J Movers

Group	Level	% same y	% same (y, k)
All J2J movers	—	18.7%	14.9%
By age bin	1	13.8%	9.5%
	2	19.1%	15.6%
	3	22.6%	18.8%
	4	22.3%	18.7%
	5	22.0%	18.7%
By gender	Male	17.3%	13.3%
	Female	20.2%	16.6%
By worker type	$x = 1$	18.1%	14.6%
	$x = 2$	21.5%	17.2%
	$x = 3$	15.4%	11.9%

2.5.3.2 Destination choice regressions

We now ask whether workers' destination choices align with lifetime value, current wages, or the option-value component of NPV. The analysis operates at the cell-pair level, where a cell is a labor-market state ($x, y, k, f,$ age) and a cell pair (s, s') is an origin–destination combination.

Excess transition probability. The key dependent variable is the *excess transition probability* $P_{\text{excess}}(s \to s')$, defined as the difference between the observed transition rate from origin s to destination s' and the marginal (population-average) destination frequency:

$$P_{\text{excess}}(s \to s') = \underbrace{\frac{n_{s\to s'}}{\sum_{s''} n_{s\to s''}}}_{\text{observed share}} - \underbrace{\frac{\sum_{s} n_{s\to s'}}{\sum_{s,s''} n_{s\to s''}}}_{\text{marginal share}}. \tag{2.5.1}$$

The marginal share measures how frequently destination s' appears across all origin cells—it is the destination frequency that would arise if workers chose destinations at random, independently of their origin. A positive P_{excess} means workers in state s flow to s' more often than this benchmark; a negative value means they flow there less often. By construction, $\sum_{s'} P_{\text{excess}}(s \to s') = 0$ for each origin: excess probabilities sum to zero within each origin cell, so any positive deviation toward one destination implies negative deviations toward others.

Regressors. The independent variables are cell-level value differences between destination and origin: $\Delta\text{NPV} = \text{NPV}(s') - \text{NPV}(s)$, $\Delta w = \mu_{s'} - \mu_s$ (the cell-mean wage difference), and—when NPV is decomposed—ΔV_{wage} (the capitalized current-wage component) and ΔV_{option} (the option-value component). All regressors vary at the cell-pair level, not at the individual level. This ensures that the wage and NPV measures operate at the same level of aggregation—the (x, y, k, f) cell—making their coefficients directly comparable. Individual-level wage variation (tenure, match quality, bargaining) is absorbed into within-cell dispersion and does not enter the regressions.

Interpretation. The regression $P_{\text{excess}}(s \to s') = \alpha + \beta\,\Delta V + \varepsilon$ asks: do workers flow to destinations with higher value *more often than the marginal destination frequency predicts*? A positive β means destination choices are systematically tilted toward higher-value states, beyond what random reallocation would produce. A zero β means the observed transition pattern is consistent with destination choice being independent of the value measure—workers end up in high- and low-value destinations at rates that match the population average.

Results. Table 2.15 reports the estimates. The sample is employed-to-employed transitions with $\Delta\text{NPV} \neq 0$. Standard errors are robust; observations are weighted by transition counts.

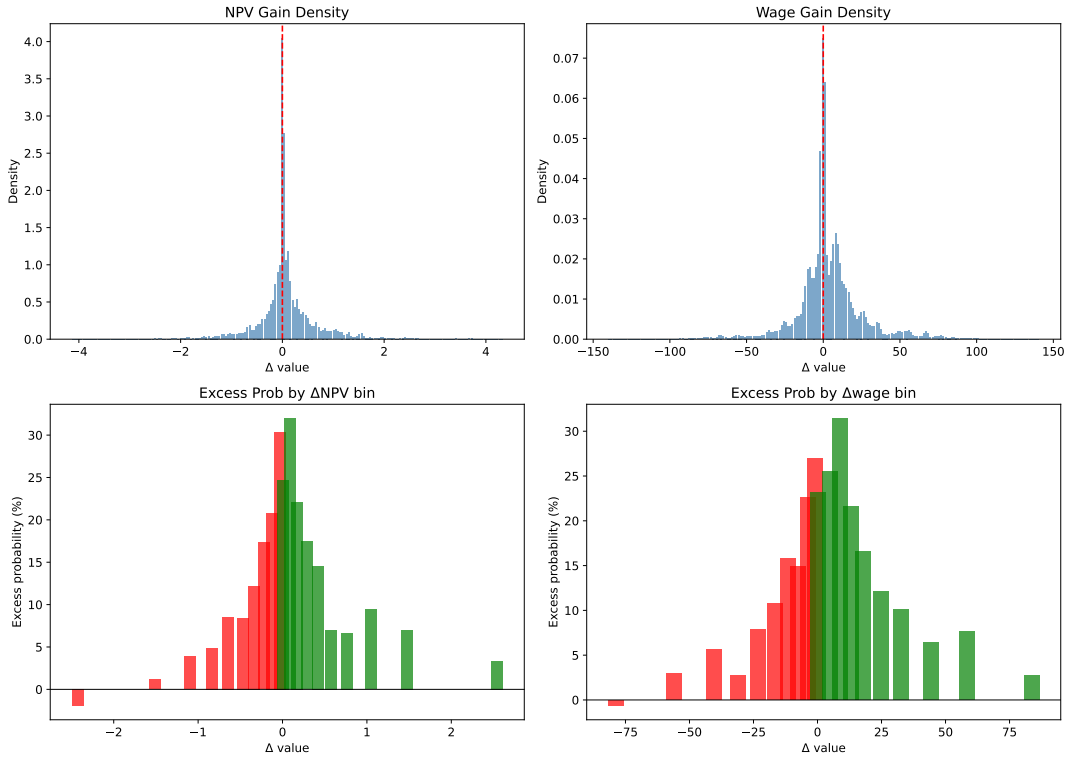


Figure 2.9. Distribution of Job-to-Job Transition Gains. *Top panels*: density of NPV gains (left) and wage gains (right); red dashed line marks zero. *Bottom panels*: excess transition probability relative to the marginal destination distribution (green = above random, red = below random), binned by NPV gain (left) and wage gain (right). Excess probability leans rightward for wages but is roughly symmetric for NPV; both panels spike at zero (lateral mobility).

The results confirm the asymmetry visible in Figure 2.9. First, the linear response of P_{excess} to overall ΔNPV (column t2a) is small and statistically insignificant: workers do not flow disproportionately to destinations with higher lifetime value. Second, when ΔNPV is decomposed into its wage and option components (column t2c), only the wage component loads significantly ($\hat{\beta} = 0.037$, SE 0.018). The option component loads negatively and insignificantly. Workers deviate from the marginal destination distribution in the direction of higher current wages, not higher option value. This is visible in Figure 2.9: the excess probability bars lean sharply rightward for wages—workers move to higher-wage destinations far more often than the marginal benchmark—whereas for NPV the excess distribution is nearly symmetric around zero. Third, the model has markedly higher R^2 when the wage/option

Table 2.15. Destination Choice and NPV Regression Results

Variable	(t2a)	(t2b)	(t2c)	(t2d)	(t3a)	(t3b)	(t4)
ΔNPV_{100k}	0.0144				−0.0015		
	(0.0177)				(0.0036)		
Δw		0.0022				0.0002	
		(0.0006)				(0.0001)	
$\Delta V_{\text{wage},100k}$			0.0369				
			(0.0175)				
$\Delta V_{\text{opt},100k}$			−0.0114				
			(0.0172)				
$\text{NPV}_{\text{dest},100k}$				0.0117			
				(0.0179)			
$\text{NPV}_{\text{origin},100k}$				−0.0252			
				(0.0179)			
$\Delta\text{NPV}^2_{100k}$					−0.0046		
					(0.0006)		
Δw^2						0.0000	
						(0.0000)	
$\Delta\text{NPV}_{\text{losses}}$							−0.0925
							(0.0085)
$\Delta\text{NPV}_{\text{gains}}$							0.1252
							(0.0112)
Constant	0.2138	0.2071	0.1874	0.2554	0.1860	0.1843	0.2287
	(0.0103)	(0.0095)	(0.0073)	(0.0239)	(0.0093)	(0.0092)	(0.0114)
Observations	36,151	35,475	36,151	36,151	63,436	63,436	63,436
R^2	0.0004	0.0151	0.1380	0.0087	0.0092	0.0058	0.1105

decomposition replaces the aggregate ΔNPV (14% vs 0.04%), indicating that the sharpness of the wage response was diluted by a noisy option component in the aggregate regression. Fourth, column t4 shows strongly asymmetric slopes on ΔNPV above and below zero. The slope on positive gains is negative (−0.093, SE 0.009) and the slope on negative gains is positive (0.125, SE 0.011). Read together with the intercept, this means P_{excess} peaks at $\Delta\text{NPV} = 0$ and declines in both directions—lateral mobility to destinations of similar lifetime value is the most common transition type, exactly as the spike at zero in Figure 2.9 shows.

Heterogeneity regressions, reported in the appendix table A2.23, confirm this pattern. The NPV-origin coefficient is consistently negative and significant: among workers in higher-NPV origins, excess choice probabilities on destinations are weaker. Interactions with worker type, gender, and age bin are all individually insignificant, though the age pattern is suggestive of more NPV responsiveness in mid-career (age bins 3–5) than in youth (bins 1–2). The non-employment regression (t7) shows a positive coefficient on origin NPV, meaning workers in higher-NPV states are more likely to transition to non-employment. This reflects hazard compounding in high-NPV cells rather than voluntary exit.

Table A2.24 reports structural log-share regressions that absorb origin-cell fixed effects. These estimates are the conditional-logit analogs of the destination choice model. Column t8a shows a positive and significant response to destination NPV ($\hat{\beta} = 0.086$, SE 0.028). Column t8b shows a positive and significant response to destination wage. Column t8c includes both; the NPV coefficient flips sign, indicating that wage and NPV are collinear conditional on origin FE. Column t8d is the definitive test: conditional on destination fixed effects (absorbed by (*x, female, age,* origin state) FE), workers respond strongly to the wage component of value ($\hat{\beta} = 0.112$, SE 0.031) and not at all to the option component ($\hat{\beta} = 0.002$, SE 0.034). The combined test statistic on ΔV_{opt} is close to zero across specifications.

The consistent message across reduced-form and structural specifications is that workers respond to current pay differences across destinations, not to lifetime-value differences. The decomposition results rule out the interpretation that this pattern is an artifact of measurement error in career value: the (x, y, k, f) classification explains 71.0% of NPV entropy, with all four dimensions synergistic and the within-cell residual only 10.7%. The option-value signal is large, structured, and informationally rich. Workers do not respond to it.

A caveat is in order. The NPV values are computed under the estimated transition matrix, which embeds the observed mobility behavior. The option-value component therefore measures the expected gain from future mobility *as it actually occurs in the data*, not as it would occur under counterfactual optimal behavior. The finding is thus internally consistent: workers do not deviate from average transition patterns in the direction of higher option value, and the option values themselves are computed under those same average patterns. A structural model with endogenous forward-looking mobility—as in Lentz et al. (2023)—would yield different NPV values and potentially different conclusions, but would require the behavioral assumptions our reduced-form approach is designed to avoid.

2.5.3.3 Flow decomposition

To probe this further, we decompose job transitions into four quadrants based on the sign of Δw (current wage change) and ΔV_{opt} (option-value change).

Figure 2.10 visualizes the decomposition. Panel A traces quadrant shares across age bins. Harvest moves (w+ opt−) rise modestly with age, consistent with older workers settling into stable, well-paid jobs at the cost of future mobility. Ladder-climbing moves (w− opt+) are rare at all ages but slightly elevated among the youngest workers. Panel B shows that women have a higher share of harvest moves than men (roughly 39% vs 28%), consistent with the within-firm-type mobility gap in Table 2.14: women move within the same firm type at higher rates (20% vs 17%), suggesting they are more concentrated in firms with structured internal labor markets where wage progression is available but cross-firm mobility is limited. Panel C plots the structural trade-off directly: each point is a destination cell, sized by transition count. The strong negative slope ($r = -0.67$) means that destinations offering a wage premium over the origin tend to offer a lower option value, and vice versa. Workers predominantly

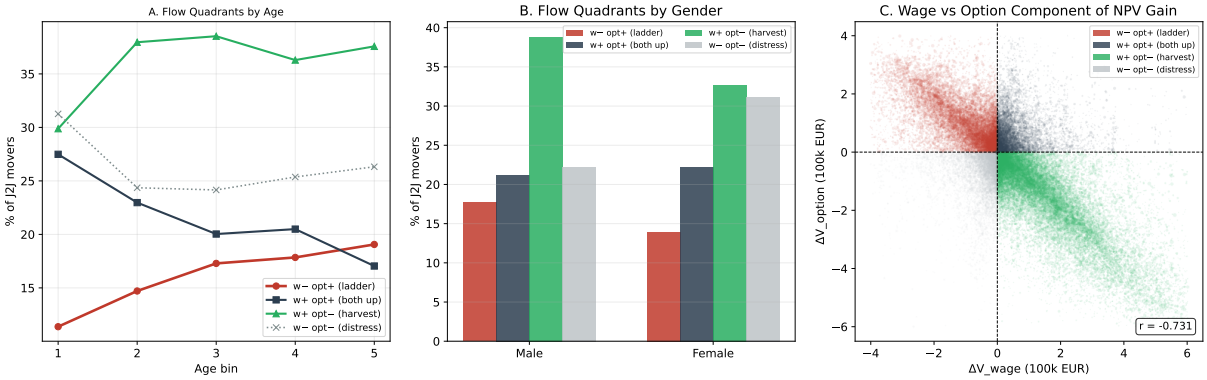


Figure 2.10. Flow Quadrant Decomposition of job Transitions. *Panel A*: quadrant shares by age bin; harvest moves (w+ opt−) dominate throughout. *Panel B*: same decomposition by gender. *Panel C*: scatter of ΔV_{wage} against ΔV_{opt} for all employed-to-employed transitions; the strong negative correlation ($r = -0.73$) confirms a systematic wage–option trade-off.

move into the harvest quadrant (bottom-right, 36.1% of all moves), suggesting that if a trade-off exists, current wages win.

Table 2.16. Job to job Flow Decomposition: NPV vs Wage Quadrants

Quadrant	Count	Share
Wage +, Option + (both improve)	990,201	21.7%
Wage +, Option − (harvest)	1,646,958	36.1%
Wage −, Option + (ladder climb)	730,924	16.0%
Wage −, Option − (distress move)	1,199,286	26.3%

Of all J2J moves, 57.8% are wage-up and 37.7% are option-up. The single most common type of move is the "harvest" (36.1%): a worker takes a current-wage increase while accepting lower option value. Distress moves—where both wages and option value fall—account for 26.3%. Pure career-building moves (wage down, option up) are relatively rare (16.0%). This composition reinforces the regression finding: when workers face a tradeoff between current wages and option value, they choose current wages.

2.5.3.4 Within-type flows and the aggregate picture

Table 2.17 reports the share of job moves that are "uphill" by NPV, broken down by worker type and gender.

Table 2.17. Share of J2J Moves Uphill by NPV, by Worker Type and Gender

Worker type	Male (%)	Female (%)
$x = 1$	50.5	46.6
$x = 2$	48.2	44.3
$x = 3$	46.8	36.4

Uphill-by-NPV move shares cluster around 45–50%, marginally below the 50% one would get by pure random reshuffling. Men have a higher propensity to make uphill moves, dominated by type $x = 1$

(50.5% uphill by NPV), and females of type $x = 3$ make the least uphill moves (36.4%). Across all groups, the fraction uphill is close to one-half, consistent with the reduced-form regression result that NPV does not strongly drive choice.

2.5.4 Summary

Four findings define the paper's empirical contribution. First, occupation is the single largest contributor to both current wages and lifetime value across all three decomposition methods: 26.6% of wage variance (Shapley), 17.8% of own wage dispersion with 7.1% of firm–occupation sorting (additive), 8.2% of wage entropy (MI), and 40.8% of NPV variance, 27.7% of NPV entropy. The dominant sorting channel runs between firms and occupations ($\rho = 0.201$), not between firms and workers ($\rho = -0.065$). Second, the MI decomposition reveals a qualitative shift from wages to NPV: sorting among worker types, firm types, and occupations is approximately additive for wages (net synergy +1.5 pp) but deeply complementary for lifetime value (+20.2 pp). Over a fifth of NPV entropy is explained by information that exists only in the interaction among dimensions and cannot be recovered from any factor in isolation. The within-cell wage residual (36.3%) captures forces that are large for current pay but average out over a career of transitions governed by the discrete classification. Third, gender operates not through its own main effect but through non-additive complementarities with the (x, y, k) cell, accounting for roughly 11 percentage points of wage variance. Fourth, despite the strong structure of lifetime value, workers' destination choices respond to the capitalized current-wage component but not to the option-value component. The wage component predicts choice with $\hat{\beta} = 0.112$ in the structural logit; the option component is indistinguishable from zero.

2.6 Conclusion

This paper jointly estimates sorting among workers, firms, and occupations, conditional on gender and age, in the French administrative record. Using a Classification EM algorithm on the universe of salaried employment from 2015–2019, we recover three latent worker types, twelve latent firm types, and a full transition structure over the eight PCS occupations. A job is the pair (y, k); a worker's labor-market state is (x, y, k), with gender and age conditioning all model parameters. We compute the net present value of lifetime earnings in every state via exact backward induction on the estimated Markov chain, and decompose both wage and NPV variation using three complementary methods: Shapley variance decomposition, additive main-effects decomposition with pairwise sorting covariances, and Shapley-valued mutual information decomposition.

Four key findings emerge: first, occupation is the single largest contributor to both wage and NPV variation, regardless of the decomposition method. Its Shapley contribution to age-adjusted wage variance is 26.6%, exceeding firm type (23.0%), worker type (7.8%), and gender (6.3%). Its Shapley contribution to NPV variance is 40.8%, and to NPV entropy 27.7%—in both cases the largest of any factor. The additive decomposition reveals that occupation and firm type have nearly identical own wage dispersion (17.8% vs. 17.4%), but the dominant sorting channel runs between firms and

occupations ($\rho = 0.201$, 7.1% of wage variance), not between firms and workers ($\rho = -0.065$). Classification frameworks that collapse occupation into worker or firm heterogeneity miss both the largest source of heterogeneity and the primary sorting channel.

Second, the MI decomposition reveals a qualitative shift in the structure of sorting between current wages and lifetime value. For wages, the contributions of worker type, firm type, occupation, and gender are approximately additive (net synergy +1.5 pp), with occupation and firm type mildly redundant and worker type and gender synergistic. For NPV, all four dimensions are synergistic, and net synergy reaches +20.2 pp—over a fifth of NPV entropy exists only in the interaction among factors and cannot be recovered from any factor in isolation. The within-cell wage residual (36.3%) captures forces—tenure, match quality, bargaining—that are large for current pay but average out over a career of transitions governed by the discrete classification. This pervasive complementarity reflects the forward-looking structure of lifetime value: NPV integrates over transition probabilities, separation hazards, and occupational switching rates that are governed by the full (x, y, k, f) state jointly. Two-way sorting models appear adequate when evaluated on wage fit, because the missing dimensions contribute additively; for lifetime value, omitting dimensions destroys synergistic information that cannot be recovered.

Third, gender operates primarily through non-additive complementarities rather than through its own main effect. Gender's own variance is small (1.5% of wage variance, 5.6% of NPV variance), but the non-additive interaction term—which captures cell-specific premia that cannot be decomposed into separate main effects—is 16.9% for wages and 15.9% for NPV, with gender responsible for roughly 11 percentage points of the wage interaction. The MI decomposition tells the same story: gender's unconditional MI is small (0.7% for wages, 4.8% for NPV), but its conditional MI after the other three factors triples (2.3% for wages) or more than doubles (10.2% for NPV). Gender–occupation segregation deepens over the lifecycle ($I(k; f)$ rises from 0.137 to 0.177 bits), even as gender's additive occupation premium $\hat{\gamma}_k$ is uncorrelated with gender ($\rho = -0.002$ for wages). The gender differences in wages is not about women being in low-premium occupations; it is about gender interacting with the specific (x, y, k) cell.

Fourth, workers' destination choices are indistinguishable from myopic with respect to career-building value. In structural log-share regressions with origin fixed effects, the option-value component of lifetime earnings is indistinguishable from zero ($\hat{\beta} = 0.002$, SE 0.034), while the capitalized current-wage component loads positively and significantly ($\hat{\beta} = 0.112$, SE 0.031). In reduced-form regressions, the same asymmetry holds: the wage component loads at 0.037 (significant) and the option component at −0.011 (insignificant). When workers face a pure tradeoff—36% of observed J2J moves involve a current-wage increase paired with an option-value decline—they take the wage. This 36% harvest-move share is itself a lower bound on myopic behavior, since it counts only transitions where the wage–option tradeoff is unambiguous. Observed mobility responds to the immediate wage signal but not to the career-building signal that the model reveals is present.

From a measurement perspective, these findings imply that conventional variance decompositions of wages—in particular the AKM framework and its discrete-type extensions, which collapse occupation into worker or firm fixed effects—understate the informational content of the joint classification. The redundancy analysis reinforces this: occupation's contribution to wage information is largely unique, overlapping only modestly with firm type, meaning that collapsing occupation into the other dimensions does not merely redistribute information—it destroys it. The additive decomposition further shows that the primary sorting channel (firm–occupation, $\rho = 0.201$) is invisible to two-way frameworks. Within-cell wage dispersion is large, but forward-looking value is highly structured by the (x, y, k, f) classification, with pervasive synergy among all dimensions.

From a policy perspective, the weak response of mobility to option value raises a concrete question: are workers choosing suboptimal careers because they cannot perceive the option-value signal, or because they face frictions that prevent them from acting on it? Policies that improve labor-market information—such as wage transparency or career-trajectory tools—would be effective only under the first interpretation; policies that reduce mobility frictions—such as portable benefits, improved job-search infrastructure, or reduced occupational licensing barriers—would be effective under the second. The gender findings sharpen this question: the large non-additive gender interaction (11 pp of wage variance) suggests that career-trajectory tools would need to be gender-specific to be informative, since the option-value landscape differs substantially between men and women in the same (x, y, k) cell. Distinguishing these interpretations is a natural next step.

Several extensions suggest themselves, each addressing a limitation of the current framework. Most importantly, the model treats the worker type as time-invariant. This is a substantive restriction: skill accumulation, career-stage transitions, and human capital depreciation during non-employment could all manifest as changes in the effective worker type over the life cycle. The lifecycle analysis partially addresses this concern —worker type's MI share of NPV is remarkably stable at 20% across all age bins, suggesting that the time-invariant classification captures persistent heterogeneity rather than a transient feature of early-career sorting—but allowing x to transition slowly would enrich the lifecycle dynamics and sharpen the distinction between persistent worker heterogeneity and accumulated experience. Second, adding firm productivity data (when available) would enable tests of whether firm type reflects complementarity-driven sorting or rent sharing. Third, extending the analysis to multiple European countries using the LEED infrastructure would reveal whether the leading role of occupations and the pervasive NPV synergy are specific to French labor-market institutions or a more general feature of skill-based economies.

Bibliography

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

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

Martyn J Andrews, Leonard Gill, Thorsten Schank, and Richard Upward. High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A*, 171(3):673–697, 2008.

Miguel Azkarate-Askasua and Manuel Zerecero. Occupational mobility and wage dynamics. *Journal of Labor Economics*, 42(2):413–456, 2024.

Damien Babet, Olivier Godechot, and Marco Palladino. In the land of akm: The performance of the akm estimator in finite samples. *Labour Economics*, 83:102388, 2023.

Gary S Becker. A theory of marriage: Part i. *Journal of Political Economy*, 81(4):813–846, 1973.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer-employee data. *Econometrica*, 83(3):699–739, 2019.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643, 2022. doi: https://doi.org/10.3982/ECTA15238. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15238.

Katarína Borovičková and Robert Shimer. Assortative matching and wages. *Journal of Political Economy*, 132(4):1009–1054, 2024.

Kenneth Burdett and Dale T. Mortensen. Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–273, 1998.

David Card, Jorg Heining, and Patrick Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015, 2013.

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

Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Leave-Out Estimation of Variance Components. *Econometrica*, 88(5):1859–1898, September 2020.

Rasmus Lentz, Suphanit Piyapromdee, and Jean-Marc Robin. The anatomy of sorting—evidence from danish data. *Econometrica*, 91(6):2409–2455, 2023. doi: https://doi.org/10.3982/ECTA16425. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16425.

Ilse Lindenlaub. Sorting Multidimensional Types: Theory and Application. *The Review of Economic Studies*, 84(2):718–789, 01 2017. doi: 10.1093/restud/rdw063.

Ilse Lindenlaub and Fabien Postel-Vinay. The worker-job surplus. Working Paper 28402, National Bureau of Economic Research, January 2021. URL http://www.nber.org/papers/w28402.

Jeremy Lise and Fabien Postel-Vinay. Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–76, August 2020. doi: 10.1257/aer.20162002. URL https://www.aeaweb.org/articles?id=10.1257/aer.20162002.

Jeremy Lise, Costas Meghir, and Jean-Marc Robin. Matching, sorting and wages. *Review of Economic Dynamics*, 19:63–87, 2016.

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

Lloyd S. Shapley. A value for n-person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Volume II*, pages 307–317. Princeton University Press, 1953.

Robert Shimer and Lones Smith. Assortative matching and search. *Econometrica*, 68(2):343–369, 2000.

Jae Song, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. Firming up inequality. *The Quarterly Journal of Economics*, 134(1):1–50, 2019.

CHAPTER 3

Signals, Noise, and Sorting

Abstract

This paper investigates the structural reliability of the AKM variance decomposition (developed by Abowd et al. (1999)) within sparse bipartite worker-firm networks. I demonstrate that the incidental-parameter bias inherent in plug-in estimators is driven by the spectral properties of the graph Laplacian, where a low average node degree ($\bar{d}$) serves as a primary diagnostic for estimation fragility. Crucially, the analysis reveals that distortions in sorting covariance and worker variance shares are not merely artifacts of limited mobility; rather, they arise from the interaction of idiosyncratic noise with the weighting structure of the bipartite graph.Through Monte Carlo simulations, performance of some bias correction methods are evaluated. The results indicate that while these corrections reduce bias, they remain susceptible to noise-amplification when the network's spectral density is low, a phenomenon that systematically attenuates evidence of positive assortative sorting. By identifying $\bar{d}$ as a sufficient statistic for the severity of spectral collapse, this research provides a framework for characterizing the structural sensitivity of variance component estimators to the the underlying worker-firm network.

3.1 Introduction

Since the seminal work of Abowd et al. (1999), the two-way fixed effects decomposition of log wages into worker and firm components has become a standard tool for understanding the sources of wage inequality. Applications of the AKM framework span many countries and time periods, from West Germany (Card et al., 2013) to the United States (Song et al., 2019), and the methodology has been extended to study gender wage gaps (Card et al., 2016), the role of firms in rising inequality (Card et al., 2018), and the dynamics of worker sorting across contexts (Bonhomme et al., 2023). The appeal of the AKM approach lies in its interpretation as a linear projection of wages onto a high-dimensional bipartite worker-firm network, allowing researchers to summarize wage dispersion through the geometry of this network via worker effects, firm effects, and their covariance.

However, the estimation of these variance components is complicated by the fact that fixed effects are themselves high-dimensional projections of noisy outcomes on the worker-firm graph. As a result, plug-in estimators suffer from bias due to sampling error that is propagated through the projection operator, leading to a well-documented overstatement of dispersion in estimated fixed effects (Abowd et al., 1999, 2002). In response, a vibrant literature has developed bias-correction methods. Andrews

et al. (2008) proposed a degrees-of-freedom correction under the assumption of homoscedastic errors. More recently, Kline et al. (2020) introduced a leave-out estimator that is unbiased and consistent under arbitrary heteroscedasticity, provided the worker-firm mobility network is sufficiently connected. Babet et al. (2025) developed a split-sample estimator that, by randomly partitioning workers within each firm, delivers unbiased estimates of the firm variance and the sorting covariance while circumventing the need to estimate the full leave-out correction. These advances have greatly improved the credibility of the firm variance component, and many applied researchers now report bias-corrected estimates alongside traditional plug-in figures.

Yet a closer examination of these corrections reveals an important but underappreciated fact: although the leave-out estimator of Kline et al. (2020) provides direct bias corrections for all three components—$\mathrm{Var}(\psi)$, $\mathrm{Var}(\alpha)$, and $2\,\mathrm{Cov}(\alpha, \psi)$—the finite-sample precision of the worker and covariance corrections is governed by a different asymptotic regime than that of the firm variance. The firm variance correction benefits from the pooling of information across all movers; its variance decays as the total number of firm-to-firm transitions grows. In contrast, the bias correction for $\mathrm{Var}(\alpha)$ depends on the precision with which each individual worker effect $\hat{\alpha}_i$ is estimated, which is fundamentally limited by the number of time periods T_i observed per worker. When T_i is small, as it is in virtually all matched employer-employee panels, the error in $\hat{\alpha}_i$ remains non-negligible regardless of how many firms J are in the sample. The split-sample estimator of Babet et al. (2025) adopts a different approach: it delivers direct, unbiased estimates of $\mathrm{Var}(\psi)$ and $2\,\mathrm{Cov}(\alpha, \psi)$ by orthogonalizing the estimation errors across independent subsamples, but it does not provide a direct estimate of $\mathrm{Var}(\alpha)$; instead, $\mathrm{Var}(\alpha)$ is backed out by subtracting the other estimated components from the total wage variance. As a result, any noise in the estimates of $\mathrm{Var}(\psi)$, $\mathrm{Var}(\varepsilon)$, or $2\,\mathrm{Cov}(\alpha, \psi)$ is mechanically transmitted to the implied worker variance share. In both the leave-out and split-sample frameworks, therefore, the reliability of the worker variance and the sorting covariance hinges critically on the amount of information available to estimate each worker's fixed effect. As I demonstrate below, this information is severely limited in the sparse, low-workers-per-firm settings that characterize some empirical applications of the AKM model.

Matched employer-employee datasets typically contain a small number of time periods per worker (often 2–5 years) and, perhaps more importantly, a small (unweighted) average number of workers per firm in the sample used for analysis. For instance, in Abowd et al. (2002), the ratio of the number of workers to the number of firms in the largest connected set is approximately 3 in France and 4 in the United States. In Portugal, the sample used in Card et al. (2018) has a ratio of about 9 in the largest connected set. While there are empirical studies where this ratio is higher (for example, Woodcock (2015) has an average close to 30, Sorkin (2018), 191, Abowd et al. (2019), 22, Song et al. (2019), 15, Kline et al. (2020), 16, Babet et al. (2025), 32), heterogeneity of this ratio is given scant attention and not always reported. Although the literature has emphasized the role of mobility in connecting firms, the properties of the projection operator also depend critically on the average firm size, which governs the leverage of individual workers and the effective normalization of observations.

When workers per firm is low, the mobility network linking firms through worker transitions is inherently sparse. Even if the overall mobility rate per period is moderately high (e.g., 10%-40%), the number of movers connecting any given firm to the rest of the network remains small. Increasing the number of movers strengthens connections, but does not fully eliminate the problem: identification of firm effects may still rely on relatively few independent comparisons, and the graph Laplacian of the firm network—the matrix that governs the precision of the firm effect estimates—possesses a small spectral gap. Jochmans and Weidner (2019) showed that the convergence rate of fixed effect estimators in a graph-structured panel is determined by the second smallest eigenvalue of the normalized Laplacian, linking it to the Cheeger constant of the graph (Friedland and Nabben, 2002). Kline et al. (2020) extended this insight to variance components, proving that the asymptotic distribution of the leave-out estimator involves non-central χ^2 components when the spectral gap is small. Intuitively, a small spectral gap indicates bottlenecks—subsets of firms that are only weakly connected to the rest of the network—which slow the rate at which information about firm effects accumulates across the sample.

The present paper makes a distinct but complementary point: the distortions I document arise not from weak identification in the sense of Kline et al. (2020), nor from a failure of invertibility of the underlying system, but from the interaction between idiosyncratic noise and the weighting structure induced by the bipartite worker-firm graph. When $\sigma_\varepsilon^2 = 0$—that is, when wages are observed without error—all four estimators (plug-in, Andrews, KSS, and split-sample) recover the true variance shares exactly, regardless of network structure. The problem is therefore not one of identification or ill-conditioning, but of how sampling noise is propagated through the projection. When $\sigma_\varepsilon^2 > 0$, estimation errors $\hat{\psi}_j - \psi_j$ are mapped into firm effects through the pseudoinverse of the graph Laplacian, and firms with few workers assign disproportionately large weight to each mover. The resulting amplification of noise is governed not only by the spectral gap λ_2 of the induced firm network, but also by the heterogeneity of node degrees that determines the effective normalization of the projection.

The bias in the firm variance, $\sigma^2 B_\psi = \sigma^2 \operatorname{tr}(\mathbf{AS}^+)$, is absorbed asymmetrically across the variance decomposition: OLS back-substitution assigns the firm-effect error to worker effects for stayers, inflating $\widehat{\operatorname{Var}}(\alpha)$ and attenuating $2\widehat{\operatorname{Cov}}(\alpha, \psi)$ by an amount proportional to $\sigma^2 B_\psi$. Importantly, the bias factor B_ψ is driven primarily by the joint structure of firm size and mobility, not by mobility alone. A low number of workers per firm mechanically increases the leverage of each observation, so that even with substantial mobility the induced firm-firm graph does not average idiosyncratic shocks effectively.

From a spectral perspective, degree-based bounds show that λ_2 is constrained by the lower tail of the degree distribution (Fiedler, 1973; Mohar, 1991). In the firm-firm projection, degrees are themselves endogenously determined by the interaction of firm size and mobility: the number of effective connections of a firm is bounded above by firm size times the mobility rate. When firms are small, this induces a structural limitation on effective connectivity even when aggregate mobility is high, so that increasing the number of movers does not proportionally increase the amount of independent information available for averaging noise.

In this paper, I provide a systematic Monte Carlo evaluation of four leading estimators—plug-in, the homoskedastic correction (HO) proposed by Andrews et al. (2008), the Leave-one-out estimator introduced by Kline et al. (2020) and the split-sampler by Babet et al. (2023)—under a range of relevant designs. I vary the number of firms, the average workers per firm N/J, the mobility rate (5%, 10%, 20%, and 40% per period), and the firm size distribution (uniform and Zipf). The data generating process incorporates sorting ($\rho = 0.3$) and idiosyncratic noise. Across all specifications, I find that none of the estimators delivers a reliable firm-effect variance when the average node degree $\bar{d}$ is low; the bias corrections reduce the plug-in distortion but leave economically significant errors, and the split-sample estimator can even undershoot because the common Largest Connected Set ceases to be representative. The firm variance share becomes trustworthy only once $\bar{d}$ surpasses a modest threshold. The worker variance share and the sorting covariance remain severely distorted even when the firm variance is well estimated; at low $\bar{d}$ both components are unreliable, and the sorting covariance in particular is often driven to zero or negative values, completely masking the true positive assortative sorting. For instance, in the Zipf design with 10 % mobility, the KSS estimate of $\mathrm{Var}(\alpha)$ is biased upward by roughly 20 % of the true share, and the split-sample estimate of $2\,\mathrm{Cov}(\alpha, \psi)$ is distorted by a similar magnitude. The sufficient statistic for the severity of the problem is the average node degree itself (or, equivalently, the workers-per-firm ratio and the per-period mobility rate). Because the bias factor B_ψ is inversely related to $\bar{d}$, a low average node degree implies a large bias factor and hence unreliable worker and covariance shares. Applied researchers should therefore routinely report $\bar{d}$ alongside their variance decompositions; when $\bar{d}$ is small, the worker variance share and the sorting covariance should be treated as unreliable.

The remainder of the paper is organized as follows. Section 3.2 develops the connection between the AKM model, the Laplacian structure of the normal equations, and the spectral characterization of bias. Section 3.3 reviews the four bias-correction methods. Section 3.4 presents the Monte Carlo design and results. Section 3.5 discusses implications for empirical work and section 2.6 concludes.

3.2 The Geometry of Noise Amplification

3.2.1 The bipartite network and the firm Laplacian

Consider a balanced panel of N workers observed over T periods at J firms. Let $X_{i,t} \in \{1, \dots, J\}$ denote the firm employing worker i at time t. Stacking all $n = NT$ observations, the canonical wage equation is:

$$\mathbf{w} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{F}\boldsymbol{\psi} + \boldsymbol{\varepsilon}, \tag{3.2.1}$$

where $\mathbf{D} \in \{0,1\}^{n \times N}$ is the worker-indicator design matrix, $\mathbf{F} \in \{0,1\}^{n \times J}$ is the firm-indicator design matrix, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)'$, and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_J)'$. After concentrating out worker effects via the within-worker projector $\mathbf{M}_D = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$, the firm effects solve the reduced normal equations $\mathbf{S}\hat{\boldsymbol{\psi}} = \mathbf{F}'\mathbf{M}_D\mathbf{w}$, where $\mathbf{S} = \mathbf{F}'\mathbf{M}_D\mathbf{F}$ is the firm-level Gram matrix.

Proposition 1 (Laplacian structure). *The matrix* $\mathbf{S}$ *is a weighted graph Laplacian of the firm–firm projection of the bipartite worker–firm mobility network. Its entries are given by:*

$$S_{jk} = \begin{cases} \sum_i n_{ij}\left(1 - \dfrac{n_{ij}}{T_i}\right) & \text{if } j = k, \\ -\sum_i \dfrac{n_{ij}\, n_{ik}}{T_i} & \text{if } j \neq k, \end{cases}$$

where n_{ij} *is the total periods worker* i *spends at firm* j*, and* $T_i = \sum_j n_{ij}$*.* $\mathbf{S}$ *is positive semidefinite with* $\ker(\mathbf{S}) = \operatorname{span}(\mathbf{1}_J)$ *in a connected network. Proof in Appendix A3.2.1.*

Definition 1 (Node degree and average connectivity). *The* node degree *of firm* j *is given by the diagonal entry* $d_j = S_{jj}$ *of the Laplacian. It measures the identifying mass incident to firm* j*, i.e. the effective number of mover-periods contributing to its identification. A stayer contributes zero, while a mover who spends* s *periods at firm* j *contributes* $s(T-s)/T$*.*

The average connectivity *of the network is the mean node degree,*

$$\bar{d} = \frac{1}{J}\sum_{j=1}^{J} d_j = \frac{\operatorname{tr}(\mathbf{S})}{J},$$

where the second equality follows from the fact that the trace equals the sum of diagonal elements.

3.2.2 The estimand: variance as a quadratic functional

The target of the AKM decomposition is the worker-weighted variance of firm effects:

$$\theta = \boldsymbol{\psi}'\mathbf{A}\boldsymbol{\psi},$$

where the $J \times J$ observation-weighted centering matrix is

$$\mathbf{A} = \frac{1}{n}\left(\mathbf{N}_F - \frac{1}{n}\mathbf{n}_F\mathbf{n}_F'\right). \tag{3.2.2}$$

Here $n_j = \sum_{i=1}^{N} n_{ij}$ is the total number of observations at firm j, $\mathbf{N}_F = \operatorname{diag}(n_1, \ldots, n_J)$, and $\mathbf{n}_F = (n_1, \ldots, n_J)'$.

The nullspace structure of $\mathbf{A}$ matches that of the Laplacian $\mathbf{S}$: in a connected network $\ker(\mathbf{S}) = \operatorname{span}(\mathbf{1}_J)$ (Proposition 1), and $\mathbf{A}\mathbf{1}_J = \mathbf{0}$ by construction. The Moore-Penrose pseudoinverse $\mathbf{S}^+$ maps into $\ker(\mathbf{S})^\perp = \{\mathbf{1}_J\}^\perp$, so the estimated firm effects satisfy the minimum-norm solution $\hat{\boldsymbol{\psi}} = \mathbf{S}^+\mathbf{b} \perp \mathbf{1}_J$ and therefore lie in $\ker(\mathbf{A})^\perp$, the subspace on which $\mathbf{A}$ is positive definite. This ensures that the plug-in variance $\widehat{\operatorname{Var}}_n(\psi) = \hat{\boldsymbol{\psi}}'\mathbf{A}\hat{\boldsymbol{\psi}}$ is strictly positive whenever $\hat{\boldsymbol{\psi}} \neq \mathbf{0}$.

Under homoskedasticity ($\sigma_\varepsilon^2 > 0$), the plug-in estimator $\hat{\theta} = \hat{\boldsymbol{\psi}}'\mathbf{A}\hat{\boldsymbol{\psi}}$ is upwardly biased (Andrews et al., 2008; Kline et al., 2020):

$$\mathrm{E}[\hat{\theta}] = \theta + \sigma_\varepsilon^2 \underbrace{\operatorname{tr}(\mathbf{A}\mathbf{S}^+)}_{B_\psi}. \tag{3.2.3}$$

The bias factor B_ψ can be expressed in two equivalent ways[1]. First, using the spectral decomposition $\mathbf{S}^+ = \sum_{k=2}^{J} \lambda_k^{-1} \mathbf{v}_k \mathbf{v}_k'$ with $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_J$,

$$B_\psi = \operatorname{tr}(\mathbf{A}\mathbf{S}^+) = \sum_{k=2}^{J} \frac{1}{\lambda_k} \mathbf{v}_k' \mathbf{A} \mathbf{v}_k, \tag{3.2.4}$$

where the last equality uses $\operatorname{tr}(\mathbf{A}\mathbf{v}_k\mathbf{v}_k') = \mathbf{v}_k'\mathbf{A}\mathbf{v}_k$. Second, following Andrews et al. (2008), we can write

$$B_\psi = \frac{1}{n} \operatorname{tr}(\mathbf{S}^+ \mathbf{N}_F) - \frac{1}{n^2} \mathbf{n}_F' \mathbf{S}^+ \mathbf{n}_F. \tag{3.2.5}$$

Substituting the spectral expansion of $\mathbf{S}^+$ into the first term gives

$$\operatorname{tr}(\mathbf{S}^+ \mathbf{N}_F) = \sum_{k=2}^{J} \frac{1}{\lambda_k} \mathbf{v}_k' \mathbf{N}_F \mathbf{v}_k.$$

Equation (3.2.4) shows that distortions are driven by the *reciprocal eigenvalues*, with weights $\mathbf{v}_k'\mathbf{A}\mathbf{v}_k$ determined by the alignment of the eigenvectors with the firm-size structure. When network connectivity is low, small eigenvalues dominate the sum, amplifying idiosyncratic noise and "dumping" it into the estimated variance components.

3.2.3 Scaling law: expected connectivity and N/J

This section establishes the mechanical link between the workers-per-firm ratio and the network spectrum. I model mobility as a time-homogeneous Markov chain with transition matrix $\mathbf{Q}$. Let $\mathbf{G} = \sum_{t=0}^{T-1} \mathbf{Q}^t$ be the finite-horizon fundamental matrix.

Proposition 2 (Connectivity scaling). *Consider a balanced panel of N workers observed over T periods at J firms ($T_i = T$ for all i). Let $\mathbf{S}$ denote the firm-level Laplacian defined in Proposition 1, and define the average node degree as $\bar{d} = \operatorname{tr}(\mathbf{S})/J$.*

(a) ***Expected connectivity.*** *Let n_k^{start} be the number of workers starting at firm k at $t = 1$, with $\sum_{k=1}^{J} n_k^{start} = N$. The expected average node degree is*

$$\mathrm{E}[\bar{d}] = \frac{1}{J} \sum_{j=1}^{J} \mathrm{E}[d_j] = \frac{1}{J} \sum_{j=1}^{J} \sum_{k=1}^{J} n_k^{start} \, C_{k \to j}, \tag{3.2.6}$$

where the per-worker contribution from a start at k to firm j is

$$C_{k \to j} = \frac{T-1}{T} G_{kj} - \frac{2}{T} \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} f_j(T-s), \qquad f_j(H) = \sum_{u=1}^{H} (\mathbf{Q}^u)_{jj}. \tag{3.2.7}$$

Here $\mathbf{Q}$ is the $J \times J$ transition matrix governing worker mobility, $\mathbf{G} = \sum_{t=0}^{T-1} \mathbf{Q}^t$ is the finite-horizon fundamental matrix, and $C_{k \to j}$ is the expected contribution to the degree of firm j from a single worker starting at firm k.

[1] The two expressions are equivalent because substituting $\mathbf{A} = \frac{1}{n}(\mathbf{N}_F - \frac{1}{n}\mathbf{n}_F\mathbf{n}_F')$ into $\operatorname{tr}(\mathbf{A}\mathbf{S}^+)$ gives $\frac{1}{n}\operatorname{tr}(\mathbf{S}^+\mathbf{N}_F) - \frac{1}{n^2}\mathbf{n}_F'\mathbf{S}^+\mathbf{n}_F$, using $\operatorname{tr}(\mathbf{n}_F\mathbf{n}_F'\mathbf{S}^+) = \mathbf{n}_F'\mathbf{S}^+\mathbf{n}_F$ and the cyclic property of the trace.

(b) ***Scaling with observations.*** *If the initial allocation is balanced ($n_k^{start} = N/J$ for all k),*

$$\mathrm{E}[\bar{d}] = \frac{N}{J} \cdot \frac{1}{J} \sum_{j=1}^{J} \sum_{k=1}^{J} \left[\frac{T-1}{T} G_{kj} - \frac{2}{T} \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} f_j(T-s) \right]. \tag{3.2.8}$$

Consequently, $\mathrm{E}[\bar{d}]$ *scales linearly with the workers-per-firm ratio* N/J*. Neglecting the return-probability correction (the term involving* f_j*), the leading-order behaviour is*

$$\mathrm{E}[\bar{d}] \;\approx\; \kappa \;\cdot\; (T-1)\,\frac{N}{J}, \tag{3.2.9}$$

where κ *is a constant that depends only on the stationary distribution implied by* $\mathbf{Q}$*. Thus the average connectivity is jointly determined by mobility dynamics (through* $\mathbf{Q}$ *and* T*) and sampling intensity (through* N/J*).*

(c) ***Inverse mapping.*** *In a connected regular network where every firm has identical observation counts, the bias factor* B_ψ *is the average of the reciprocal non-zero eigenvalues of* $\mathbf{S}$*:*

$$B_\psi = \frac{1}{J} \sum_{k=2}^{J} \lambda_k^{-1}. \tag{3.2.10}$$

Moreover, B_ψ *is bounded below by the reciprocal of average connectivity:*

$$B_\psi \geq \frac{(J-1)^2}{J^2 \bar{d}} \approx \frac{1}{\bar{d}}. \tag{3.2.11}$$

(d) ***Spectral Fragility à la Mohar.*** *The spectral gap* λ_2 *is fundamentally constrained by the firm with the lowest connectivity (Mohar, 1991):*

$$\lambda_2 \;\leq\; \frac{J}{J-1}\, d_{\min}. \tag{3.2.12}$$

Complete proof of the proposition can be found in Appendix A3.2.2.

The role of the Fiedler vector. The smallest positive eigenvalue λ_2 (the spectral gap) is often cited as a measure of network connectivity. However, the bias factor B_ψ depends on the entire spectrum, not only on λ_2. In particular, the contribution of the $k = 2$ term is $\lambda_2^{-1}(\mathbf{v}_2'\mathbf{A}\mathbf{v}_2)$. Whether this term dominates depends on whether the Fiedler vector $\mathbf{v}_2$ aligns with the firm-size structure. In empirically relevant cases with Zipf-distributed firm sizes, bottlenecks typically separate large clusters, so $\mathbf{v}_2'\mathbf{A}\mathbf{v}_2$ is bounded away from zero, making the $k = 2$ term of order $1/\lambda_2$. Nevertheless, the aggregate bias B_ψ is inversely proportional to the average node degree $\bar{d}$ rather than to λ_2 alone (as seen in the Monte Carlo evidence as well), because the sum over all eigenvalues is dominated by the bulk of the spectrum, which scales with $\bar{d}$.

The spectral representation of the Andrews bias factor in equation 3.2.5 is, to my knowledge, not stated explicitly in the prior literature. It reveals that the bias is not merely a function of "the number of movers" (the usual informal description of limited mobility bias) but of the *spectral structure* of the

mobility network: two networks with the same number of movers can produce very different biases depending on whether those movers connect weakly linked clusters (contributing to small eigenvalues) or reinforce already well-connected subgraphs. To see this concretely, consider two networks of $J = 100$ firms, each with 200 total movers. In network A, the movers are distributed evenly: each firm sends 2 movers to 2 random destinations. In network B, 50 firms form a dense cluster with 196 movers circulating internally, while 4 movers bridge the two halves of the network. Both networks have the same total number of movers, but network B has a much smaller λ_2 (and hence larger B_ψ) because the bridging movers create a severe bottleneck. Moreover, the bias factor B_ψ depends on the firm-size structure through $\mathbf{N}_F$ and the centering matrix $\mathbf{A}$. Low ratios of workers to firms amplify the bias independently of limited mobility, because fewer observations per firm increase $\text{tr}(\mathbf{S}^+\mathbf{N}_F)$ relative to the centering correction.

3.2.4 Mechanisms of noise amplification

The pipeline —

$$\left.\begin{array}{l} N/J \xrightarrow{\text{linear}} \\ \text{mobility} \xrightarrow{\text{via } \mathbf{Q}} \end{array}\right\} \bar{d} \xrightarrow{\text{inverse}} B_\psi \xrightarrow{\sigma_\varepsilon^2} \text{distortion,}$$

—explains why distortions persist in the wage variance decomposition even in a setup with high aggregate mobility. The main argument is that the whole network is "soft" when the average node degree $\bar{d}$ is low, so that the entire spectrum of $\mathbf{S}$ is compressed toward zero, leading to a large bias factor B_ψ and substantial noise amplification. The spectral gap λ_2 is a key bottleneck, but it is not catastrophic in itself. The bottleneck issue is most severe if the bulk of the other eigenvalues are also close to the smallest positive eigenvalue, which is the case when the average node degree is low. In contrast, if the average node degree is high, the bulk of the spectrum is pushed away from zero, so that even if a few firms are poorly connected or the raw mobility rate is low, the aggregate bias B_ψ remains small as long as $\bar{d}$ is large. The key takeaway is that the "limited mobility" problem is best understood not as a failure of connectivity ($\lambda_2 \to 0$) by itself, but as a steeper loss of spectral stiffness if the average node degree is low, which is the case when the workers-per-firm ratio N/J is small. In this regime, even a moderate mobility rate does not produce a well-connected network, and the bias factor B_ψ grows rapidly as N/J decreases.

Power-law growth of bias. While the benchmark case lower bound $B_\psi \gtrsim \frac{1}{\bar{d}}$ provides a theoretical floor, simulation evidence (Section 3.4) across uniform and Zipf-distributed firm sizes shows that B_ψ grows much faster than $1/\bar{d}$ as connectivity drops, following an empirical power law $B_\psi \approx a\bar{d}^{-b}$. For Zipf-distributed firm sizes, b is substantially larger than 1 (e.g., $b \approx 2$), while for uniformly sized firms and sufficiently large N/J, b approaches 1. This faster-than-linear growth arises because the spectral gap λ_2 is governed by the minimum degree $d_{\min}$ via the Mohar bound $\lambda_2 \leq \frac{J}{J-1} d_{\min}$, and because both $\bar{d}$ and $d_{\min}$ decline together as N/J falls. As the workers-per-firm ratio N/J decreases, both $\bar{d}$ and $d_{\min}$ fall — the latter is observed even at a 40% mobility rate per period in the simulations presented in section 3.4 and it is the quantity responsible for the collapse of λ_2. Consequently, the

sum of reciprocal eigenvalues in (3.2.4) explodes, driving B_ψ upwards. The power-law exponent $b > 1$ reflects the sensitivity of the bias to the lowest-degree firms, whose influence is diluted in the average $\bar{d}$ but amplified through the reciprocal eigenvalues.

Fixed-scale sparsity. Under the regime of *fixed-scale sparsity*, the workers-per-firm ratio N/J acts as the fundamental lever for the precision of the variance decomposition. As established in Proposition 2(a), the identifying mass of the network scales linearly with N/J, directly shifting the expected node degree $\mathrm{E}[D_j]$. This scaling determines the "stiffness" of the entire firm Laplacian. While the Mohar bound (3.2.12) ensures that the spectral gap λ_2 is bottlenecked by the least-connected firm ($d_{\min}$), a small λ_2 is not a catastrophic failure of the estimator so long as it is strictly positive.The true mechanism of noise amplification lies in the *bulk* of the spectrum. Since $B_\psi = \frac{1}{J}\sum \lambda_k^{-1}$, the bias is driven by the density of eigenvalues near zero. When N/J is low, the average connectivity $\bar{d}$ is small, effectively compressing the entire spectral distribution toward the origin. In this "soft" network, $d_{\min}$ not only collapses λ_2, but the lack of aggregate identifying mass ensures that the subsequent eigenvalues ($\lambda_3, \lambda_4, \dots$) are also small, causing the sum of reciprocals to explode. Conversely, in a network with high average connectivity, the bulk of the eigenvalues are pushed away from zero; even if a single bottlenecked firm keeps λ_2 small, its contribution to the trace is isolated, and the aggregate bias B_ψ remains controlled. Thus, the "limited mobility" problem is best understood not as a failure of connectivity at the weakest degree ($\lambda_2 \to 0$), but as a loss of spectral stiffness driven by the local sparsity of the firm graph at low workers-per-firm ratios. Low mobility at high average node degree does not produce the same level of distortion because the bulk of the spectrum is not compressed toward zero.

Finally, note that $\sigma_\varepsilon^2 > 0$ is a necessary condition for distortion. If the data were noiseless ($\sigma_\varepsilon^2 = 0$), the true parameters would be recovered regardless of the network's spectral properties. The "limited mobility" problem is therefore not a failure of identification (connectivity) per se (as long as we are using a connected set) but a failure of precision (noise amplification) driven by the local sparsity of the firm graph.

3.3 Bias Correction Methods

3.3.1 The KSS leave-one-out correction

The KSS leave-one-out estimator removes the first-order bias of the plug-in quadratic functional by replacing one factor with a leave-one-out prediction (Kline et al., 2020). For each observation (i, t) with firm assignment $j(i, t)$, let $\hat{\varepsilon}_{it} = w_{it} - \hat{\alpha}_i - \hat{\psi}_{j(i,t)}$ be the OLS residual and P_{it} the leverage. The leave-one-out firm effect is

$$\hat{\psi}_{-(it),j(i,t)} = \hat{\psi}_{j(i,t)} - \frac{[\mathbf{S}^+ \tilde{\mathbf{f}}_{it}]_{j(i,t)}\ \hat{\varepsilon}_{it}}{1 - P_{it}}, \tag{3.3.1}$$

where $\tilde{\mathbf{f}}_{it} \in \mathbb{R}^J$ is the within-worker demeaned firm indicator for observation (i, t), with entries

$$[\tilde{\mathbf{f}}_{it}]_j = \mathbf{1}\{j(i, t) = j\} - \frac{n_{ij}}{T_i}, \qquad j = 1, \dots, J.$$

Under arbitrary heteroskedasticity, the bias of the plug-in variance of firm effects is $\mathrm{tr}(\mathbf{A}\,\mathrm{Cov}(\hat{\boldsymbol{\psi}}))$, which is no longer proportional to a single σ^2. Define the observation-level variance proxy $\hat{\sigma}^2_{it} = w_{it}\,\hat{\varepsilon}_{it}/(1-P_{it})$, where w_{it} is the observed wage and $\hat{\varepsilon}_{it}$ is the OLS residual (Kline et al., 2020). The heteroskedasticity-robust matrix is then

$$\hat{\boldsymbol{\Omega}} = \sum_{i,t} \hat{\sigma}^2_{it}\,\tilde{\mathbf{f}}_{it}\tilde{\mathbf{f}}'_{it}.$$

The sandwich covariance estimator of $\hat{\boldsymbol{\psi}}$ is $\mathbf{C} = \mathbf{S}^{+}\hat{\boldsymbol{\Omega}}\,\mathbf{S}^{+}$, and the KSS bias correction for the variance of firm effects is

$$\widehat{B}^{\mathrm{KSS}}_{\psi} = \frac{1}{n}\,\mathrm{tr}(\mathbf{N}_F\,\mathbf{C}) - \frac{1}{n^2}\mathbf{n}'_F\mathbf{C}\,\mathbf{n}_F, \tag{3.3.2}$$

so that $\widehat{\mathrm{Var}}^{\mathrm{KSS}}_n(\psi) = \widehat{\mathrm{Var}}^{\mathrm{plug}}_n(\psi) - \widehat{B}^{\mathrm{KSS}}_{\psi}$.

Corrections for worker effects and covariance. Following the leave-out logic in Kline et al. (2020, Section 4), the sensitivities of the estimated parameters to observation (i,t) are:

$$\ell^{\psi}_{it} = [\mathbf{S}^{+}\tilde{\mathbf{f}}_{it}]_{j(i,t)}, \qquad \ell^{\alpha}_{it} = \frac{1}{T_i}\left(1 - \sum_{j\in\mathcal{J}(i)} n_{ij}\,[\mathbf{S}^{+}\tilde{\mathbf{f}}_{it}]_j\right),$$

where $\mathcal{J}(i) = \{j : n_{ij} > 0\}$ is the set of firms employing worker i. The KSS bias correction for the worker–firm covariance is then

$$\widehat{B}^{\mathrm{KSS}}_{\mathrm{cov}} = \frac{1}{n}\sum_{i,t} \hat{\sigma}^2_{it}\,\ell^{\psi}_{it}\,\ell^{\alpha}_{it}, \tag{3.3.3}$$

yielding $\widehat{\mathrm{Cov}}^{\mathrm{KSS}}_n(\alpha,\psi) = \widehat{\mathrm{Cov}}^{\mathrm{plug}}_n(\alpha,\psi) - \widehat{B}^{\mathrm{KSS}}_{\mathrm{cov}}$.

For the variance of worker effects, the KSS bias correction follows an analogous two-term structure that accounts for the centering of the estimated worker effects. Let $\delta_{it} \equiv \ell^{\alpha}_{it}$ be the worker-level leverage. The total bias in the plug-in variance of $\hat{\alpha}$ is

$$\widehat{B}^{\mathrm{KSS}}_{\alpha} = \frac{1}{n}\sum_{i,t} T_i\,\delta^2_{it}\,\hat{\sigma}^2_{it} \;-\; \frac{1}{n^2}\sum_{i,t} T_i^2\,\delta^2_{it}\,\hat{\sigma}^2_{it}. \tag{3.3.4}$$

The first term corresponds to the bias in an *uncentered* sum of squared worker effects; the second term corrects for the subtraction of the mean $\bar{\alpha}$. The KSS estimator of the worker-effect variance is then

$$\widehat{\mathrm{Var}}^{\mathrm{KSS}}_n(\alpha) = \widehat{\mathrm{Var}}^{\mathrm{plug}}_n(\alpha) - \widehat{B}^{\mathrm{KSS}}_{\alpha}.$$

This direct correction avoids reliance on the variance decomposition identity and provides unbiased estimates under arbitrary heteroskedasticity and design unbalance.

Spectral representation of the correction. From Proposition 1, $\mathbf{S}$ is a graph Laplacian with eigenvalues $0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_J$ and orthonormal eigenvectors $\mathbf{v}_1 = \mathbf{1}_J/\sqrt{J}, \mathbf{v}_2, \ldots, \mathbf{v}_J$. The

pseudoinverse has the spectral decomposition

$$\mathbf{S}^{+} = \sum_{k=2}^{J} \lambda_k^{-1} \mathbf{v}_k \mathbf{v}_k'.$$

Hence the vector in the correction term expands as

$$\mathbf{S}^{+}\tilde{\mathbf{f}}_{it} = \sum_{k=2}^{J} \frac{\mathbf{v}_k' \tilde{\mathbf{f}}_{it}}{\lambda_k} \mathbf{v}_k.$$

Its $j(i,t)$-th component is therefore

$$[\mathbf{S}^{+}\tilde{\mathbf{f}}_{it}]_{j(i,t)} = \sum_{k=2}^{J} \frac{(\mathbf{v}_k' \tilde{\mathbf{f}}_{it})\, v_{k,j(i,t)}}{\lambda_k},$$

where $v_{k,j}$ is the j-th entry of $\mathbf{v}_k$. The magnitude of the correction is controlled by the smallest eigenvalue λ_2: by the Cauchy–Schwarz inequality,

$$\left|[\mathbf{S}^{+}\tilde{\mathbf{f}}_{it}]_{j(i,t)}\right| \le \|\mathbf{S}^{+}\|_2 \, \|\tilde{\mathbf{f}}_{it}\| = \frac{\|\tilde{\mathbf{f}}_{it}\|}{\lambda_2}.$$

Leverage and network bottlenecks. The leverage P_{it} is the diagonal entry of the projection matrix onto the column space of $[\mathbf{D}\ \mathbf{F}]$. In the concentrated model, it satisfies $P_{it} = \tilde{\mathbf{f}}_{it}' \mathbf{S}^{+} \tilde{\mathbf{f}}_{it}$ (see Kline et al. (2020, Appendix)). Using the spectral expansion,

$$P_{it} = \sum_{k=2}^{J} \frac{(\mathbf{v}_k' \tilde{\mathbf{f}}_{it})^2}{\lambda_k}.$$

When the network has a bottleneck (small λ_2), the term for $k = 2$ dominates both P_{it} and the correction numerator, causing $1 - P_{it}$ to be near zero for observations that straddle the bottleneck. Thus the denominator in (3.3.1) amplifies the correction.

Sampling variability of the correction. Although the KSS estimator is exactly unbiased, its sampling variance depends on the spectral structure of the Laplacian. When λ_2 is small, the sandwich matrix $\mathbf{C} = \mathbf{S}^{+} \hat{\mathbf{\Omega}}\, \mathbf{S}^{+}$ has entries that scale as $O(1/\lambda_2^2)$[2], and the variance of the bias correction can dominate the standard $O(1/n)$ rate. This leads to highly dispersed estimates of the worker and covariance shares despite unbiasedness, particularly in poorly connected networks. The KSS correction therefore cannot fully stabilize the variance decomposition when λ_2 is small; the instability is ultimately governed by the spectral gap of the mobility Laplacian.

3.3.2 The split-sample estimator

The split-sample estimator used in the simulations follows Babet et al. (2025), who split *workers* into two independent groups. For each firm j, a random half of its workers are assigned to group o and

[2]this is not catastrophic by itself unless the bulk of the eigenvalues are also small and close to λ_2 which can happen if either mobility is low or the bulk of firm sizes is low in terms of the number of weighted observations, or both.

the remaining half to group 1; all observations of a given worker travel with their group, so each worker's full career history is preserved. AKM is then estimated independently on each group, yielding $\hat{\boldsymbol{\psi}}^{(0)}, \hat{\boldsymbol{\alpha}}^{(0)}$ and $\hat{\boldsymbol{\psi}}^{(1)}, \hat{\boldsymbol{\alpha}}^{(1)}$. Because the two estimation samples are drawn from independent subsets of individuals, the estimation errors of $\hat{\boldsymbol{\alpha}}$ in one sample are orthogonal to those of $\hat{\boldsymbol{\psi}}$ in the other, which removes the limited-mobility bias on all cross-sample quadratic forms. The estimator exploits this independence to debias three components of the variance decomposition.

Variance of firm effects. Let the common Largest Connected Set (LCS) be the intersection of the largest connected sets of the two split samples (i.e., firms that belong to the LCS of group 0 and also to the LCS of group 1). The debiased estimator of $\mathrm{Var}(\psi)$ is the cross-sample covariance of firm effects, evaluated over all observations at common-LCS firms:

$$\widehat{\mathrm{Var}}_n(\psi)_{\text{split}} = \widehat{\mathrm{Cov}}_n\big(\hat{\psi}^{(0)}_{j(i,t)},\ \hat{\psi}^{(1)}_{j(i,t)}\big). \tag{3.3.5}$$

Covariance of worker and firm effects. For each observation of a group-g worker, the "own" worker effect $\hat{\alpha}^{(g)}_i$ is paired with the "other" firm effect $\hat{\psi}^{(1-g)}_{j(i,t)}$. The debiased covariance estimator is

$$\widehat{2\,\mathrm{Cov}}_n(\alpha, \psi)_{\text{split}} = 2\widehat{\mathrm{Cov}}_n\big(\hat{\alpha}^{(\text{own})}_i,\ \hat{\psi}^{(\text{other})}_{j(i,t)}\big), \tag{3.3.6}$$

where the covariance is taken over all common-LCS observations. Since $\hat{\alpha}^{(g)}$ and $\hat{\psi}^{(1-g)}$ are estimated on disjoint sets of workers, their estimation errors are independent, and the cross-product is unbiased for $\mathrm{Cov}(\alpha, \psi)$.

Variance of the residual. For each worker i in group g, define the partially residualised wage $\tilde{w}_{it} = w_{it} - \hat{\psi}^{(1-g)}_{j(i,t)}$. In population, $\tilde{w}_{it} = \alpha_i + \varepsilon_{it} + (\psi_{j(i,t)} - \hat{\psi}^{(1-g)}_{j(i,t)})$, and the last term vanishes in large samples because the firm-effect estimation error from the independent other split is $O_p(1/\sqrt{n})$. The within-worker variance of $\tilde{w}_{it}$ therefore consistently estimates $\mathrm{Var}(\varepsilon)$:

$$\widehat{\mathrm{Var}}_n(u)_{\text{split}} = \frac{1}{|\mathcal{W}|}\sum_{i\in\mathcal{W}} \widehat{\mathrm{Var}}_t(\tilde{w}_{it}), \tag{3.3.7}$$

where $\mathcal{W}$ is the set of workers with at least two observations in the common LCS and $\widehat{\mathrm{Var}}_t$ denotes the within-worker sample variance (with Bessel correction).

Variance of worker effects (by subtraction). Because each worker's effect is estimated in only one split, the plug-in $\mathrm{Var}(\hat{\alpha})$ from a single split-sample is *not* directly debiased. The variance of worker effects is instead recovered by subtraction, exploiting the accounting identity $\mathrm{Var}(w) = \mathrm{Var}(\alpha) + \mathrm{Var}(\psi) + \mathrm{Var}(\varepsilon) + 2\,\mathrm{Cov}(\alpha, \psi)$:

$$\widehat{\mathrm{Var}}_n(\alpha)_{\text{split}} = \widehat{\mathrm{Var}}_n(w) - \widehat{\mathrm{Var}}_n(\psi)_{\text{split}} - \widehat{\mathrm{Var}}_n(u)_{\text{split}} - 2\widehat{\mathrm{Cov}}_n(\alpha, \psi)_{\text{split}}, \tag{3.3.8}$$

which is consistent under the strict exogeneity assumption $\mathrm{E}[\varepsilon_{it} \mid i, j(i,t)] = 0$, which implies $\mathrm{Cov}(\varepsilon, \alpha) = \mathrm{Cov}(\varepsilon, \psi) = 0$ and hence the exact variance decomposition $\mathrm{Var}(w) = \mathrm{Var}(\alpha) + \mathrm{Var}(\psi) + \mathrm{Var}(\varepsilon) + 2\,\mathrm{Cov}(\alpha, \psi)$.

Spectral interpretation: decoupling noise from network structure. The superiority of the split-sample estimator is most transparent when viewed through the spectral lens of Section 3.2. For the plug-in estimator, the bias in $\widehat{\mathrm{Var}}_n(\psi)$ arises from the *interaction* between the idiosyncratic noise $\boldsymbol{\varepsilon}$ and the geometry of the firm mobility network. Writing $\hat{\boldsymbol{\psi}} = \boldsymbol{\psi} + \mathbf{S}^{+}\mathbf{F}'\mathbf{M}_D\boldsymbol{\varepsilon}$, the expectation of the quadratic form is

$$\mathrm{E}\left[\hat{\boldsymbol{\psi}}'\mathbf{A}\hat{\boldsymbol{\psi}}\right] = \boldsymbol{\psi}'\mathbf{A}\boldsymbol{\psi} + \sigma_\varepsilon^2 \,\mathrm{tr}\left(\mathbf{A}\mathbf{S}^{+}\right) = \boldsymbol{\psi}'\mathbf{A}\boldsymbol{\psi} + \sigma_\varepsilon^2 \sum_{k=2}^{J} \frac{\mathbf{v}_k'\mathbf{A}\mathbf{v}_k}{\lambda_k}.$$

The bias is non-zero *only because* $\sigma_\varepsilon^2 > 0$ *and* the reciprocal eigenvalues of the Laplacian $\mathbf{S}$ amplify that noise. When N/J is small or mobility is low, the average node degree $\bar{d}$ falls, the spectral gap λ_2 shrinks, and the sum of reciprocals explodes, injecting a large positive distortion into $\widehat{\mathrm{Var}}_n(\psi)$ and a corresponding negative distortion into $\widehat{\mathrm{Cov}}_n(\alpha, \psi)$.

The split-sample estimator breaks this noise-network coupling entirely. Write the firm effects from the two independent splits as $\hat{\boldsymbol{\psi}}^{(g)} = \boldsymbol{\psi} + \boldsymbol{\varepsilon}^{(g)}$, where $\boldsymbol{\varepsilon}^{(g)} = \mathbf{S}_{(g)}^{+}\mathbf{F}_{(g)}'\mathbf{M}_{D,(g)}\boldsymbol{\varepsilon}^{(g)}$. The cross-sample covariance is

$$\widehat{\mathrm{Var}}_n(\psi)_{\text{split}} = \hat{\boldsymbol{\psi}}^{(0)\prime}\mathbf{A}\hat{\boldsymbol{\psi}}^{(1)} = \boldsymbol{\psi}'\mathbf{A}\boldsymbol{\psi} + \boldsymbol{\psi}'\mathbf{A}\boldsymbol{\varepsilon}^{(1)} + \boldsymbol{\varepsilon}^{(0)\prime}\mathbf{A}\boldsymbol{\psi} + \boldsymbol{\varepsilon}^{(0)\prime}\mathbf{A}\boldsymbol{\varepsilon}^{(1)}.$$

Taking expectations, the bias term is $\mathrm{E}\left[\boldsymbol{\varepsilon}^{(0)\prime}\mathbf{A}\boldsymbol{\varepsilon}^{(1)}\right]$. Because the two splits are constructed from *independent* random halves of the workers, the error vectors $\boldsymbol{\varepsilon}^{(0)}$ and $\boldsymbol{\varepsilon}^{(1)}$ are uncorrelated. Consequently,

$$\mathrm{E}\left[\boldsymbol{\varepsilon}^{(1)}\boldsymbol{\varepsilon}^{(0)\prime}\right] = \mathbf{S}_{(1)}^{+}\mathbf{F}_{(1)}'\mathbf{M}_{D,(1)} \underbrace{\mathrm{E}[\boldsymbol{\varepsilon}^{(1)}\boldsymbol{\varepsilon}^{(0)\prime}]}_{=\mathbf{0}} \mathbf{M}_{D,(0)}\mathbf{F}_{(0)}\mathbf{S}_{(0)}^{+} = \mathbf{0}.$$

The bias vanishes identically, regardless of the value of σ_ε^2 *or the spectral properties of* $\mathbf{S}_{(g)}$. The split-sample estimator does not merely reduce the amplification of noise by the small eigenvalues of the Laplacian; it *orthogonalises* the estimation errors so that the product term has zero expectation. In the noiseless limit ($\sigma_\varepsilon^2 = 0$), all consistent estimators coincide, but whenever $\sigma_\varepsilon^2 > 0$, the split-sample approach alone achieves exact bias removal for the cross-sample quadratic forms, without requiring the estimation of leverage scores or the inversion of an ill-conditioned matrix.

The only remaining finite-sample distortion is the *composition effect* induced by the common Largest Connected Set restriction: when N/J is very small, some firms drop out of the LCS in one or both splits, and the common-LCS subsample ceases to be fully representative of the population. This selection bias is orthogonal to the noise-amplification channel analysed above; it can be monitored by reporting the fraction of firms (or worker-years) retained in the common LCS. Once the sample is large enough that the common LCS covers nearly all firms, the split-sample estimator is unbiased for all quadratic forms, even in networks whose Laplacian spectrum would cause severe distortions for the plug-in, Andrews, or KSS estimators.

3.4 Monte Carlo Evidence

3.4.1 Design

Let:

- $i = 1, \ldots, N$ index workers,
- $t = 1, \ldots, T$ index time periods ($T = 5$),
- $j = 1, \ldots, J$ index firms,
- $j(i, t)$ denote the firm where worker i works at time t.

The Monte Carlo study is built to isolate the key sources of identification weakness in two-way fixed-effect variance decompositions: the workers-per-firm ratio N/J, the share of movers (mobility rate p), the number of firms J, and the heterogeneity in firm exposure. Two mobility regimes are considered: *uniform*, where a mover's destination is drawn uniformly from all other firms, and *size-weighted*, where the destination probability is proportional to firm size under a Zipf distribution.

Firm effects

Firm effects are drawn i.i.d. and then centered:

$$\psi_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\psi^2), \quad \sigma_\psi^2 = 0.50, \qquad \widetilde{\psi}_j = \psi_j - \frac{1}{J}\sum_{k=1}^{J} \psi_k.$$

For simplicity, I denote the centered effects still as ψ_j.

Worker effects with sorting

Worker effects are correlated with the initial firm's effect:

$$\alpha_i = \rho \, \frac{\sigma_\alpha}{\sqrt{\widehat{\text{Var}}(\psi_{j(i,0)})}} \, \psi_{j(i,0)} \; + \; \sqrt{1-\rho^2} \, \eta_i,$$

where $\eta_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\alpha^2)$, $\sigma_\alpha^2 = 0.50, \rho = 0.3$, and $\widehat{\text{Var}}(\psi_{j(i,0)})$ is the realized cross-worker variance of the initial firm effects in the current draw.[3]

Mobility process

Each period $t \geq 1$, a random subset of $M = \max(\lfloor N \cdot p \rfloor, 1)$ workers is selected to move, where $p \in \{0.05, 0.10, 0.20, 0.40\}$ is the mobility rate per period. For each mover:

[3]Because $\eta_i \perp \psi_{j(i,0)}$, $\text{Var}(\alpha_i) = \rho^2\sigma_\alpha^2 + (1-\rho^2)\sigma_\alpha^2 = \sigma_\alpha^2$ exactly, for every design cell and replication, regardless of J or the firm-size distribution. The implied sorting is $\text{Cov}(\alpha_i, \psi_{j(i,0)}) = \rho\,\sigma_\alpha\sqrt{\widehat{\text{Var}}(\psi_{j(i,0)})}$.

- With **uniform mobility**, the destination firm is drawn uniformly from $\{1, \ldots, J\}$ (re-drawn if it equals the current firm).
- With **size-weighted mobility**, the destination firm is drawn with probability proportional to the current firm sizes (computed from the initial assignment).

All other workers stay in the same firm: $j(i,t) = j(i,t-1)$.

Wage equation

The log wage is constructed as:

$$w_{it} = \psi_{j(i,t)} + \alpha_i + \varepsilon_{it}, \qquad \varepsilon_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2), \quad \sigma_\varepsilon^2 = 0.25.$$

Mobility, exogeneity, and the benchmark for distortion

The mobility process described above is independent of both the worker type α_i and the destination firm effect $\psi_{j(i,t)}$ for $t \geq 1$. This guarantees strict exogeneity,

$$\mathbb{E}\left[\varepsilon_{it} \,\middle|\, \boldsymbol{\alpha}, \boldsymbol{\psi}, \{j(i,s)\}_{s=0}^{T}\right] = 0,$$

by construction. Sorting enters the design only through the initial assignment $j(i,0)$, and is mechanically attenuated in the panel as workers reshuffle: a worker who has not yet moved by period t contributes a sorted observation, while a worker who has moved contributes one in which her firm is independent of α_i. The realized panel covariance $\widehat{\text{Cov}}(\alpha_i, \psi_{j(i,t)})$ is therefore smaller in absolute value than the initial sorting parameter $\rho\,\sigma_\alpha\sqrt{\widehat{\text{Var}}(\psi_{j(i,0)})}$, with the gap widening in both p and T.

To handle this cleanly, the "true" variance shares reported in Section 3.4.2 are recomputed on the realized LCS-restricted panel:

$$\frac{\widehat{\text{Var}}(\psi_{j(i,t)})}{\widehat{\text{Var}}(w_{it})}, \qquad \frac{\widehat{\text{Var}}(\alpha_i)}{\widehat{\text{Var}}(w_{it})}, \qquad \frac{2\,\widehat{\text{Cov}}(\alpha_i, \psi_{j(i,t)})}{\widehat{\text{Var}}(w_{it})},$$

so that every distortion is measured against the same population the estimators actually see. The benchmark is the realized panel moment, not the parameter ρ.

This design follows the standard AKM Monte Carlo convention, in which mobility is drawn independently of unobservables (Andrews et al., 2008), and it is chosen deliberately. Any distortion documented in Section 3.4.2 therefore arises *within* the AKM identifying assumptions, isolating the noise-amplification mechanism analysed in Section 3.4.2.3 from confounding bias due to endogenous mobility. The chapter's central claim—that bias corrections fail when $\bar{d}$ is low even though the model is correctly specified—is in this sense conservative: it documents a failure mode that survives the most favourable environment AKM can be evaluated in.

Largest connected set and estimation sample

Before estimation, the panel is restricted to the *largest connected set* (LCS) of firms linked by worker mobility. A firm belongs to the LCS if and only if it can be reached from every other firm in the set through a chain of movers. All "true" variance shares reported below—$\mathrm{Var}(\psi)/\mathrm{Var}(w)$, $\mathrm{Var}(\alpha)/\mathrm{Var}(w)$, and $2\,\mathrm{Cov}(\alpha, \psi)/\mathrm{Var}(w)$—are computed on this LCS-restricted sample, ensuring that distortions measure bias relative to the same population used for estimation. Each replication records the fraction of firms, workers, observations, and movers retained by the LCS, which varies sharply with the design parameters.

Parameter summary

$$\sigma_\psi^2 = 0.50, \quad \sigma_\alpha^2 = 0.50, \quad \sigma_\varepsilon^2 = 0.25, \quad \rho = 0.3, \quad T = 5, \quad p \in \{0.05, 0.10, 0.20, 0.40\}.$$

The number of workers N depends on firm counts $J \in \{50, 100, 500\}$ and the chosen firm-size distribution:

- Uniform: $N = J \times$ (workers per firm), with workers per firm $\in \{5, 10, 15, 30, 50, 100, 500\}$.
- Zipf: N is the sum of Zipf-distributed firm sizes, trimmed to match the target total.

The full design crosses 2 mobility types × 4 rates × 3 firm counts × 2 firm size distribution types × 7 workers-per-firm levels = 336 design cells, each run with $R = 10$ replications, yielding 3,360 total draws.

3.4.2 Results

3.4.2.1 Average Node Degree as a Summary of Network Stiffness

Figure 3.1 displays the average node degree $\bar{d} = \mathrm{tr}(\mathbf{S})/J$, computed from the same Monte-Carlo experiments described in Section 3.4. The two horizontal axes sweep the per-period mobility rate from 5 % to 40 % and the workers-per-firm ratio N/J from 5 to 500. The vertical axis plots the $\bar{d}$ resulting from those designs. A power-law surface is fitted to the observations of the form, $\bar{d} = a \times (N/J)^b \times m\%^c$.

All the fitted surfaces have $R^2 \approx 0.999$. The coefficients b, c determine how $\bar{d}$ changes with N/J and $m\%$ respectively. Four features of the surface are immediately apparent and confirm the theoretical predictions of Section 3.2.3.

- *First*, $\bar{d}$ behaves linearly ($b \approx 1$) with N/J for a fixed $m\%$, consistent with the scaling law in Proposition 2(b): each additional worker contributes a roughly constant expected amount of identifying mass, the magnitude of which is governed by mobility dynamics through the matrix $\mathbf{Q}$. Higher the $m\%$, the steeper the slope.

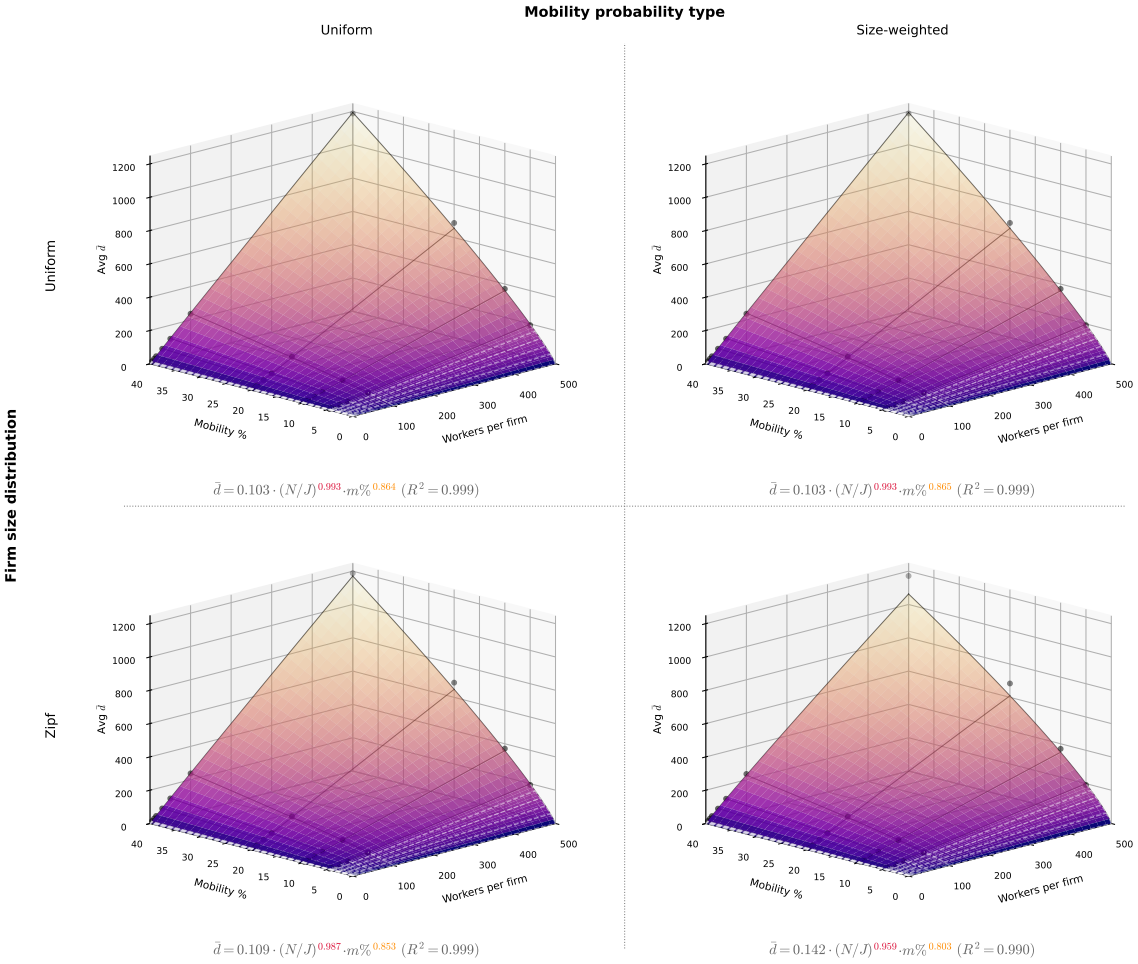


Figure 3.1. Average node degree as a function of per-period mobility and workers per firm under all designs split by two mobility types and two firm size distributions. The vertical axis represents the average of the effective realized node degree per Definition 2 The dashed region is extrapolated.

- *Second*, $\bar{d}$ also increases with the mobility rate at a given N/J. But since the coefficient $c \approx 0.85$, $\bar{d}$ rises slowly with increasing $m\%$. This also means in the very low mobility region, the returns to increased mobility is higher. At $N/J = 10$, moving from 5 % to 40 % mobility lifts $\bar{d}$ from about 40 to several hundred, reflecting the greater accumulation of mover-periods.
- *Third*, the surface approaches 0 when *both* N/J and $m\%$ are low —exactly the parameter range where the plug-in AKM estimator exhibits severe positive bias in the firm-effect variance and severe attenuation of the sorting term (see the distortion table 3.1, for example).
- *Finally*, in the low N/J region, increasing mobility barely helps recover the $\bar{d}$, but at low $m\%$ region, increasing the N/J can significantly improve $\bar{d}$.

This behavious can also be explained by the spectral bias factor $B_\psi = \mathrm{tr}(\mathbf{AS}^+)$ introduced in Equation (3.2.3). When $\bar{d}$ falls below roughly unity, the entire spectrum of the firm Laplacian $\mathbf{S}$ is compressed toward the origin: not only does the spectral gap λ_2 collapse in accordance with the Mohar bound $\lambda_2 \leq \frac{J}{J-1} d_{\min}$, but the *bulk* of subsequent eigenvalues $(\lambda_3, \lambda_4, \dots)$ are also small. Because B_ψ is the average of the

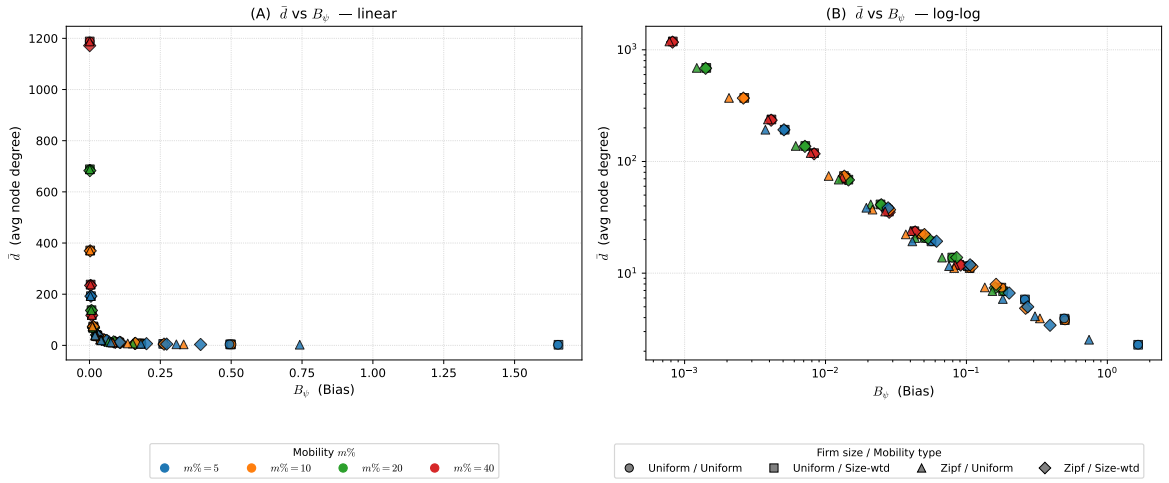


Figure 3.2. Average bias over all design cells as a function of the average node degree $\bar{d}$. The right panel uses log-log axes, revealing a power-law relationship between the bias factor B_ψ and the average node degree $\bar{d}$. Each point corresponds to a unique combination of mobility rate, firm size distribution, number of firms, and workers-per-firm ratio. The tight clustering of points along a single curve confirms that $\bar{d}$ fully captures the influence of these design parameters on the bias.

reciprocal non-zero eigenvalues (Eq. 3.2.4), this spectral compression drives B_ψ upward, amplifying the residual noise σ_ε^2 in all estimated fixed effects. Conversely, once $\bar{d}$ exceeds a modest threshold, the bulk eigenvalues are pushed sufficiently far from zero that the bias becomes negligible, even if a few peripheral firms still exhibit small individual degrees. Thus, $\bar{d}$ acts as a parsimonious sufficient statistic for the *spectral stiffness* of the firm mobility network: it summarizes the combined influence of panel length, firm size, and mobility on the precision of the AKM variance decomposition.

Figure 3.1 therefore provides the empirical foundation for the pipeline

$$\left.\begin{array}{l} N/J \xrightarrow{\text{linear}} \\ \text{mobility} \xrightarrow{\text{via } \mathbf{Q}} \end{array}\right\} \bar{d} \xrightarrow{\text{inverse}} B_\psi \xrightarrow{\sigma_\varepsilon^2} \text{distortion},$$

which explains why the "limited-mobility" problem is, fundamentally problematic for inference under a low-average-degree setup rather than a mere matter of the raw number of movers. Both the workers-per-firm ratio and the per-period mobility rate contribute to $\bar{d}$, but through distinct channels: the former enters linearly, as shown in Proposition 2(b), while the latter operates through the transition matrix $\mathbf{Q}$ and the finite-horizon path sums. Crucially, their combined effect on the bias is fully mediated by the average node degree, as the scatter in Figure 3.2 will confirm. Two networks with the same total number of mover-periods can have dramatically different average node degrees—and hence dramatically different biases—depending on how those movers are distributed across firms (the concentration effect already illustrated in the Zipf-vs-uniform comparison of Section 3.2.3). The 3-D surface makes this connection visually explicit and motivates the remainder of the empirical analysis, which traces the propagation of low $\bar{d}$ into biases in each term of the variance decomposition.

Figure 3.2 collapses the full design space of the Monte-Carlo exercise—varying mobility rates, firm-size distributions, numbers of firms, and workers-per-firm ratios—into a single scatter of the empirical bias factor B_ψ against the average node degree $\bar{d}$. The right panel uses log-log axes, revealing that the points lie on a tight, monotonically decreasing curve that is well approximated by a power law $B_\psi \propto \bar{d}^{-b}$. The remarkable feature of the plot is that *every* design combination falls onto the same locus: conditional on $\bar{d}$, neither the per-period mobility rate nor N/J carries any independent predictive power for the bias. A network with 5% mobility and a large workforce per firm can therefore exhibit a smaller bias than one with 40% mobility but a very thin workforce, because it is the resulting average node degree—not the raw fraction of movers—that determines the spectral stiffness of the firm Laplacian. The log-log scale makes the fragility of the low-$\bar{d}$ regime especially vivid: when $\bar{d}$ drops below unity, B_ψ exceeds 1, so that the bias alone injects more variance into the between-firm component than the signal itself. Conversely, once $\bar{d}$ surpasses a modest threshold (roughly 5–10), further increases in mobility or sample size yield sharply diminishing returns, as the bulk of the Laplacian's eigenvalues are already comfortably away from zero. The figure thus provides strong empirical support for the theoretical pipeline of Section 3.2.3, in which N/J and mobility operate exclusively through their product-like contribution to $\bar{d}$, and $\bar{d}$ summarises all relevant information about the network's capacity to resist noise amplification.

3.4.2.2 Variance Decomposition Distortions

Table 3.1 reports the average distortion (difference in the estimated variance component from true variance component, expressed in percentages) under 5% per-period mobility (which aggregates to about 18.5% aggregate mobility) with Zipf-distributed firm sizes and size-weighted mobility. This is a more realistic configuration since firm sizes are known to follow Zipf's law and mobility is size-weighted assuming these unequal firm sizes are more or less preserved over a finite time horizon. Three findings stand out:

Finding 1: HO and KSS correct the firm-effect bias relative to the plug-in estimator but show relatively similar distortions in worker-effect and covariance shares. At $J = 100$ with 5 workers per firm, the plug-in overstates the firm-effect share by 23%. The homoskedastic correction (HO) proposed by Andrews et al. (2008) reduces this to 2% and KSS to 7%. Yet the worker-effect distortion remains at +34% (plug-in) versus +33% (KSS), and the covariance distortion stays at −100% for all three traditional estimators. Although KSS computes corrections for all three components, the worker-effect correction $\widehat{B}_\alpha^{\text{KSS}} = \frac{1}{n}\,\text{tr}(\mathbf{C})$ and the covariance correction $\widehat{B}_{\text{cov}}^{\text{KSS}}$ accumulate only over movers; at 5% mobility, nearly all observations are stayers for whom these terms are zero. The corrections are therefore negligible relative to the dominant distortion, which has two distinct sources identified in Section 3.4.2.3: the bias B_ψ dumped into the worker variance through the OLS back-substitution, and the incidental-parameter bias σ_ε^2/T from estimating N worker effects from T observations each.

Table 3.1. Variance-share distortions from true shares at 5% mobility (Zipf firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	HO	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	0.80	0.26	3.01	+34	0	+9	-43	+28	+27	-8	-108	-109	-9
10	0.80	0.38	4.61	+14	-6	-1	-21	+27	+26	-1	-52	-52	+10
15	0.80	0.50	6.32	+11	-3	-1	-6	+17	+17	-4	-38	-38	-6
30	1.04	0.93	11.74	+10	+3	+3	+2	+14	+14	-2	-24	-24	-4
50	1.72	1.61	19.23	+5	+1	+1	0	+12	+12	-1	-13	-13	-1
100	5.28	5.21	38.39	+2	0	0	0	+11	+11	-1	-5	-5	+1
500	36.80	36.80	192.03	+1	+1	+1	+1	+10	+10	0	-2	-2	-1
Panel B: $J = 100$													
5	0.80	0.22	3.37	+23	+2	+7	+10	+30	+29	-5	-75	-75	-18
10	0.80	0.28	4.93	+15	-2	+2	-7	+24	+24	-2	-54	-54	-2
15	0.80	0.36	6.58	+11	-1	0	-6	+21	+21	-1	-38	-39	-4
30	0.80	0.73	11.77	+7	0	+1	-4	+15	+15	-1	-21	-22	-1
50	1.00	0.91	19.27	+5	+1	+1	+1	+12	+12	-2	-12	-13	-3
100	3.44	3.23	38.44	+1	0	0	-1	+12	+11	0	-5	-5	+1
500	29.64	29.35	192.33	0	0	0	0	+10	+10	0	-1	-1	0
Panel C: $J = 500$													
5	0.80	0.13	3.87	+17	-3	+2	-14	+27	+26	-3	-63	-64	-2
10	0.80	0.17	5.44	+11	-1	+2	-4	+21	+20	-2	-42	-43	0
15	0.80	0.25	7.00	+12	+1	+3	-2	+19	+19	-1	-36	-36	-4
30	0.80	0.32	12.11	+6	0	+1	-2	+15	+15	-1	-19	-19	0
50	0.80	0.51	19.43	+3	0	0	0	+13	+13	-1	-12	-12	-1
100	1.24	1.17	38.49	+2	0	0	0	+11	+11	0	-5	-5	0
500	18.00	17.89	192.35	0	0	0	0	+10	+10	0	-1	-1	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_\psi = \sigma^2_\alpha = 0.50$, $\sigma^2_\varepsilon = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct* $\widehat{\mathrm{Var}}(\alpha)$ *by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.*

Finding 2: A persistent ~10% worker-effect bias survives at all sample sizes. Even at 5,000 workers per firm—where the mobility network is dense and the firm-effect bias vanishes ($B_\psi \approx 0$) thanks to the inverse relationship with average node degree—every estimator except the split reports a worker-effect distortion of approximately +10%. This is the $O(1/T)$ incidental-parameter component identified below in Section 3.4.2.3: with $\sigma^2_\varepsilon = 0.25$ and $T = 5$, the excess is $\sigma^2_\varepsilon/T = 0.05$, which as a share of $\mathrm{Var}(w)$ is roughly 4%. This bias does not vanish as $N \to \infty$ with T fixed, and neither Andrews nor KSS removes it. Only the split-sample estimator eliminates it, because it recovers $\mathrm{Var}(\alpha)$ by subtraction through an independent estimate of $\mathrm{Var}(\varepsilon)$.

Finding 3: The covariance is the most severely distorted component. The plug-in $2\widehat{\mathrm{Cov}}(\alpha, \psi)$ share is biased downward by up to 109% ($J = 50$, 5 workers per firm), entirely obscuring the true positive sorting ($\rho = 0.3$). This distortion is a direct consequence of the zero-sum OLS mechanics derived in Section 3.4.2.3: the upward bias in both the firm and worker shares must be offset by a downward bias in the covariance, since the three shares sum to $\mathrm{Var}(w)$. Andrews and KSS do not

correct this channel. The split-sample estimator substantially reduces the covariance distortion (to −61% at the worst cell), though residual attenuation persists when the half-sample networks are severely fragmented.

The average node degree $\bar{d}$ as the operative diagnostic.

Tables A3.4-A3.6 in the appendix report the corresponding results at different per-period mobility rates under size-weighted mobility for both uniform distribution of firm sizes and zipf firm sizes.

3.4.2.3 Sparse networks and the allocation of bias

The bias factor B_ψ governs the magnitude of the firm-effect bias, but it does not fully characterize how this estimation error propagates through the variance decomposition. To explain the empirical regularity that the upward bias in the worker-effect variance is positively correlated with the firm-effect bias, I examine the local algebraic constraints imposed by the OLS normal equations.

By the Frisch–Waugh–Lovell theorem, the estimated worker effects $\hat{\boldsymbol{\alpha}}$ are backed out as residuals after solving for $\hat{\boldsymbol{\psi}}$. For any worker i, the first-order condition with respect to α_i dictates that the mean residual for that worker is zero:

$$\hat{\alpha}_i = \bar{w}_i - \frac{1}{T_i} \sum_{t=1}^{T_i} \hat{\psi}_{j(i,t)},$$

where $\bar{w}_i = T_i^{-1} \sum_t w_{it}$ is the worker's time-average log wage.

In sparse or low-mobility networks, the vast majority of workers are stayers who never change firms. Let $\mathcal{N}_S \subset \{1, \dots, N\}$ denote the set of stayers, such that for all $i \in \mathcal{N}_S, j(i,t) = j$ for all t. For these workers, the back-substitution simplifies to $\hat{\alpha}_i = \bar{w}_i - \hat{\psi}_j$. Substituting the true DGP $\bar{w}_i = \alpha_i + \psi_j + \bar{\varepsilon}_i$ and defining the firm-effect estimation error $\eta_j = \hat{\psi}_j - \psi_j$ yields the exact finite-sample error allocation for stayers:

$$\hat{\alpha}_i - \alpha_i = -\eta_j + \bar{\varepsilon}_i. \tag{3.4.1}$$

This identity reveals a strict zero-sum mechanical linkage: for stayers, any estimation error in the firm effect (η_j) maps deterministically and inversely into the worker effect.

We can now examine how this error allocation distorts the quadratic functionals. Evaluating the variance over the set of stayers (which dominates the observation-weighted variance in low-mobility samples), and using the orthogonality of $\bar{\varepsilon}_i$ to the fixed effects under strict exogeneity, the estimated variance of worker effects expands as

$$\widehat{\mathrm{Var}}(\alpha) \;\approx\; \mathrm{Var}(\alpha) + \mathrm{Var}(\eta) + \mathrm{Var}(\bar{\varepsilon}) - 2\,\mathrm{Cov}(\alpha, \eta).$$

Because OLS forces η to absorb the finite-sample noise, $\mathrm{Var}(\eta) \approx \sigma^2 B_\psi$ as defined in (3.2.3). Under the independence of α and η, the bias in the worker variance share has two additive components:[4]

$$\mathbb{E}\big[\widehat{\mathrm{Var}}(\alpha)\big] - \mathrm{Var}(\alpha) \;\approx\; \underbrace{\sigma^2 B_\psi}_{\text{limited-mobility bias}} \;+\; \underbrace{\sigma_\varepsilon^2 / T}_{\text{incidental-parameter bias}} . \tag{3.4.2}$$

The first term is the bias factor derived under homoskedasticity by Andrews et al. (2008) in (3.2.3). Its magnitude is governed by the spectral sum $B_\psi = \sum_{k=2}^{J} \lambda_k^{-1} (\mathbf{v}_k' \mathbf{A} \mathbf{v}_k)$. When $\bar{d}$, the average node degree is low, the entire spectrum of the firm Laplacian $\mathbf{S}$ is compressed toward zero, so that not only the spectral gap λ_2 but also the bulk of subsequent eigenvalues are small. Because B_ψ is the average of the reciprocal non-zero eigenvalues, this spectral compression drives B_ψ upward, amplifying the residual noise σ_ε^2 in all estimated fixed effects. In the extreme case where $\lambda_2 \to 0$, B_ψ grows as $O(1/\lambda_2)$. Consequently, these small eigenvalues directly amplify the bias dumped into $\widehat{\mathrm{Var}}(\alpha)$ and $2\widehat{\mathrm{Cov}}(\alpha, \psi)$. The second term is an $O(1/T)$ incidental-parameter bias from estimating N worker effects from T observations each: $\mathrm{Var}(\bar{\varepsilon}_i) = \sigma_\varepsilon^2 / T$, which does not vanish as $N \to \infty$ with T fixed. Table 3.1 confirms this decomposition: at large N/J where $B_\psi \approx 0$, a residual +10% worker-effect distortion persists, corresponding to $\sigma_\varepsilon^2 / T = 0.05$ as a share of $\mathrm{Var}(w) \approx 4\%$.

The algebraic dumping also attenuates the covariance. For stayers, under the same independence assumptions and $\mathrm{Cov}(\psi, \eta) = 0$ (which holds under strict exogeneity since $\eta = \mathbf{S}^{+}\mathbf{F}'\mathbf{M}_D\boldsymbol{\varepsilon}$):

$$\mathbb{E}\big[\widehat{\mathrm{Cov}}(\alpha, \psi)\big] \;\approx\; \mathrm{Cov}(\alpha, \psi) - \sigma^2 B_\psi .$$

Consequently, in sparse networks where node degrees are low and B_ψ is large, correcting only the firm variance leaves the system fundamentally unbalanced. The uncorrected noise $\sigma^2 B_\psi$ remains dumped in the worker variance, and the covariance is driven toward zero or negative, entirely obfuscating the true sorting parameter.

3.5 Discussion

The Monte Carlo experiments deliver three main results. First, the bias in the plug-in AKM variance decomposition is governed not by the raw number of movers but by the *average node degree* $\bar{d} = \mathrm{tr}(\mathbf{S})/J$, a single scalar that captures the joint influence of firm size, panel length, and mobility. When $\bar{d}$ is low, the entire eigenvalue spectrum of the firm Laplacian is compressed toward zero, so that the bias factor B_ψ is large and a substantial fraction of the idiosyncratic noise is dumped into the estimated worker-effect variance and the sorting covariance is biased downwards.

[4] The decomposition (3.4.2) is exact for stayers. For movers, additional cross-terms between $\hat{\alpha}_i$ and the firm effects at multiple employers appear, but their contribution is of order $O(p)$ where p is the mobility rate, and is therefore negligible in the low-mobility regimes that are the focus of this paper. The incidental-parameter component σ_ε^2 / T is the standard finite-T bias from estimating N individual intercepts; see Nickell (1981) for the general result and Andrews et al. (2008) for its application to two-way fixed-effect models.

Second, none of the correction methods eliminates this bias when $\bar{d}$ is low. The homoskedastic correction and the KSS leave-out estimator both reduce the firm-effect bias relative to the plug-in, yet Table 3.1 shows that at $J = 50$ and $N/J = 5$ ($\bar{d} \approx 3$) the KSS firm-effect distortion is still +9%, and the worker-effect distortion remains +27% under the given design. The split-sample estimator is designed to avoid the spectral amplification entirely, but it introduces a different finite-sample problem: when N/J is very small, many firms drop out of the common Largest Connected Set, so the cross-sample covariance estimates the variance of a non-representative subset of firms, yielding a *negative* distortion (-43% in the same cell). For the worker-effect variance and the sorting covariance, the situation is worse: the KSS corrections for $\mathrm{Var}(\alpha)$ and $2\,\mathrm{Cov}(\alpha, \psi)$ accumulate almost no bias correction from stayers, so they leave the OLS dumping mechanism essentially intact. Consequently, at $\bar{d} \lesssim 5$ in the simulations, the estimated sorting term is so severely attenuated that a true $\rho = 0.3$ positive association is completely masked.

Third, even when $\bar{d}$ is large enough that the firm-effect bias becomes negligible ($B_\psi \approx 0$), a $O(1/T)$ incidental-parameter bias of about 10% persists in $\widehat{\mathrm{Var}}(\alpha)$ for all methods except the split-sample estimator, which backs out the worker variance by subtraction using an independent estimate of $\mathrm{Var}(\varepsilon)$. This floor cannot be reduced by adding more firms; only longer panels or smaller residual variance can attenuate it.

These findings have direct practical consequences. Researchers should report the average node degree $\bar{d}$ along with the bias factor B_ψ as standard pre-estimation diagnostics. Given what we believe the noise variance to be in applied settings, a small $\bar{d}$ renders the estimates of $\mathrm{Var}(\alpha)$ and $2\,\mathrm{Cov}(\alpha, \psi)$ are unreliable, and the sorting covariance in particular should not be interpreted as evidence of assortative sorting. In such settings, restricting the sample to firms above a minimum size or explicitly modelling the noise amplification (e.g. through a regularized Laplacian) may be necessary to obtain credible decompositions.

Several caveats apply. The simulations use homoskedastic Gaussian errors and an additive wage model; real earnings data contain heteroskedasticity, measurement error, and worker-firm interactions. The KSS estimator is robust to heteroskedasticity in principle, but its finite-sample variance can be large when $\bar{d}$ is low and the bulk of eigenvalues are close to zero, making the corrections noisy. The split-sample estimator circumvents this at the cost of halving the sample which is sometimes not feasible in smaller datasets. Neither approach directly resolves the incidental-parameter bias without additional information on $\mathrm{Var}(\varepsilon)$. Future work could investigate whether graph-based regularization or the use of auxiliary moments (e.g. from stayers) can stabilize the worker and covariance components.

3.6 Conclusion

This paper shows that the reliability of the AKM variance decomposition is fundamentally limited by the average node degree of the implied firm-mobility network. When the workers-per-firm ratio is low, the Laplacian's spectrum collapses toward zero, and all plug-in variance components are severely

distorted. State-of-the-art correction methods—Andrews, KSS, and BGP split-sample—reduce but do not eliminate these distortions at low $\bar{d}$; each method faces its own finite-sample fragility, whether from noisy leverage corrections, poor connectivity, or common-LCS selection. The split-sample estimator uniquely breaks the spectral noise-amplification channel and additionally removes the incidental-parameter bias through its subtraction-based recovery of $\mathrm{Var}(\alpha)$, making it the most accurate of the four estimators when the common LCS is representative.

The key takeaway is that "limited mobility" as a "problem" is a symptom of a deeper structural feature: low average firm size. Two easy-to-compute diagnostics, $\bar{d}$ and B_ψ, summarize the severity of the problem before estimation. Whenever $\bar{d}$ is small, empirical studies should treat the AKM sorting covariance with extreme caution and consider sample restrictions or supplementary methods to separate signal from noise. These recommendations are increasingly relevant as administrative matched employer-employee datasets expand in scale.

Bibliography

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

John M. Abowd, Robert H. Creecy, and Francis Kramarz. Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau, Mar 2002. URL https://ideas.repec.org/p/cen/tpaper/2002-06.html.

John M. Abowd, Kevin L. McKinney, and Ian M. Schmutte. Modeling endogenous mobility in earnings determination. *Journal of Business & Economic Statistics*, 37(3):405–418, July 2019. doi: 10.1080/07350015.2017.1356727. URL https://ideas.repec.org/a/taf/jnlbes/v37y2019i3p405-418.html.

Martyn J. Andrews, Leonard Gill, Thorsten Schank, and Richard Upward. High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A*, 171(3):673–697, 2008.

Damien Babet, Olivier Godechot, and Marco Palladino. In the land of akm: The performance of the akm estimator in finite samples. *Journal of Labor Economics*, 83:102388, 2023.

Damien Babet, Olivier Godechot, and Marco G. Palladino. In the land of akm: Explaining the dynamics of wage inequality in france. Technical Report 987, Banque de France, 2025. URL https://ideas.repec.org/p/bfr/banfra/987.html.

Stéphane Bonhomme, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41(2):291–322, 2023.

David Card, Jörg Heining, and Patrick Kline. Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics*, 128(3):967–1015, 2013.

David Card, Ana Rute Cardoso, and Patrick Kline. 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):633–686, None 2016. doi: None. URL https://ideas.repec.org/a/oup/qjecon/v131y2016i2p633-686..html.

David Card, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):13–70, None 2018. doi: 10.1086/694153. URL https://ideas.repec.org/a/ucp/jlabec/doi10.1086-694153.html.

Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

Shmuel Friedland and Reinhard Nabben. On Cheeger-type inequalities for weighted graphs. *Journal of Graph Theory*, 41(1):1–17, 2002.

Koen Jochmans and Martin Weidner. Fixed-effect regressions on network data. *Econometrica*, 87(5): 1543–1560, 2019.

Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898, September 2020.

Bojan Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, editors, *Graph Theory, Combinatorics, and Applications*, volume 2, pages 871–898. Wiley, 1991.

Stephen Nickell. Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–1426, 1981.

Jae Song, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. Firming up inequality. *Quarterly Journal of Economics*, 134(1):1–50, 2019.

Isaac Sorkin. Ranking firms using revealed preference. *The Quarterly Journal of Economics*, 133(3): 1331–1393, 2018.

Simon D. Woodcock. Match effects. *Research in Economics*, 69(1):100–121, 2015.

Appendix to Chapter 1

A1.1 Panel Construction

- Keep observations of people aged 18-60 and atleast 3 observed spells (including non-employment).
- Keep employment spells that are at least 6 months (180 days) in a year. Else, code as non-employment. Recall in DADS this employment durations include paid days/time off.
- Non-employment is imputed for the two following cases:
  (1) if someone has missing years in between employment spells.
  (2) if someone who would be younger than 60 in 2019 vanishes from the panel.
  Every year from 2012-2019, about half of the non-employment spells have some non-employment compensation (chomage indemnisee).
- In constructing a panel that has one spell per individual per year, the highest paying spell was kept (note only employment spells of atleast 6 months are kept).
- Winsorize log net real annual earnings (variable `s_net`) by state and year: cut and replace top and bottom 1%.
- **Note:** Full-time hours in DADS is 1820 hours annually (this figure translates to a 35-hour workweek). This is because in DADS paid time off is also included in employment durations and in hours; so, 35-hour workweeks over the 52 weeks in a year gives us $35 * 52 = 1820$ hours annually.

Table A1.2. Classification of French Degrees and U.S. Equivalents

Category	Degree	U.S. Equivalent	Sample Classification
1	Aucun Diplôme déclaré	No Degree declared	
2	CEP, DFEO	Elementary or Middle School	
3	BEPC, BE, BEPS	High School	
4	CAP, BEP, EFAA, BAA, BPA, FPA 1er	Vocational Technical School (Basic)	
5	Baccalauréats technique et professionnel, brevet professionnel, autres brevets (BEA, BEC, BEH, BEI, BES, BATA), baccalauréat général, brevet supérieur, CFES	High school diploma (General or Technical/Vocational)	Low
6	Santé, BTS, DUT, DEST, DEUL, DEUS, DEUG	Associate's degree or equivalent vocational degree	Medium
7	2ème cycle, 3ème cycle, Grande école, CAPES, CAPET	Graduate School and Other Post-Secondary Education	High

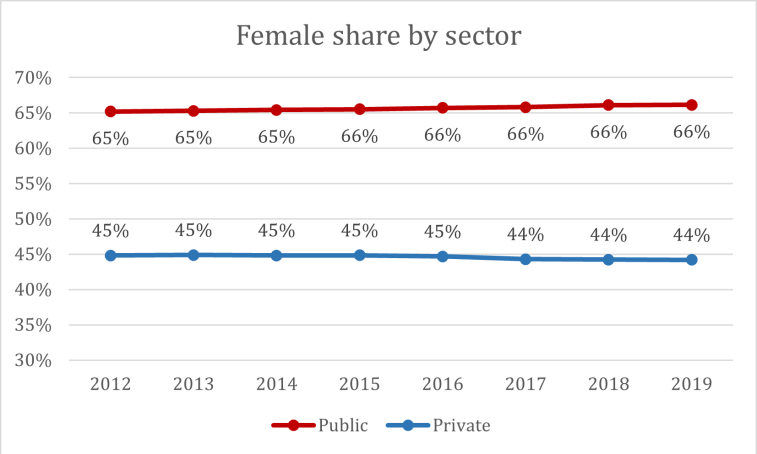
A1.2 Descriptives

Public Share of Employment

We observe that the public share of total employment in the sample remains fairly stable with an overall gradual increase of approximately 1.4% rising from a little under 23% close to 24.6% over the 8-year period. The slight but stable increase in the public share of employment is in-line with official INSEE statistics and corroborated in this sample panel.

A1.2.1 Additional demographics in the Public and Private Sectors

Figure A1.1. Share of women in the Public and Private Sectors over 2012-19



The share of employment in both sectors is stable across the years observed. The public sector, including part-time and full-time work averages at 24.3%, with small deviations. Since this article focuses on the intersection with gender, the sex ratio in both sectors, by full-time status is of key importance. Women hold 60% of the full-time public sector jobs in contrast to the 39% in the private sector. Interestingly, the part-time work landscape is very different as most of these jobs are done by women. Over the years, the female share of part-time jobs rises in both sectors .

The stability of the sex ratio (of females to males) in the public and private sectors irrespective of part-time full-time worker status is captured by Figure A1.1 generated from the sample. In France, the sex ratio is roughly 65:35 in the public sector and 45:55 in private. Curiously, women's earnings in the public sector is more compressed than that of men and also peaks later for the mean and above-mean income earners. The picture in the private sector by sex is even more grim. The dispersion of women's earnings far exceeds that of men. The lowest decile earners across sex experience a significant fall in earnings towards retirement. However, the fall not only starts relatively earlier in women's trajectories, but is far steeper across income percentiles. The below table depicts which sector, gender, and age group all the individuals observed in the panel (using their first spell) belong to:

Table A1.3. Distribution of Education Levels by Sector, Sex and Age Groups at first observed spell

Sex	Education	Public Sector			Private Sector		
		Under 30	31-45	Over 45	Under 30	31-45	Over 45
Females	Low	15	36	49	23	37	40
	Medium	19	48	34	27	51	22
	High	17	58	26	31	52	17
Males	Low	15	38	46	26	39	35
	Medium	15	48	37	26	51	23
	High	12	53	35	27	50	23

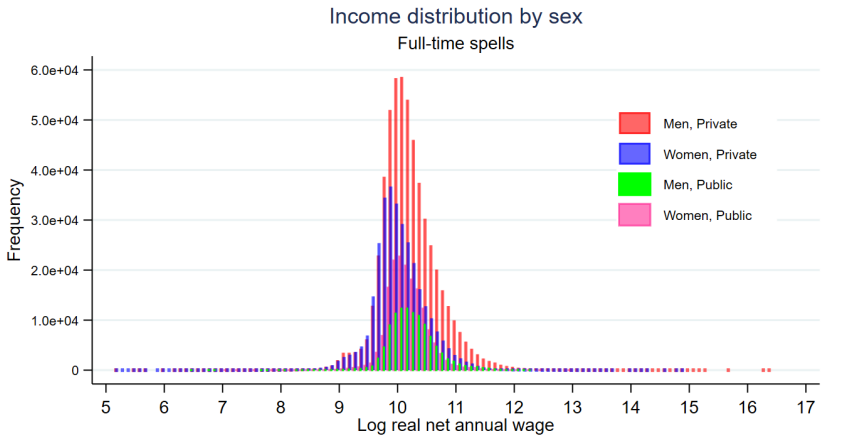


Figure A1.2. Histogram of log net annual wages by sex and sector for full-time work

Income and non-employment Statistics

The non-employment or inactivity remained mostly stable during this period at about 30%. Since it is imputed from the dataset it underestimates the rate for the first two years but catches up thereafter. Figure A1.2 shows the income distribution of men and women in the public and private sectors over the whole sample. Recall that this includes part-time workers. There are many noteworthy points in this distribution : firstly, the dominance of males in the private sector and of females in the public sector is evident. Secondly, as expected, the spread of income is more dispersed in the private sector while it is relatively more compressed in the public sector. Thirdly, the private sector distribution of income of males to the right of the mean dominates that of women; implying that a higher number of men earn more than the average than do women: notice the blue bars to the right of the mean in the figure. This phenomenon is reversed in the public sector distributions where more women earn the mean and above mean incomes. But, since the public sector is relatively smaller than the private, the public distributions are much shorter than their private counterparts. This income distribution provides initial evidence for the average earnings of females being lower than for men in the private sector.

Table A1.4. OLS regressions on real net wages - alternative specification

VARIABLES	Net annual	Net hourly
Pub FT	0.100***	-0.006
	(0.005)	(0.004)
Pvt PT	-0.636***	-0.118***
	(0.018)	(0.010)
Pub PT	-0.443***	-0.097***
	(0.009)	(0.005)
Female	-0.238***	-0.171***
	(0.011)	(0.008)
Pub FT#1.female	0.117***	0.059***
	(0.011)	(0.008)
Pvt PT#1.female	0.117***	0.071***
	(0.028)	(0.016)
Pub PT#1.female	0.337***	0.121***
	(0.013)	(0.009)
Experience (over a decade)	0.374***	0.275***
	(0.002)	(0.001)
Experience over a decade squared	-0.065***	-0.042***
	(0.000)	(0.000)
Medium educ	0.288***	0.264***
	(0.001)	(0.001)
High educ	0.537***	0.529***
	(0.001)	(0.001)
Constant	9.498***	2.229***
	(0.006)	(0.004)
Observations	1,389,551	1,385,227
Adjusted R-squared	0.454	0.387
Year fixed effects	✓	✓
Contract x state x female fixed effects	✓	✓

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A1.3 Model

This subsection lays out the equations that are similar in spirit to Postel-Vinay and Turon (2007) and Dickson et al. (2014) - the only difference being the data structure and the employment *state* variable.

A1.3.1 Unobserved Heterogeneity

The probability of an individual belonging in a particular class given their observed individual heterogeneity z_i^f is given as

$$\Pr\{k_i^y, k_i^m | z_i^f\} = \Pr\{k_i^y | k_i^m, z_i^f\} \cdot \Pr\{k_i^m | z_i^f\} \tag{A1.3.1}$$

In this analysis, given four employment states, 4 transition classes & 3 income classes are settled upon after trying various other combinations, which give the best compromise between analytic efficiency and accuracy of model fit. Note that both k^m and k^y are categorical variables. Both components of (A1.3.1) are expressed as multinomial logits with $K^y = 3$ and $K^m = 4$ outcomes respectively. Based on a standard multinomial logit classification, the functional form of the probabilities of an individual

belonging to these latent classes can be described as follows:

$$\Pr\{k_i^m = k^m \mid z_i^f\} = \frac{\exp\left(z_i^{f\prime}\kappa_{k^m}^m\right)}{\sum_{k=1}^{K^m}\exp\left(z_i^{f\prime}\kappa_k^m\right)},$$
$$\Pr\{k_i^y = k^y \mid k_i^m, z_i^f\} = \frac{\exp\left(\begin{pmatrix} z_i^f \\ k_i^m \end{pmatrix}'\kappa_{k^y}^y\right)}{\sum_{k=1}^{K^y}\exp\left(\begin{pmatrix} z_i^f \\ k_i^m \end{pmatrix}'\kappa_k^y\right)}, \tag{A1.3.2}$$

where k_i^m takes the values 1, 2, 3 or 4 for an individual i. Similarly, k_i^y takes the values 1, 2 or 3 for an individual i. Here, κ_k^m and κ_k^y are the vectors of coefficients of respective multinomial logits for the k^{th} outcome. Taking 1 as the base outcome in each multinomial logit, κ_1^m and κ_1^y are normalized to zero.

A1.3.2 Transitions in labor market states and Transition classes: K^m

Moving on to the second component of the individual likelihood equation (1.3.1), which encompasses the information in the model relating to the transitions between the various labor market states. Labor market states at year-t are assumed to depend on the individual's previous labor market state and their observed and unobserved heterogeneity, i.e., it follows a conditional first-order Markov chain. As there are five possible states, the probability of the individual being in one of the five states is modeled as a multinomial logit where the explanatory variables are the previous labor market state and the individual's unobserved and observed heterogeneity. Thus, at any given year (t > 1),

$$\Pr\{S_{i,t} = S_s | S_{i,t-1}, z_{i,t-1}^v, z_i^f, k_i^m\} = \frac{\exp\left(\begin{pmatrix} S_{i,t-1} \\ z_{i,t-1}^v \\ z_i^f \\ k_i^m \end{pmatrix}'\chi_{S_s}\right)}{\sum_{S=0}^{4}\exp\left(\begin{pmatrix} S_{i,t-1} \\ z_{i,t-1}^v \\ z_i^f \\ k_i^m \end{pmatrix}'\chi_S\right)} \tag{A1.3.3}$$

This equation is valid for all years except for the first year as the first state is the initial value. The initial state is specified as being dependent only on the observed and unobserved heterogeneity and thus, similarly modeled as a multinomial logit:

$$\Pr\{S_{i,1} = S_s | z_i^f, k_i^m\} = \frac{\exp\left(\begin{pmatrix} z_i^f \\ k_i^m \end{pmatrix}'\chi_{S_s}^o\right)}{\sum_{S=0}^{4}\exp\left(\begin{pmatrix} z_i^f \\ k_i^m \end{pmatrix}'\chi_S^o\right)} \tag{A1.3.4}$$

$[\chi_s]_{s=0}^{S-1}$ and $[\chi_s^o]_{s=0}^{S-1}$ are the vectors of coefficients of the respective multinomial logits for the S^{th} state outcome. χ_0 and χ_0^o are the base outcomes and normalized at zero.

Given these two specifications, we can now construct the (conditional) likelihood function of observing an individual's employment state trajectory over the years as follows:

$$\ell_i(S_{it}|z^v_{it}, z^f_i, k^m_i) = \Pr\{S_{i1}|z^f_i, k^m_i\} \times \prod_{t=2}^{T_i} \Pr\{S_{it}|S_{i,t-1}, z^v_{i,t-1}, z^f_i, k^m_i\} \tag{A1.3.5}$$

A1.4 Estimating the mobility and income parameters using sequential EM

A sequential, limited information version of the EM algorithm inspired by Bonhomme and Robin (2006) is employed to obtain a consistent set of estimates. In the previous sections, one can see all the parameters of the model that need to be estimated. The entire list of parameters can be split into two: $\Theta^m = \left\{(\kappa^m_k)^{K^m-1}_{k=0}, (\chi_s)^4_{s=0}, (\chi^0_s)^4_{s=0}\right\}$ and $\Theta^y = \left\{(\kappa^y_k)^{K^y-1}_{k=0}, \mu, \sigma, \xi\right\}$. To recollect, we get the Θ^m parameters from the state mobility equations (A1.3.2), (A1.3.3) and (A1.3.4); and we get the Θ^y parameters from the income process equations (A1.3.2), (1.3.4), (1.3.5), (1.3.8). The structure of the complete individual likelihood equation (1.3.1) means that it can be decomposed into two parts as shown below:

$$\mathcal{L}_i(\mathbf{x}_i, k_i; \Theta^m, \Theta^y) = \mathcal{L}^m_i(\mathbf{x}_i, k^m_i; \Theta^m) \times \mathcal{L}^y_i(\mathbf{x}_i, k^m_i, k^y_i; \Theta^y) \tag{A1.4.1}$$

$$\text{where } \mathcal{L}^m_i(\mathbf{x}_i, k^m_i; \Theta^m) = \ell^m_i\left(\mathbf{S}_i|\mathbf{z}^v_i, z^f_i, k^m_i; \Theta^m\right) \cdot \Pr\left\{k^m_i|z^f_i; \Theta^m\right\},$$
$$\text{and } \mathcal{L}^y_i(\mathbf{x}_i, k^m_i, k^y_i; \Theta^y) = \ell^y_i\left(\mathbf{y}_i|S_i, \mathbf{z}^v_i, z^f_i, k^m_i, k^y_i; \Theta^y\right) \cdot \Pr\left\{k^y_i|k^y_i, z^f_i; \Theta^m\right\}$$

This structure makes it easier to separate income sequences ($\mathbf{y_i}$) and income classes (k^y_i) from the *state* transition part of the likelihood function, $\mathcal{L}^m_i(\mathbf{x}_i, k^m_i; \Theta^m)$. The parameters pertaining to the job mobility process and job transition classes can thus be recovered by separately considering the likelihood of observed job sector mobility: $\sum_{i=1}^N \ln\left(\sum_{k^m_i=1}^{K^m} \mathcal{L}^m_i(\mathbf{x}_i, k^m_i; \Theta^m)\right)$. The maximization can be achieved by applying the EM algorithm for finite mixtures. To make the process more efficient, the parameters for the *state* transitions part of the likelihood equation are first estimated. This yields an initial estimate of $\hat{\Theta}^m$ for the mobility parameters. Then, Θ^m is fixed at $\hat{\Theta}^m$, and the estimation proceeds to the income parameters Θ^y. After reaching a suitable convergence in terms of distance between the parameters from subsequent iterations, an initial estimate $\hat{\Theta}^y$ of Θ^y is obtained. Finally, using $\hat{\Theta}^m$ and $\hat{\Theta}^y$ as initial values, the complete likelihood function (given by equation (1.3.1) and (A1.4.1)) is estimated, allowing all parameters to vary, to arrive at the final estimates of all the mobility and income parameters.

A1.4.1 Calibrating the initial job mobility parameters Θ^m

- **E-Step**: For an initial value Θ^m_n of Θ^m, for each transition class $k^m = 1, ..., K^m$ and for each individual i in the sample, the posterior probability that i belongs to transition class k^m given

$\mathbf{x}_i$ and Θ_n^m would be:

$$\Pr\{k_i^m = k^m | \mathbf{x}_i; \Theta_n^m\} = \frac{\mathcal{L}_i^m(\mathbf{x}_i, k_i^m; \Theta_n^m)}{\sum_{k=1}^{K^m} \mathcal{L}_i^m(\mathbf{x}_i, k_i^m; \Theta_n^m)} \tag{A1.4.2}$$

These probabilities are calculated by estimating the likelihoods using using the functional forms of equations (A1.3.2), (A1.3.3) and (A1.3.4).

- **M-Step**: We now update Θ_n^m into Θ_{n+1}^m by maximizing the following augmented sample log-likelihood, weighted by the E-Step equation above:

$$\Theta_{n+1}^m = \arg\max_{\Theta^m} \sum_{i=1}^{N} \sum_{k_i^m=1}^{K^m} \Pr\left\{k_i^m = k | \mathbf{x}_i; \Theta_n^m\right\} \cdot \ln\left[\mathcal{L}_i^m(\mathbf{x}_i, k_i^m; \Theta_n^m)\right] \tag{A1.4.3}$$

This maximization is carried out by creating 4 copies of out data set, one for each possible transition class that an individual can be in. Then we carry out weighted multinomial logits of k^m, *State*, and $State_{initial}$ using equations (A1.3.2), (A1.3.3) and (A1.3.4) weighing each copy of the data by its prior probability obtained in the E-step for the corresponding transition class k^m.

We continue iterations till the Euclidean distance between the coefficients in subsequent iterations falls below 10^{-3} as is standard in the literature. We can now use this initial estimate $\hat{\Theta}^m$ to calibrate the income process estimations in the next step.

A1.4.2 Calibrating the income parameters Θ^y

- **E-Step**: For an initial value Θ_n^y of Θ^y, for each class index $k = (k^m, k^y)$, $k^m = 1, ..., K^m$; $k^y = 1, ..., K^y$, and for each individual i in the sample, compute the posterior probability that i belongs to transition class k^m and income class k^y given $\mathbf{x}_i$, Θ_n^y and $\hat{\Theta}^m$:

$$\Pr\left\{k_i^m = k^m, k_i^y = k^y | \mathbf{x}_i; \hat{\Theta}^m, \Theta_n^y\right\} = \frac{\mathcal{L}_i\left(\mathbf{x}_i, k^m, k^y; \hat{\Theta}^m, \Theta_n^y\right)}{\sum_{\ell^m=1}^{K^m} \sum_{\ell^y=1}^{K^y} \mathcal{L}_i\left(\mathbf{x}_i, \ell^m, \ell^y; \hat{\Theta}^m, \Theta_n^y\right)} \tag{A1.4.4}$$

- **M-Step**: In this step we update the μ, σ, ξ and κ^y coefficients. The following process is followed:
  (1) We first create 12 copies of our dataset. One for each k^y and k^m class combination that every individual can be in. Each class combination is weighted by the class probabilities obtained from equation (A1.4.4).
  (2) We then update the income mean parameter $\mu(\cdot)$ using weighted OLS regressions of y_{it} on $\left(S_{it}, z_{it}^v, z_i^f, k_i^y\right)$, using the 12 class probabilities obtained from equation (A1.4.4) in the E-Step as weights for each of the 12 copies of the data. The vector of coefficients of this regression thus becomes the updated $\hat{\mu}^{n+1}(\cdot)$.

(3) Similarly, the log squared residuals from the regression in step 2 are taken and regressed on $\left(S_{it}, z_{it}^{v}, z_{i}^{f}, k_{i}^{y}\right)$, again using weighted OLS, to update income variance parameter $\hat{\sigma}^{n+1}(\cdot)$.

(4) Log income disturbances are updated as:

$$\tilde{y}_{it}^{n+1} = \frac{y_{it} - \hat{\mu}_{n+1}(S_{it}, z_{it}^{v}, z_{i}^{f}, k_{i}^{y})}{\hat{\sigma}_{n+1}(S_{it}, z_{it}^{v}, z_{i}^{f}, k_{i}^{y})}$$

Next, we can generate the $\tau_{i,t,t-1}$ of $\tilde{y}_{it}$ by taking into account all past income streams up until the year t. So, for each year, the $Cov(\tilde{y}_{it}^{n+1}, \tilde{y}_{i,t-1}^{n+1})$ measure is expected to be the outcome of the $\tau_{1}^{n+1}(\cdot)$ function. Thus, we can update $\hat{\tau}_{1}^{n+1}(\cdot)$ by taking the inverse functional form of τ_{1}^{n+1} that was given in (1.3.8). To do this, a new variable f_{it} is defined such that:

$$f_{it} = ln\left(\frac{1 + Cov(\tilde{y}_{it}^{n+1}, \tilde{y}_{i,t-1}^{n+1})}{1 - Cov(\tilde{y}_{it}^{n+1}, \tilde{y}_{i,t-1}^{n+1})}\right)$$

We then regress f_{it} on $S_{it}, S_{i,t-1}, z_{it}^{v}, k_{i}^{m}, k_{i}^{y}, S_{it} * k_{i}^{y}, S_{i,t-1} * k_{i}^{y}$, again using equation (A1.4.4) as weights. The vector of coefficients thus obtained is the updated ξ^{n+1}.

(5) Finally we can update the set of income class assignment parameters, $\left(\kappa_{k}^{y}\right)_{k=0}^{K^{y}-1}$ by running a weighted multinomial logit regression of class indices on $\left(z_{i}^{f}, k_{i}^{m}\right)$, again by using the (A1.4.4) as weights.

A1.4.3 Likelihood Maximization

After obtaining the initial estimates of the Θ^{m} and Θ^{y} parameters separately, we then proceed to update the combined log-likelihood given by the equation:

$$\text{likelihood: } \prod_{i=1}^{N} \prod_{k_{i}^{m}=1}^{K^{m}} \prod_{k_{i}^{y}=1}^{K^{y}} \mathcal{L}_{i}\left[\mathbf{x}_{i}, (k_{i}^{m}, k_{i}^{y})\right] \tag{A1.4.5}$$

$$\text{log-likelihood: } \sum_{i=1}^{N} \ln\left(\sum_{k_{i}^{m}=1}^{K^{m}} \sum_{k_{i}^{y}=1}^{K^{y}} \mathcal{L}_{i}\left[\mathbf{x}_{i}, (k_{i}^{m}, k_{i}^{y})\right]\right) \tag{A1.4.6}$$

We carry out this estimation using a similar methodology employed in the initialization phase.

Table A1.5. Demographic and Educational Composition by Transition Class (*km*)

Transition	Female	Low ed	Med ed	High ed	Age <31	Age 31–45	Age >45
$km = 0$	0.65	0.51	0.16	0.32	0.25	0.43	0.32
$km = 1$	0.66	0.71	0.29	0.00	0.32	0.40	0.28
$km = 2$	0.33	0.86	0.14	0.00	0.32	0.39	0.29
$km = 3$	0.56	0.68	0.10	0.22	0.23	0.41	0.36

Note: Means by class from the latent model output. Classes differ markedly in observed characteristics—e.g., gender and education—but overlap substantially across those dimensions, suggesting compositional clustering rather than latent behavioral types.

A1.5 Model Fit

A1.5.1 Additional Tables and Figures for Model Fit and Unobserved Heterogeneity

Table A1.6. Transition Probabilities by Latent Transition Class (km) – Observed Data

From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
Transition Class 0					
NE	0.81	0.10	0.02	0.05	0.01
Pvt FT	0.08	0.89	0.00	0.03	0.00
Pub FT	0.02	0.01	0.93	0.00	0.04
Pvt PT	0.15	0.12	0.00	0.72	0.01
Pub PT	0.08	0.01	0.14	0.01	0.76
Transition Class 1					
NE	0.82	0.10	0.02	0.05	0.01
Pvt FT	0.09	0.88	0.00	0.03	0.00
Pub FT	0.03	0.01	0.92	0.00	0.05
Pvt PT	0.17	0.11	0.00	0.71	0.00
Pub PT	0.10	0.01	0.11	0.01	0.76
Transition Class 2					
NE	0.83	0.11	0.01	0.04	0.01
Pvt FT	0.09	0.89	0.00	0.02	0.00
Pub FT	0.03	0.01	0.93	0.00	0.03
Pvt PT	0.19	0.13	0.00	0.67	0.00
Pub PT	0.12	0.01	0.13	0.01	0.72
Transition Class 3					
NE	0.82	0.10	0.02	0.05	0.01
Pvt FT	0.08	0.89	0.00	0.03	0.00
Pub FT	0.02	0.00	0.93	0.00	0.04
Pvt PT	0.17	0.11	0.00	0.71	0.01
Pub PT	0.09	0.01	0.13	0.01	0.76

Note: Entries show the probability of remaining or moving between states from period t to $t + 1$, where NE = non-employment, Pvt FT = private full-time, Pub FT = public full-time, Pvt PT = private part-time, and Pub PT = public part-time. Transition probabilities across the four latent transition classes ($km = 0$–3) show extremely similar magnitudes. The persistence in key employment states (0.81–0.83 for non-employment, 0.88–0.89 for private full-time, and 0.92–0.93 for public full-time) varies by less than one percentage point across classes. This similarity implies that the unobserved heterogeneity component (km) primarily reflects compositional differences in demographics and sectoral attachment rather than distinct behavioral adjustment regimes.

Table A1.7. Summary of Latent Classes

Latent Class	k^y	k^m	Class Size (%)	Female Share	Age Group	Educ. Level	Income Mobility	Labor Market Dynamics
1	1	1	3.9%	49.3%	51.9% over 45	Mixed	Upward (Q5: 35.9%)	High non-employment, mixed gender stability in Pvt FT, upward income mobility (Q5: 35.9%)
2	1	2	0.1%	50.9%	54.7% over 45	Low	Moderate (Q5: 20.4%)	High non-employment, women slower to transition to Pvt FT, moderate upward mobility (Q5: 20.4%)
3	1	3	2.0%	20.8%	53.1% over 45	Low	Moderate (Q5: 22.4%)	High non-employment, men stable in Pvt FT, moderate upward mobility (Q5: 22.4%)
4	1	4	2.0%	41.7%	Balanced	Mixed	Upward (Q5: 30.2%)	Balanced employment, more transitions to Pvt FT, upward mobility (Q5: 30.2%)
5	2	1	13.8%	71.1%	42.7% over 45	Low-Mid	Moderate (Q5: 23.9%)	Diverse transitions, more women shifting to Pvt PT, moderate income mobility (Q5: 23.9%)
6	2	2	1.8%	68.6%	Balanced	Low	Limited (Q5: 15.3%)	High non-employment, women remain in Pvt PT longer, limited upward mobility (Q5: 15.3%)
7	2	3	12.4%	36.7%	41.5% over 45	Low	Moderate (Q5: 18.2%)	Men stable in Pvt FT, limited transitions to Pvt PT, moderate income mobility (Q5: 18.2%)
8	2	4	4.1%	62.6%	36.2% over 45	Low	Limited (Q5: 19.4%)	High non-employment, women more likely to shift to Pvt PT, limited mobility (Q5: 19.4%)
9	3	1	24.3%	64.4%	77.2% under 45	Mixed	Upward (Q5: 25.5%)	Diverse transitions, women shift to/from public roles, upward income mobility (Q5: 25.5%)
10	3	2	2.5%	65.0%	80.2% under 45	Low-Mid	Limited (Q5: 13.1%)	High non-employment, younger workforce, women less upwardly mobile (Q5: 13.1%)
11	3	3	32.2%	32.4%	77.6% under 45	Low	Moderate (Q5: 14.4%)	Men stable in Pvt FT, fewer transitions to public roles, moderate mobility (Q5: 14.4%)
12	3	4	1.1%	55.9%	83.3% under 45	Mixed	Moderate (Q5: 20.1%)	Diverse transitions, balanced gender, younger workforce, moderate upward mobility (Q5: 20.1%)

Table A1.8. Predicted Transition Probabilities (Aggregate, Men, Women)

From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
Aggregate					
Non-employment (NE)	0.82	0.10	0.02	0.04	0.01
Private full-time (Pvt FT)	0.08	0.89	0.0025	0.03	0.0005
Public full-time (Pub FT)	0.02	0.01	0.93	0.0015	0.04
Private part-time (Pvt PT)	0.17	0.13	0.005	0.69	0.005
Public part-time (Pub PT)	0.09	0.01	0.14	0.01	0.74
Total	*0.31*	*0.41*	*0.14*	*0.10*	*0.04*

	Men					**Women**				
From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
NE	0.84	0.11	0.015	0.03	0.005	0.82	0.09	0.02	0.06	0.02
Pvt FT	0.07	0.91	0.002	**0.02**	0.0003	0.09	0.86	0.003	**0.05**	0.001
Pub FT	0.03	0.007	0.94	0.001	**0.02**	0.02	0.005	0.93	0.002	**0.05**
Pvt PT	0.24	**0.21**	0.006	**0.54**	0.003	0.15	**0.10**	0.005	**0.74**	0.01
Pub PT	0.17	0.02	0.20	0.01	**0.60**	0.08	0.01	0.13	0.01	**0.77**
Total	*0.32*	*0.52*	*0.11*	*0.04*	*0.01*	*0.31*	*0.31*	*0.17*	*0.15*	*0.06*

Max distance from observed: Aggregate=0.01 , Men=0.07 , Women=0.01

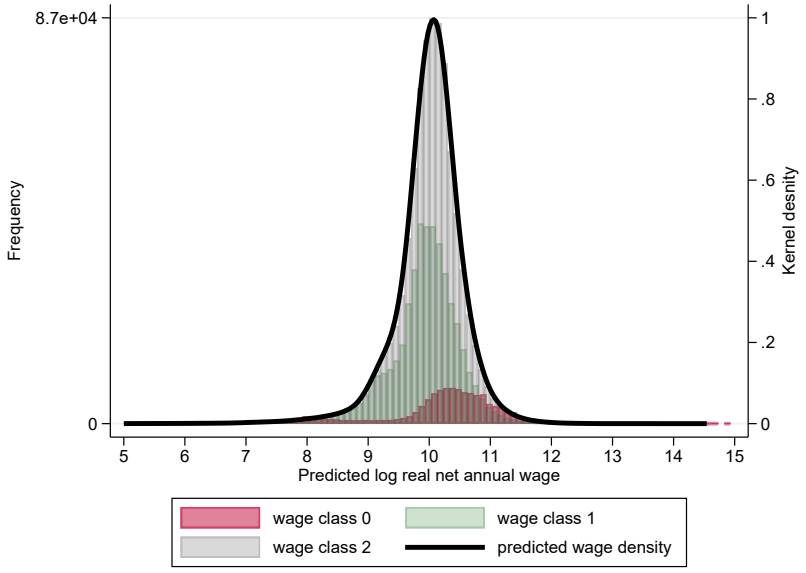


Figure A1.3. Predicted latent wage class distribution

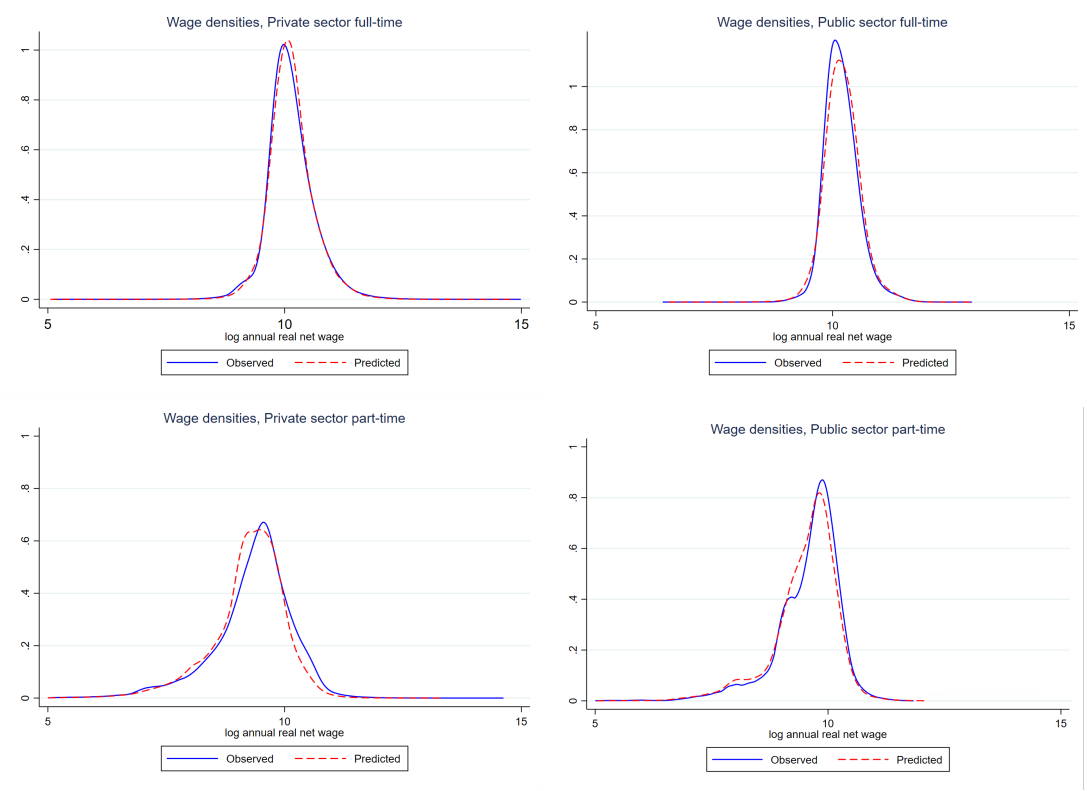


Figure A1.4. Wage Densities by employment state: Observed vs Predicted

A1.5.2 Prediction: Out-of-sample predictions of wages and transitions

Transition probabilities

Table A1.9. Transition Probabilities (Aggregate, Men, Women)

From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
			Aggregate		
Non-employment (NE)	0.83	0.10	0.02	0.04	0.01
Private full-time (Pvt FT)	0.08	0.89	0.0024	0.03	0.0005
Public full-time (Pub FT)	0.02	0.01	0.93	0.0014	0.04
Private part-time (Pvt PT)	0.17	0.13	0.01	0.69	0.01
Public part-time (Pub PT)	0.09	0.01	0.14	0.01	0.74
Total	*0.31*	*0.41*	*0.14*	*0.10*	*0.04*

	Men					**Women**				
From/To	NE	Pvt FT	Pub FT	Pvt PT	Pub PT	NE	Pvt FT	Pub FT	Pvt PT	Pub PT
NE	0.84	0.11	0.02	0.03	0.01	0.82	0.09	0.02	0.06	0.02
Pvt FT	0.07	0.91	0.0019	0.02	0.0002	0.09	0.86	0.0034	0.05	0.0009
Pub FT	0.03	0.01	0.94	0.0011	0.02	0.02	0.0045	0.93	0.0015	0.05
Pvt PT	0.24	0.22	0.01	0.54	0.0033	0.14	0.10	0.0049	0.74	0.01
Pub PT	0.17	0.02	0.20	0.01	0.59	0.08	0.01	0.13	0.01	0.77
Total	*0.32*	*0.52*	*0.11*	*0.04*	*0.01*	*0.31*	*0.31*	*0.17*	*0.15*	*0.06*

Max distance from observed: Aggregate=0.01 , Men=0.08 , Women=0.01

Wage prediction fit

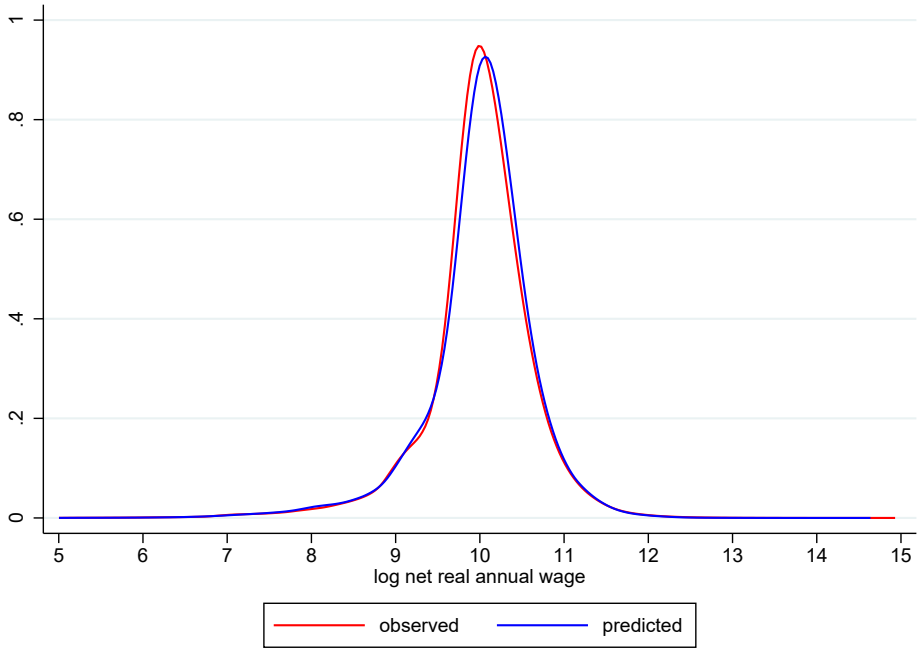


Figure A1.5. Observed and predicted wage densities

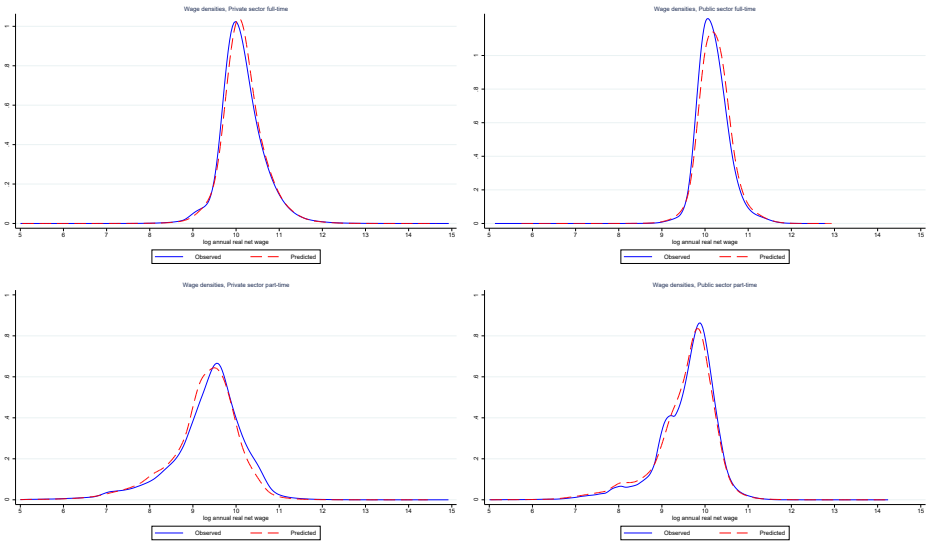


Figure A1.6. Observed and predicted wage densities by employment state



Figure A1.7. Observed and predicted wage densities and wage classes

A1.6 Estimated model parameters

Table A1.10. Parameters of unobserved mobility heterogeneity (mlogit models)

Mobility heterogeneity: $\kappa_i^m \mid z_i^f$			
$\kappa_i^m = 1$			
female	-0.233	first xp	-0.267
educ=1	0.198	constant	-1.361
educ=2	-20.756		
$\kappa_i^m = 2$			
female	-1.581	first xp	-0.263
educ=1	-0.626	constant	1.856
educ=2	-20.980		
$\kappa_i^m = 3$			
female	-0.458	first xp	0.060
educ=1	-0.758	constant	-1.307
educ=2	-0.679		

Table A1.11. Parameters of unobserved income heterogeneity (mlogit models)

Income heterogeneity: $\kappa_i^y \mid \kappa_i^m, z_i^f$			
$\kappa_i^y = 1$			
female	0.752	km=1	1.438
educ=1	-0.203	km=2	0.535
educ=2	-0.968	km=3	-0.628
first xp	-0.354	constant	1.835
$\kappa_i^y = 2$			
female	0.535	km=1	1.310
educ=1	-0.372	km=2	0.984
educ=2	-0.497	km=3	-2.588
first xp	-0.902	constant	3.312

Table A1.12. Parameters of state mobility (mlogit models)

Initial state selection: Pr{S_{i1}}							
$S_{i1} = 1$				$S_{i1} = 3$			
female	-0.609	k^m=1	-0.147	female	0.958	k^m=1	-0.203
med educ	0.653	k^m=2	0.108	med educ	0.170	k^m=2	-0.197
high educ	0.695	k^m=3	-0.917	high educ	0.068	k^m=3	0.596
first xp	0.600	constant	-0.281	first xp	0.480	constant	-1.912
$S_{i1} = 2$				$S_{i1} = 4$			
female	0.233	k^m=1	-0.059	female	1.298	k^m=1	0.004
med educ	1.175	k^m=2	-0.259	med educ	0.866	k^m=2	-0.381
high educ	1.635	k^m=3	-2.324	high educ	0.925	k^m=3	-0.440
first xp	0.958	constant	-2.676	first xp	0.608	constant	-3.715
State selection parameters: Pr{S_{it}} t>1							
$S_{it} = 1$				$S_{it} = 3$			
$S_{i,t-1}$=1	3.119	xp^2	-0.119	$S_{i,t-1}$=1	1.379	xp^2	-0.059
$S_{i,t-1}$=2	0.231	female	-0.264	$S_{i,t-1}$=2	-0.587	female	0.731
$S_{i,t-1}$=3	1.391	med educ	0.370	$S_{i,t-1}$=3	3.282	med educ	0.175
$S_{i,t-1}$=4	-0.171	high educ	0.439	$S_{i,t-1}$=4	0.368	high educ	0.134
$xp_{i,t-1}$	1.675	first xp	-1.545	$xp_{i,t-1}$	0.984	first xp	-0.819
$xp_{i,t-1} * S_{i,t-1}$=1	0.704	k^m=1	-0.110	$xp_{i,t-1} * S_{i,t-1}$=1	0.298	k^m=1	-0.110
$xp_{i,t-1} * S_{i,t-1}$=2	0.081	k^m=2	0.024	$xp_{i,t-1} * S_{i,t-1}$=2	0.353	k^m=2	-0.144
$xp_{i,t-1} * S_{i,t-1}$=3	0.235	k^m=3	-0.601	$xp_{i,t-1} * S_{i,t-1}$=3	0.485	k^m=3	0.183
$xp_{i,t-1} * S_{i,t-1}$=4	-0.242	constant	-2.065	$xp_{i,t-1} * S_{i,t-1}$=4	0.104	constant	-3.506
$S_{it} = 2$				$S_{it} = 4$			
$S_{i,t-1}$=1	-0.270	xp^2	-0.170	$S_{i,t-1}$=1	-1.568	xp^2	-0.124
$S_{i,t-1}$=2	5.307	female	0.261	$S_{i,t-1}$=2	3.686	female	0.947
$S_{i,t-1}$=3	0.237	med educ	0.494	$S_{i,t-1}$=3	0.363	med educ	0.462
$S_{i,t-1}$=4	2.930	high educ	0.769	$S_{i,t-1}$=4	4.669	high educ	0.447
$xp_{i,t-1}$	1.063	first xp	-0.906	$xp_{i,t-1}$	1.254	first xp	-0.936
$xp_{i,t-1} * S_{i,t-1}$=1	0.333	k^m=1	-0.023	$xp_{i,t-1} * S_{i,t-1}$=1	0.381	k^m=1	0.018
$xp_{i,t-1} * S_{i,t-1}$=2	1.229	k^m=2	-0.130	$xp_{i,t-1} * S_{i,t-1}$=2	0.514	k^m=2	-0.296
$xp_{i,t-1} * S_{i,t-1}$=3	-0.067	k^m=3	-1.058	$xp_{i,t-1} * S_{i,t-1}$=3	0.154	k^m=3	-0.125
$xp_{i,t-1} * S_{i,t-1}$=4	0.745	constant	-4.010	$xp_{i,t-1} * S_{i,t-1}$=4	0.782	constant	-5.101

Table A1.13. Parameters of cross-sectional income means and standard deviations

Income means: μ							
$S_{it} = 1$	0	xp	0.631	$S_{it} \times k^y$			
$S_{it} = 2$	-0.079	$xp * S_{it} = 2$	-0.055	**1 1**	-1.061	**3 1**	1.075
$S_{it} = 3$	-3.031	$xp * S_{it} = 3$	0.096	**1 2**	-0.636	**3 2**	1.841
$S_{it} = 4$	-2.683	$xp * S_{it} = 4$	-0.018	**2 1**	-0.416	**4 1**	1.828
xp^2	-0.080	female	-0.139	**2 2**	-0.711	**4 2**	1.281
med educ	0.283	high educ	0.524				
first xp	-0.164	constant	10.182				
Income standard deviation: σ							
$S_{it} = 2$	0.085	$xp^2 * S_{it} = 1$	0.076	$S_{it} \times k^y$			
$S_{it} = 3$	1.842	$xp^2 * S_{it} = 2$	0.187	**1 1**	-0.348	**3 1**	-0.582
$S_{it} = 4$	2.082	$xp^2 * S_{it} = 3$	0.122	**1 2**	-0.224	**3 2**	-0.632
xp	-0.590	$xp^2 * S_{it} = 4$	0.259	**2 1**	-0.432	**4 1**	-1.552
$xp * S_{it} = 2$	-0.538	first xp	0.232	**2 2**	-0.341	**4 2**	-0.686
$xp * S_{it} = 3$	-0.327	$k^m = 1$	-0.006				
$xp * S_{it} = 4$	-0.788	$k^m = 2$	-0.054				
$k^m = 3$	1.879	female	-0.019				
med educ	-0.021	high educ	0.178				
constant	-3.524						

Table A1.14. Parameters of income mobility

First-order income autocorrelation: τ_1							
$k^y = 1$	-0.139	$S_{i,t-1} = 1$	0.151	$k^y \times S_{i,t}$		$xp_{i,t-1} \times S_{i,t-1}$	
$k^y = 2$	0.008	$S_{i,t-1} = 2$	0.131	**1 2**	0.077	**1**	-0.019
$S_{i,t} = 2$	0.019	$S_{i,t-1} = 3$	0.149	**1 3**	0.126	**2**	-0.057
$S_{i,t} = 3$	-0.066	$S_{i,t-1} = 4$	0.096	**1 4**	-0.110	**3**	0.039
$S_{i,t} = 4$	0.204	$xp_{i,t-1}$	-0.610	**2 2**	-0.069	**4**	0.008
xp	0.717	$k^m = 1$	-0.033	**2 3**	0.068		
xp^2	-0.017	$k^m = 2$	-0.035	**2 4**	-0.186		
constant	0.310	$k^m = 3$	0.194				

A1.7 Results: Figures

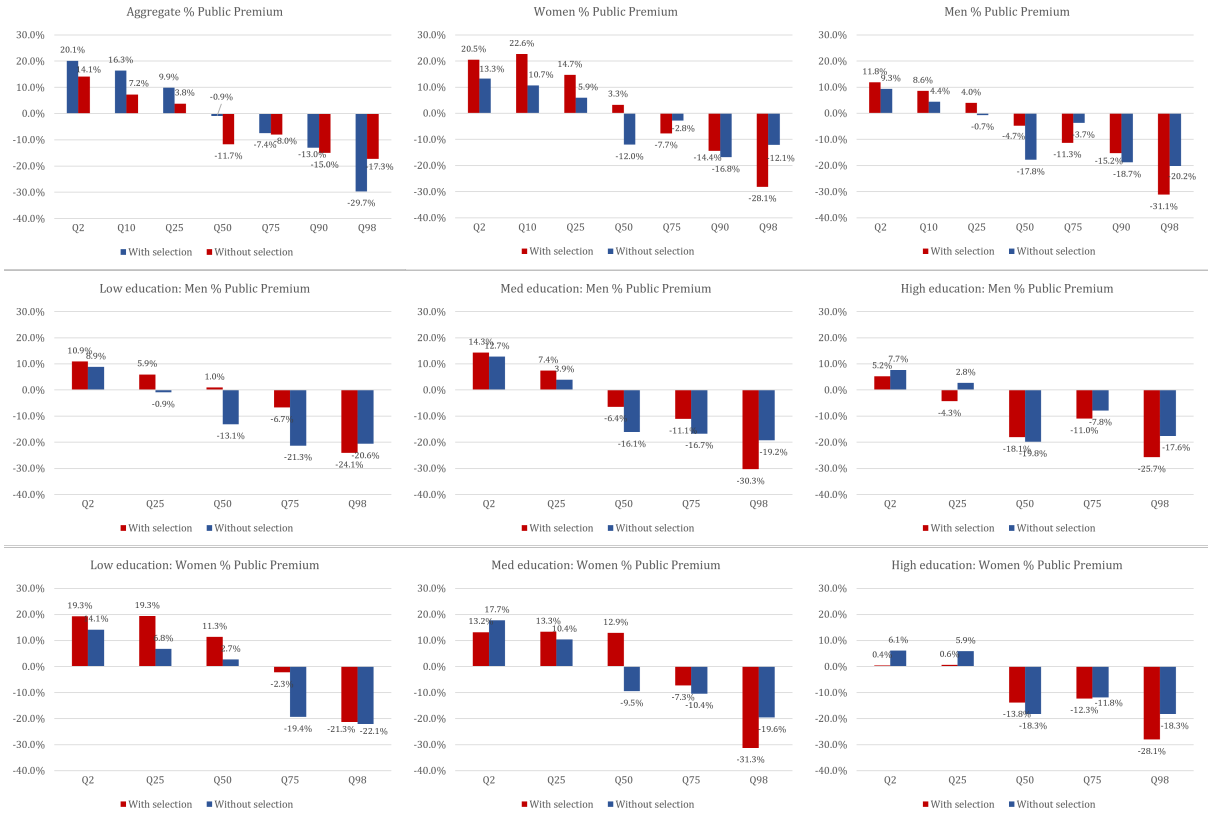


Figure A1.8. Public premium in lifetime earnings: "job for life" counterfactuals with and without selection

Bibliography

Stéphane Bonhomme and Jean-Marc Robin. Modeling individual earnings trajectories using copulas: France, 1990-2002. In Henning Bunzel, Bent J. Christensen, George R. Neumann, and Jean-Marc Robin, editors, *Structural Models of Wage and Employment Dynamics*, volume 275 of *Contributions to Economic Analysis*, pages 441–479. Elsevier, 2006.

Matt Dickson, Fabien Postel-Vinay, and Hélène Turon. The lifetime earnings premium in the public sector: The view from europe. *Labour Economics*, 29:179–190, 2014. doi: 10.1016/j.labeco.2014.07.003.

Fabien Postel-Vinay and Hélène Turon. The public pay gap in britain: Small differences that (don't?) matter. *Economic Journal*, 117(523):F422–F442, 2007. doi: 10.1111/j.1468-0297.2007.02067.x.

Appendix to Chapter 2

A2.1 Data

We use the DADS (Déclarations Annuelles de Données Sociales) Postes administrative dataset, which covers the universe of salaried workers in France for the years 2015–2019. Building the analysis sample from the raw DADS requires three stages: (i) constructing a consistent worker identifier across annual files and chaining workers over years, (ii) assembling yearly records into a single panel, and (iii) cleaning and collapsing the panel into worker-spells, and recoding occupations.

Stage 1: Cross-year worker chaining and occupation recoding. The DADS are released as a set of annual "yearfiles", each containing all employment records for a given calendar year. Crucially, the worker identifier (`IDENT_S`) is specific to each yearfile and cannot be used directly to track individuals across years. We therefore apply the pseudo-identifier procedure of Babet et al. (2023), which exploits the overlapping structure of the DADS: yearfile y reports not only spells in year t but also the previous-year spells ($t-1$) of the same workers. By matching the year-t records of yearfile $y-1$ to the year-($t-1$) records of yearfile y on a key composed of gender, employer (SIREN and NIC codes), gross wage, hours worked, spell start and end dates, age, and region of work, the procedure links workers across consecutive yearfiles. The matching imposes that the wage difference is minimized and the absolute age difference is at most two years; after 2014, hours worked are added as a third matching criterion. Only unique matches are retained; ambiguous cases (multiple candidates) are discarded. The procedure chains all pairwise links into a single time-invariant identifier (`IDENT_ALL`) that tracks each worker across the entire panel. A companion procedure computes firm- and establishment-level seniority (year of entry) from the chained records, and a separate correction script recovers reliable foreign-born and citizenship indicators for years where these variables are missing or erroneous.

Stage 2: Panel assembly. For the years 2015–2019, the raw DADS files are stored as regional SAS datasets. For each yearfile we extract all employment records with positive gross wages and positive hours, dropping records with missing birth year or a null employer identifier (SIREN = "000000000"). We keep the following variables: worker identifier, gender, birth year, employer SIREN and NIC, occupation (PCS code), gross wage, hours, spell start and end dates, spell duration in days, region and municipality of work, and the sampling weight. The regional extracts are appended into a single national file per year, merged with the pseudo-identifier crosswalk, and exported to Stata.

Stage 3: Spell construction and cleaning. Starting from the annual panel—where each observation is a worker × year × job—we construct spells through the following steps. First, we define a *job* z as

a unique (firm, occupation) pair. Second, we order jobs within each worker-year using continuity flags: a record is flagged as continuing from the previous year (respectively, into the next year) if the same worker held the same job z in the adjacent year. Jobs continuing from the prior year are placed first; jobs continuing into the next year are placed last; all others are in between. Third, for interior calendar years (excluding the worker's first and last observed year), any shortfall between 360 days and total observed spell duration within the year is imputed as an intervening non-employment spell. Fourth, consecutive observations sharing the same job z are collapsed into a single spell, summing their durations.

Occupations are directly observed in the data (PCS). It contains over 400 detailed 4-digit occupational codes, that are aggregated into 8 categories[5] tabulated below in table 2.1. Workers may change occupation both within and across firms.

Several filters are applied to ensure data quality. Workers whose calendar year contains three or more records yet totals fewer than 360 days—making the ordering of employment and non-employment ambiguous—are dropped. Workers exhibiting a "boomerang" pattern, where the same job appears in non-consecutive spells (indicative of simultaneous part-time jobs that cannot be sequenced), are removed. Workers with incomplete interior years within a continuous spell (duration < 360 days in a year that is neither the first nor the last year of the spell) are also excluded, as are workers whose first or last observed year contains two or more records summing to fewer than 360 days.

The resulting dataset is a panel of worker-spells, where each spell corresponds to a contiguous period during which a worker is employed at the same firm in the same occupation. A new spell begins whenever a worker changes employer, changes occupation, or transitions into or out of non-employment. Each spell records the worker's identity, firm, occupation, total duration in days, age at spell start, and gender. Non-employment spells are coded with a null firm and occupation identifier from the last employment spell. The final estimation sample comprises approximately 18 million workers, 1.6 million firms, and 27 million worker-spells.

A2.2 Methodology Details

A2.2.1 Dimensions and Parameter Counts

The model parameters comprise:

- $\mu_{x,y,k,fem}$: 624 cell-specific wage means
- $h_{x,y,k,fem}$: 624 cell-specific baseline hazards
- β_1, β_2: 2 common wage–age slopes
- γ_1, γ_2: 2 common hazard–age slopes
- σ: 1 common wage residual s.d.

[5]We thank Francis Kramarz for this classification of occupations.

Symbol	Value	Description
X	3	Latent worker types
Y	12	Latent firm types ($y = 0$: non-employment)
K	8	Occupations (PCS 2-digit)
Fem	2	Gender
Age	5	Age bins (quintiles)
Derived dimensions		
Base cells (x, y, k, fem)	624	$3 \times 13 \times 8 \times 2$
Extended cells (x, y, k, fem, age)	3,120	624×5

- $\pi(x \mid fem)$: $X \times Fem = 6$ type prevalences
- $P_1(\cdot)$: 3,120 initial match probabilities (multinomial over $13 \times 8 = 104$ states per extended cell)
- $P_M(\cdot)$: 3,120 destination multinomials ($Y + 2 = 14$ outcomes each)
- $P_{C\neg M}(\cdot)$: 3,120 stayer occupation multinomials (K outcomes each)
- $P_{CM}(\cdot)$: 3,120 mover occupation multinomials (K outcomes each)

A2.2.2 Full CEM Description

This appendix provides the closed-form M-step updates, exact numerical sub-routines, and concrete schedule parameters for the algorithm summarized in Section 2.2.

M-step parameter updates

The M-step updates eight parameter blocks sequentially. All multinomial estimates use Laplace smoothing with $\alpha = 1$ throughout.

M.1 — Wages. The age slopes β_1, β_2 are estimated by pooled within-cell demeaned OLS. For each base cell $c = (x, y, k, fem)$, the within-cell means $\bar{w}_c, \overline{age}_c, \overline{age^2}_c$ are computed over all employed observations in that cell. Deviations from these cell means are stacked and (β_1, β_2) is obtained by OLS of $w_{it} - \bar{w}_c$ on $(age_{it} - \overline{age}_c,\ age_{it}^2 - \overline{age^2}_c)$, equivalent to a within-cell fixed-effects regression pooled across all cells. Age-adjusted cell means are then set to

$$\hat{\mu}_c = \frac{1}{n_c} \sum_{i,t:\, c_{it}=c} \left(w_{it} - \hat{\beta}_1\, age_{it} - \hat{\beta}_2\, age_{it}^2\right),$$

where n_c is the spell count in cell c. Cells with no employed observations are imputed at the global (cross-cell) mean of age-adjusted wages. The common residual standard deviation is

$$\hat{\sigma} = \sqrt{\frac{1}{N_e} \sum_{i,t:\, y_{it}>0} \left(w_{it} - \hat{\mu}_{c_{it}} - \hat{\beta}_1\, age_{it} - \hat{\beta}_2\, age_{it}^2\right)^2},$$

where N_e is the total number of employed spells.

M.2 — Worker-type prior $\pi(x \mid fem)$. Counts $n_{x,fem}$ of workers of type x and gender *fem* are accumulated at each worker's first observed spell. Laplace-smoothed estimates are

$$\hat{\pi}(x \mid fem) = \frac{n_{x,fem} + \alpha}{n_{fem} + \alpha X},$$

where $n_{fem} = \sum_{x'} n_{x',fem}$.

M.3 — Initial matching $P_1(y, k \mid x, fem, age)$. Cell counts $n_{x,y,k,fem,a}$ are accumulated over all first spells and smoothed against the (x, fem, a) marginal:

$$\hat{P}_1(y, k \mid x, fem, a) = \frac{n_{x,y,k,fem,a} + \alpha}{n_{x,fem,a} + \alpha\,(Y{+}1)K},$$

where the denominator sums over all $(Y{+}1) \times K$ destination cells (including non-employment, $y = 0$).

M.4 — Firm sampling probability $q(j \mid y)$. For each employed type $y \in \{1, \ldots, Y\}$, the number of distinct firms currently assigned to type y is counted as N_y. The structural firm-draw probability is set to $q(j \mid y) = 1/N_y$. For non-employment, $q(0 \mid 0) = 1$.

M.5 — Separation hazard $h_{x,y,k,fem}$, γ_1, γ_2. Let $e_c = \sum_{n:\, c_n = c} \mathbf{1}[F_n = 1]$ denote the event count in cell c and define age-weighted exposure $S_c(\boldsymbol{\gamma}) = \sum_{n:\, c_n = c} D_n \exp(\gamma_1 \widetilde{age}_n + \gamma_2 \widetilde{age}_n^2)$, where $\widetilde{age} = age - \overline{age}$ is the demeaned age. The common age slopes (γ_1, γ_2) are updated by Newton–Raphson on the Poisson log-likelihood with the baseline hazards concentrated out. At each Newton step, the baseline is updated as $\hat{h}_c = (e_c + \frac{1}{2})/S_c(\boldsymbol{\gamma})$ (Jeffreys half-event smoothing for cells with zero events), and the score and observed information are

$$g_r = \sum_{F_n = 1} \widetilde{age}_n^r - \sum_n \mu_n \widetilde{age}_n^r, \qquad H_{rs} = -\sum_n \mu_n \widetilde{age}_n^r \widetilde{age}_n^s,$$

where $\mu_n = \hat{h}_{c_n} D_n \exp(\gamma_1 \widetilde{age}_n + \gamma_2 \widetilde{age}_n^2)$ is the fitted event count for observation n, and $r, s \in \{1, 2\}$. Newton steps are taken up to 20 iterations with convergence declared when $\max |g_r| < N \times 10^{-5}$. Updates are box-constrained to $|\gamma_1| \le 1$ and $|\gamma_2| \le 0.1$ to prevent explosive estimates. Cells with zero events are assigned the global hazard rate $\hat{h}_0 = \sum_c e_c / \sum_n D_n \exp(\hat{\gamma}_1 \widetilde{age}_n + \hat{\gamma}_2 \widetilde{age}_n^2)$.

M.6 — Destination-type distribution $P_M(y' \mid x, y, k, fem, age)$. For each spell ending in a separation ($F_{it} = 1$), the destination code y' takes values in $\{-1, 0, 1, \ldots, Y\}$: -1 encodes a within-firm spell (stayer, $M_{it} = 0$), 0 denotes non-employment, and $1, \ldots, Y$ denote the destination firm type. Counts $n_{x,y,k,fem,a,y'}$ are accumulated and the Laplace-smoothed multinomial over $N_{j2j} = Y + 2$ outcomes is:

$$\hat{P}_M(y' \mid x, y, k, fem, a) = \frac{n_{x,y,k,fem,a,y'} + \alpha}{n_{x,y,k,fem,a} + \alpha\, N_{j2j}}.$$

M.7 — Occupation transition for stayers $P_{C\neg M}(k' \mid x, y, k, fem, age)$. Accumulated from all spells where $F_{it} = 1$ and $M_{it} = 0$ (separation without employer change), conditioning on the current firm

type y:

$$\hat{P}_{C\neg M}(k' \mid x, y, k, fem, a) = \frac{n_{x,y,k,fem,a,k'} + \alpha}{n_{x,y,k,fem,a} + \alpha K}.$$

M.8 — Occupation transition for movers $P_{CM}(k' \mid x, y', k, fem, age)$. Accumulated from all spells where $F_{it} = 1$ and $M_{it} = 1$ (employer change), conditioning on the *destination* firm type y':

$$\hat{P}_{CM}(k' \mid x, y', k, fem, a) = \frac{n_{x,y',k,fem,a,k'} + \alpha}{n_{x,y',k,fem,a} + \alpha K}.$$

E-step: candidate likelihoods and sampling

Let

$$L_i(x') = \sum_{t=1}^{T_i} \ell_{it}(x', y_{j_{it}})$$

denote worker i's total log-likelihood under hypothetical type x'. The candidate log-likelihoods are computed by evaluating the likelihood with x_i replaced by $x' \in \{1, \ldots, X\}$ for all of i's spells, holding all other workers' types and the parameters fixed. In the deterministic phase ($\tau = 0$), $x_i \leftarrow \arg\max_{x'} L_i(x')$. In the simulated annealing phase ($\tau > 0$), the worker is sampled from

$$\Pr(x_i = x') = \frac{\exp\left[(L_i(x') - L_i^{\max})/\tau\right]}{\sum_{x''=1}^{X} \exp\left[(L_i(x'') - L_i^{\max})/\tau\right]},$$

where $L_i^{\max} = \max_{x'} L_i(x')$ is subtracted before exponentiation for numerical stability (the standard log-sum-exp trick that prevents overflow). Sampling is performed by inverse-CDF lookup against a uniform random draw. In the rare event of numerical underflow producing a degenerate probability vector, the worker is reassigned to a uniformly random type.

C-step: candidate likelihoods and sampling

The C-step parallels the E-step but operates on firms. For each firm j, the per-worker-spell average log-likelihood under type y' is

$$\bar{L}_j(y') = \frac{1}{n_j} \sum_{\{(i,t):\, j_{it} = j\}} \ell_{it}(x_i, y'),$$

where n_j is the number of worker-spells at firm j. The deterministic and softmax updates have the same form as the E-step, with $\bar{L}_j$ in place of L_i:

$$y_j \leftarrow \arg\max_{y' \in \{1,\ldots,Y\}} \bar{L}_j(y') \quad (\tau = 0), \qquad \Pr(y_j = y') \propto \exp\left[(\bar{L}_j(y') - \bar{L}_j^{\max})/\tau\right] \quad (\tau > 0),$$

with the same row-max subtraction for numerical stability. The economic motivation for the per-worker normalization $1/n_j$ is given in Section 2.2.

Iteration order

Each iteration executes M-step → log-likelihood evaluation → convergence check → E-step → C-step. The convergence check uses the log-likelihood evaluated immediately after the M-step (with the previous iteration's classifications). The E- and C-steps then update classifications based on the freshly updated parameters, and the next iteration's M-step uses these new classifications. Placing the convergence check before the classification updates ensures that convergence is declared on the basis of parameter stability rather than classification stability.

Schedule and convergence parameters

The geometric cooling schedule (Section 2.2) is implemented as

$$\tau_{\text{iter}} = \tau_{\text{start}} \cdot \exp\left[\frac{\log(\tau_{\text{end}}/\tau_{\text{start}}) \cdot (\text{iter} - 1)}{C - 1}\right],$$

with $\tau_{\text{start}} = 0.5$, $\tau_{\text{end}} = 0.01$, and $C = 200$ cooling iterations. Once τ falls below the threshold $\tau_{\text{argmax}} = 0.02$, the algorithm switches to deterministic arg max updates ($\tau = 0$) for the remaining iterations, up to a maximum of 400 iterations total. In practice the threshold is crossed around iteration 168.

The convergence criterion declares termination when the absolute change in log-likelihood between consecutive argmax-phase iterations satisfies $|\Delta \log L| < 100$ for 5 consecutive iterations. Given a total log-likelihood on the order of −450 million, this is equivalent to a per-spell change below $4 \cdot 10^{-6}$ nats, and corresponds to fewer than a handful of effective worker reassignments per iteration. The convergence check is only active during the deterministic phase, since log-likelihood naturally fluctuates during the stochastic annealing phase as workers and firms are reassigned probabilistically.

A2.3 Additional Results

A2.3.1 Detailed Decompositions by Subgroup

This appendix reports the full 4-factor decomposition for each age bin (Tables A2.15–A2.19), the pairwise sorting structure across all six factor pairs (Table A2.20), the NPV additive decomposition (Table A2.21), and a summary comparison across all decomposition methods (Table A2.22).

Table A2.15. 4-Factor Decomposition: Age Bin 1 (ages 20–27)

	Wages (% of subgroup)				NPV (% of subgroup)			
Factor	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}
Occ. k	4.8	4.9	4.7	13.7	32.1	36.3	34.7	44.8
Firm y	5.1	4.3	4.7	19.8	28.5	35.1	32.2	28.4
Worker x	1.2	3.9	2.5	12.6	17.9	21.3	20.6	13.8
Gender f	0.3	2.1	1.1	6.0	10.8	13.1	12.5	12.9
I_{j} / Btw		13.0%		52.0		100.0%		100.0
$\sum I_j$ / Wth		11.3%		48.0		89.3%		0.0
Synergy		+1.7 pp				+10.7 pp		

Notes. $I(j)$ = unconditional MI; $I(j|\cdot)$ = full conditional MI (entering last). ϕ^{MI} = MI Shapley; ϕ^{Var} = variance Shapley. $H(\tilde{w}) = 4.322$ bits, $\text{Var}(\tilde{w}) = 0.369$. $H(\text{NPV}) = 4.141$ bits, $\text{Var}(\text{NPV}) = 1.171$.

Table A2.16. 4-Factor Decomposition: Age Bin 2 (ages 28–34)

	Wages (% of subgroup)				NPV (% of subgroup)			
Factor	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}
Occ. k	8.8	6.7	8.0	26.2	40.2	34.7	38.6	47.5
Firm y	6.8	4.8	5.9	25.3	23.3	30.7	27.7	29.9
Worker x	0.5	2.1	1.6	6.7	19.2	19.1	20.8	10.9
Gender f	0.9	1.4	1.3	5.5	11.4	12.3	12.9	11.7
I_{j} / Btw		16.8%		63.8		100.0%		100.0
$\sum I_j$ / Wth		17.0%		36.2		94.0%		0.0
Synergy		-0.2 pp				+6.0 pp		

Notes. $I(j)$ = unconditional MI; $I(j|\cdot)$ = full conditional MI (entering last). ϕ^{MI} = MI Shapley; ϕ^{Var} = variance Shapley. $H(\tilde{w}) = 4.322$ bits, $\text{Var}(\tilde{w}) = 0.289$. $H(\text{NPV}) = 4.101$ bits, $\text{Var}(\text{NPV}) = 2.333$.

Table A2.17. 4-Factor Decomposition: Age Bin 3 (ages 35–41)

	Wages (% of subgroup)				NPV (% of subgroup)			
Factor	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}
Occ. k	10.2	6.7	9.1	28.9	37.6	32.4	38.0	44.3
Firm y	6.5	5.2	6.0	24.7	22.4	33.1	29.4	33.2
Worker x	0.4	2.6	2.2	7.3	15.4	18.1	20.3	10.6
Gender f	0.9	1.6	1.7	6.2	9.0	10.6	12.2	11.9
I_{j} / Btw		18.9%		67.1		100.0%		100.0
$\sum I_j$ / Wth		18.0%		32.9		84.3%		-0.0
Synergy		+0.9 pp				+15.7 pp		

Notes. $I(j)$ = unconditional MI; $I(j|\cdot)$ = full conditional MI (entering last). ϕ^{MI} = MI Shapley; ϕ^{Var} = variance Shapley. $H(\tilde{w}) = 4.322$ bits, $\text{Var}(\tilde{w}) = 0.340$. $H(\text{NPV}) = 4.030$ bits, $\text{Var}(\text{NPV}) = 2.628$.

Table A2.18. 4-Factor Decomposition: Age Bin 4 (ages 42–49)

	Wages (% of subgroup)				NPV (% of subgroup)			
Factor	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}
Occ. k	10.9	7.1	9.7	30.5	38.7	32.5	38.6	44.0
Firm y	6.2	5.0	5.7	23.7	23.7	33.5	29.6	35.3
Worker x	0.5	3.1	2.5	7.4	16.4	17.1	20.1	9.6
Gender f	1.0	1.9	2.0	6.8	8.5	9.7	11.6	11.1
I_{j} / Btw		19.8%		68.5		100.0%		100.0
$\sum I_j$ / Wth		18.6%		31.5		87.2%		-0.0
Synergy		+1.2 pp				+12.8 pp		

Notes. $I(j)$ = unconditional MI; $I(j|\cdot)$ = full conditional MI (entering last). ϕ^{MI} = MI Shapley; ϕ^{Var} = variance Shapley. $H(\tilde{w}) = 4.322$ bits, $\text{Var}(\tilde{w}) = 0.381$. $H(\text{NPV}) = 3.988$ bits, $\text{Var}(\text{NPV}) = 1.552$.

Table A2.19. 4-Factor Decomposition: Age Bin 5 (ages 50–60)

	Wages (% of subgroup)				NPV (% of subgroup)			
Factor	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}	$I(j)$	$I(j\|\cdot)$	ϕ^{MI}	ϕ^{Var}
Occ. k	10.5	6.9	9.3	29.6	39.9	34.7	39.9	44.0
Firm y	5.7	4.6	5.2	22.4	21.1	33.0	28.5	35.0
Worker x	1.0	3.7	3.0	8.5	16.9	16.9	20.0	9.9
Gender f	1.0	2.6	2.2	7.4	8.4	10.0	11.6	11.2
I_{j} / Btw		19.6%		67.9		100.0%		100.0
$\sum I_j$ / Wth		18.2%		32.1		86.3%		0.0
Synergy		+1.4 pp				+13.7 pp		

Notes. $I(j)$ = unconditional MI; $I(j|\cdot)$ = full conditional MI (entering last). ϕ^{MI} = MI Shapley; ϕ^{Var} = variance Shapley. $H(\tilde{w}) = 4.322$ bits, $\text{Var}(\tilde{w}) = 0.434$. $H(\text{NPV}) = 4.068$ bits, $\text{Var}(\text{NPV}) = 0.437$.

Table A2.20. Pairwise Sorting MI by Age Bin: All Six Factor Pairs (bits)

Age	$I(x;y)$	$I(x;k)$	$I(x;f)$	$I(y;k)$	$I(y;f)$	$I(k;f)$
1	0.093	0.221	0.101	0.473	0.082	0.137
2	0.051	0.439	0.077	0.392	0.094	0.173
3	0.030	0.381	0.059	0.364	0.088	0.162
4	0.033	0.411	0.071	0.344	0.096	0.175
5	0.036	0.413	0.092	0.320	0.095	0.177

Notes. Each cell reports $I(A;B)$ between two factors, measured in bits, computed over all employment spells in the age bin.

Table A2.21. Additive NPV Variance Decomposition

Component	Value	% of Var(NPV)	ρ
Pure heterogeneity			
Var($\hat{\gamma}_k$) — occupation	0.573	35.0%	
Var($\hat{\psi}_y$) — firm type	0.318	19.4%	
Var($\hat{\delta}_f$) — gender	0.091	5.6%	
Var($\hat{\alpha}_x$) — worker type	0.023	1.4%	
Pairwise sorting			
2Cov($\hat{\psi}$, $\hat{\gamma}$) — firm–occ.	0.154	9.4%	0.181
2Cov($\hat{\psi}$, $\hat{\delta}$) — firm–gender	0.083	5.1%	0.245
2Cov($\hat{\alpha}$, $\hat{\gamma}$) — worker–occ.	0.018	1.1%	0.080
2Cov($\hat{\alpha}$, $\hat{\psi}$) — worker–firm	−0.018	−1.1%	-0.103
2Cov($\hat{\gamma}$, $\hat{\delta}$) — occ.–gender	−0.016	−1.0%	-0.035
2Cov($\hat{\alpha}$, $\hat{\delta}$) — worker–gender	−0.027	−1.7%	-0.294
Non-additive components			
Var(interaction)	0.261	15.9%	
Var(within-cell residual)	0.176	10.7%	
Var(additive)	1.201	73.3%	
Total between-cell	1.462	89.3%	

Table A2.22. Wage vs. NPV: Summary Comparison Across Decomposition Methods

	Var. Shapley (%)		MI Shapley (%)		Additive own (%)	
Factor	Wage	NPV	Wage	NPV	Wage	NPV
Occupation k	26.6	40.8	8.2	27.7	17.8	35.0
Firm type y	23.0	29.5	5.2	20.5	17.4	19.4
Worker type x	7.8	8.4	2.1	14.2	1.4	1.4
Gender f	6.3	10.5	1.6	8.7	1.5	5.6
Between / I_j	63.7	89.3	17.2	71.0	46.7	73.3
Within-cell	36.3	10.7	—	—	—	—
Net synergy	—	—	+1.5	+20.2	—	—
Interaction	—	—	—	—	16.9	15.9
ρ firm–occ.	—	—	—	—	0.201	0.181
ρ firm–gender	—	—	—	—	0.221	0.245
ρ worker–firm	—	—	—	—	−0.065	−0.103
ρ worker–gender	—	—	—	—	−0.245	−0.294

Notes. Var. Shapley and MI Shapley: % of total variance and entropy respectively. Additive own = Var($\hat{\beta}_j$) as % of total variance. Between / I_j: total between-cell variance (Var. Shapley), joint MI (MI Shapley), or Var(additive) (Additive). Sorting ρ: correlation between demeaned main-effect components.

A2.3.2 Destination Choice Regressions

Table A2.23. Heterogeneity and Non-Employment

Variable	(t6a)	(t6b)	(t6c)	(t7)
$x = 1 \times \Delta\text{NPV}$	0.0514			
	(0.0602)			
$x = 2 \times \Delta\text{NPV}$	0.0244			
	(0.0222)			
$x = 3 \times \Delta\text{NPV}$	−0.0527			
	(0.0261)			
$\text{NPV}_{\text{origin},100k}$	−0.0136	−0.0135	−0.0131	
	(0.0072)	(0.0072)	(0.0071)	
Male $\times\Delta$NPV		0.0179		
		(0.0208)		
Female $\times\Delta$NPV		0.0037		
		(0.0324)		
Age bin 1 $\times\Delta$NPV			−0.0205	
			(0.0251)	
Age bin 2 $\times\Delta$NPV			−0.0039	
			(0.0359)	
Age bin 3 $\times\Delta$NPV			0.0245	
			(0.0447)	
Age bin 4 $\times\Delta$NPV			0.0480	
			(0.0564)	
Age bin 5 $\times\Delta$NPV			0.0449	
			(0.0552)	
NPV_{100k}				0.0158
				(0.0044)
Constant	0.2554	0.2555	0.2540	0.2076
	(0.0238)	(0.0239)	(0.0237)	(0.0183)
Observations	36,151	36,151	36,151	2,550
R^2	0.0113	0.0088	0.0101	0.0196

Table A2.24. Structural Log-Share (Conditional Logit)

Variable	(t8a)	(t8b)	(t8c)	(t8d)
$\text{NPV}_{\text{dest},100k}$	0.0862		−0.1735	
	(0.0284)		(0.0651)	
w_{dest}		0.0037	0.0080	
		(0.0008)	(0.0019)	
$\Delta V_{\text{wage},100k}$				0.1118
				(0.0313)
$\Delta V_{\text{opt},100k}$				0.0021
				(0.0337)
Constant	−2.4033	−2.4302	−2.2168	−2.1722
	(0.1012)	(0.0778)	(0.1069)	(0.0326)
Observations	63,436	63,436	63,436	63,436
R^2	0.2807	0.2823	0.2832	0.2843

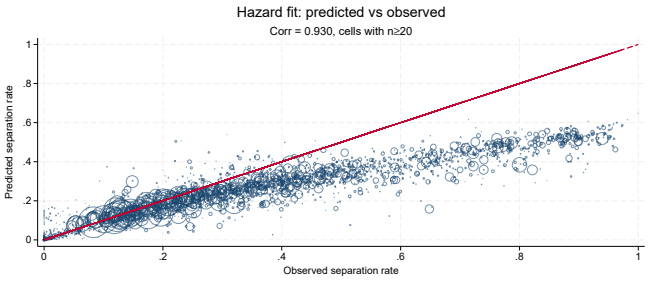


Figure A2.1. Predicted vs. observed separation probabilities across 2,731 cells defined by (k, female, age bin, x, y). Cells with fewer than 20 spells are excluded. The 45-degree line indicates perfect fit.

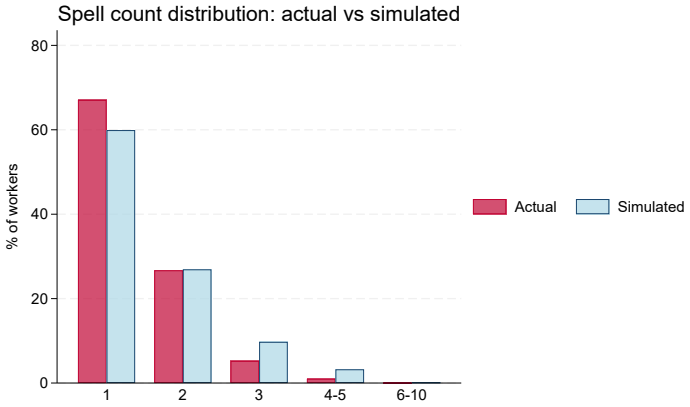


Figure A2.2. Distribution of workers by number of employment spells within the 5-year panel window: data (dark) vs. model simulation (light).

A2.3.3 Model Fit

A2.3.3.1 Separation Hazard Fit

The model parameterizes separation hazards as exponential rates $h(x, y, k, \text{age})$. Across 2,731 cells, the correlation between observed and predicted separation probabilities is $\rho = 0.93$ (Figure A2.1). The model captures the rank ordering of hazards across cells but systematically underpredicts observed separation rates (mean predicted 0.24 vs. observed 0.33). This gap reflects the exponential (memoryless) assumption: actual spell durations exhibit positive duration dependence, so the constant-rate model underpredicts the share of short spells that terminate quickly. Despite this limitation, the rank preservation across cells is what matters for the classification and sorting analysis.

A2.3.3.2 Career-Level Fit

Table A2.25 compares the distribution of workers by spell count between the data and the simulated panel. The model reproduces the dominance of single-spell careers and the steep decline in multi-spell frequencies (Figure A2.2). It overpredicts the share of workers with three or more spells (13.1% vs. 6.2%), a direct consequence of the exponential hazard: the memoryless property generates excess short spells. Within each spell-count bin, the model captures the qualitative patterns of wage growth (increasing

Table A2.25. Career-level comparison: data vs. simulation.

	Data			Model		
Spells	Share (%)	Mean Δw	Mean occ.	Share (%)	Mean Δw	Mean occ.
1	67.1	—	1.00	59.9	—	1.00
2	26.7	0.038	1.52	26.9	0.063	1.47
3	5.2	0.113	1.84	9.7	0.100	1.99
4–5	1.0	0.154	2.23	3.3	0.162	2.50

Notes: Mean Δw is mean log wage growth across spells. Mean occ. is the mean number of distinct occupations held. Data includes 18.5 million workers; simulation includes 50,000 workers.

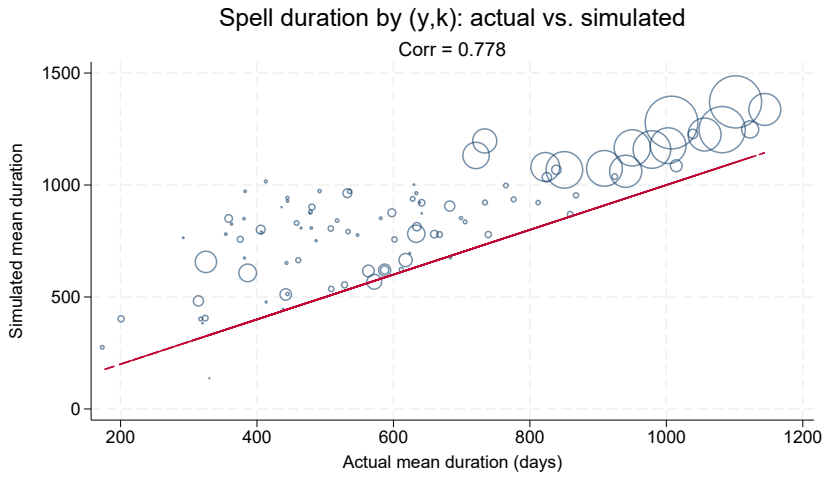


Figure A2.3. Mean spell duration by (y, k) cell: simulated vs. actual. Each point is one of 96 populated cells.

with mobility), occupation switching (more frequent for higher-spell workers), and non-employment incidence (rising with mobility).

A2.3.3.3 Spell Duration Fit

Spell durations by (y, k) cell exhibit a correlation of $\rho = 0.78$ between data and simulation, with the model overpredicting mean duration (839 vs. 610 days; Figure A2.3). This is the weakest of all validation metrics, reflecting the exponential assumption: the model matches cell-level hazard rates but not the shape of the duration distribution. A Weibull or piecewise-constant hazard would improve duration fit at the cost of substantially more parameters.

A2.3.3.4 Move-Type Composition

Table A2.26 decomposes year-to-year transitions into four categories defined by whether the worker changes firm type (y) and/or occupation (k). The model accurately captures the dominant role of within-firm-type occupation transitions (60% of all year-to-year transitions) and the relative rarity of simultaneous firm-type and occupation switches. The slight overprediction of joint (y, k) switches

Table A2.26. Move-type composition: data vs. simulation.

Transition type	Data (%)	Model (%)
Same y, same k	17.2	16.1
Same y, diff k	60.1	60.3
Diff y, same k	15.4	13.2
Diff y, diff k	7.3	10.4

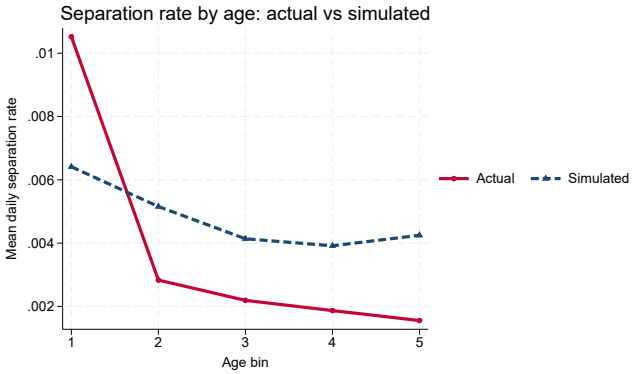


Figure A2.4. Mean daily separation rate by age bin: data vs. simulation.

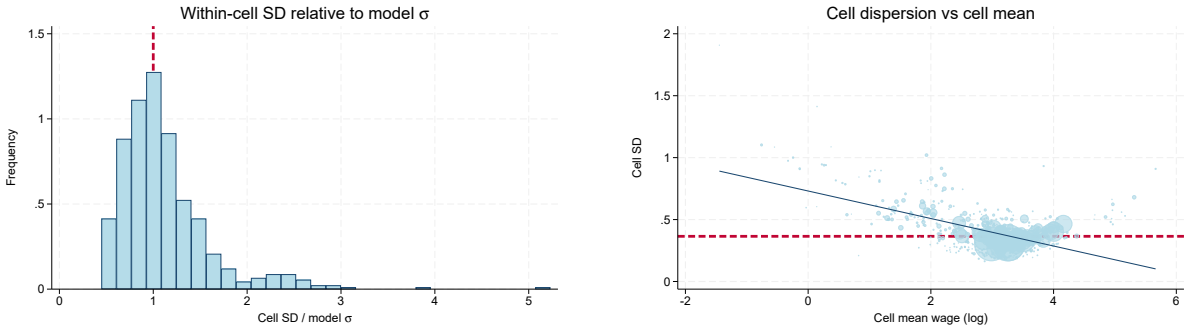
Table A2.27. Daily separation rates by age bin.

Age bin	Data (%)	Model (%)	N (data)
1 (youngest)	1.05	0.64	5,081,701
2	0.28	0.52	5,211,112
3	0.22	0.41	5,499,048
4	0.19	0.39	5,004,587
5 (oldest)	0.16	0.42	5,219,145

(10.4% vs. 7.3%) suggests mild misspecification in the independence structure between firm-type and occupation transitions, but the overall composition is well-matched.

A2.3.3.5 Age-Specific Separation Rates

The model incorporates age-dependent hazard scaling through parameters γ_1 and γ_2. Table A2.27 and Figure A2.4 compare the age profiles. Both data and model show declining separation rates with age, but the model underpredicts the youngest workers' separation rate while overpredicting rates for older workers, producing a flatter age profile than the data. This reflects unmodeled heterogeneity in the youngest cohort (probationary periods, short-term contracts) and the limitation of a two-parameter age adjustment.



(a) Distribution of SD ratios

(b) Cell SD vs. cell mean wage

Figure A2.5. Within-cell wage dispersion. Panel (a): distribution of the ratio of empirical cell standard deviation to the model's common σ. Panel (b): scatter of cell standard deviation against cell mean log wage, showing modest heteroscedasticity (lower-wage cells have higher dispersion).

Table A2.28. Within-cell standard deviation diagnostics.

Statistic	Value
Number of populated cells	576
Median ratio (SD_{cell} / σ)	1.01
Mean ratio	1.12
Share within [0.8, 1.2]	46.7%
Share within [0.5, 2.0]	93.4%

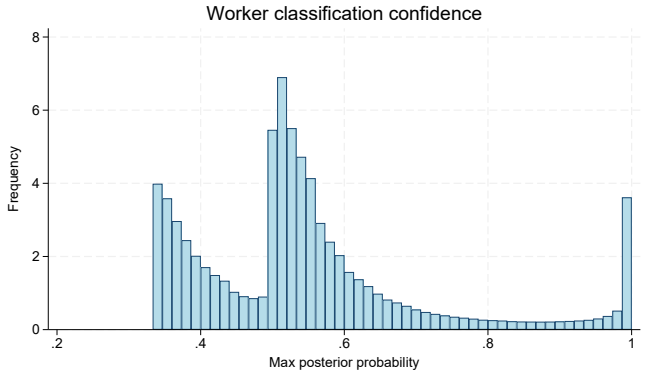


Figure A2.6. Distribution of maximum posterior probabilities across 18.5 million workers. With $X = 3$ latent types, the uniform prior assigns $p = 0.33$; values near 1.0 indicate near-certain classification.

A2.3.3.6 Within-Cell Wage Diagnostics

A2.3.3.7 Worker Classification

Table A2.29 reports the distribution of posterior entropy across workers. With $X = 3$ latent types, the maximum entropy is $\ln(3) \approx 1.10$, corresponding to a uniform posterior. We report approximate maximum posteriors computed under the assumption that the two minor-type posteriors are equal.

Table A2.29. Worker classification by posterior entropy.

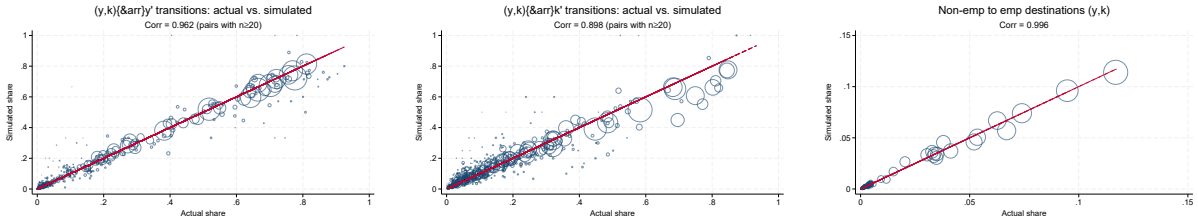
Entropy range	Workers	Share (%)	Approx. max post.	Mean spells
$H > 0.90$	6,138,099	33.1	~ 0.58	1.23
$H \in [0.55, 0.90]$	9,899,133	53.4	~ 0.79	1.41
$H \in [0.40, 0.55]$	1,108,855	6.0	~ 0.86	1.72
$H \in [0.20, 0.40]$	228,614	1.2	~ 0.94	1.67
$H \in [0.05, 0.20]$	323,301	1.7	~ 0.98	1.72
$H < 0.05$	845,460	4.6	~ 1.00	1.98

Notes: Entropy is Shannon entropy $H = -\sum_x p_i(x) \ln p_i(x)$, where $p_i(x)$ is the posterior probability that worker i is type x. Maximum entropy for $X = 3$ is $\ln(3) \approx 1.10$. Approximate maximum posteriors are computed assuming the two non-modal posteriors are equal.

Table A2.30. Firm classification accuracy by size (full likelihood).

Size bin	Firms	Accuracy (%)	Mean Δ log-L	Mean spells
1 (smallest)	942,011	45.1	−0.376	1.7
2	417,184	51.0	−0.177	6.0
3	209,696	71.4	−0.069	21.5
4	47,846	88.6	−0.024	91.3
5	12,115	90.6	−0.019	404.6
6 (largest)	2,558	92.0	−0.018	3,190.5

Notes: The full likelihood incorporates wages, separation hazards, and transition destinations. A wage-only classification achieves 25–63% accuracy across the same size bins.



(a) $(y, k) \to y'$ **(b)** $(y, k) \to k'$ **(c)** NE → E destinations

Figure A2.7. Additional simulated vs. actual transition shares: (a) firm-type destination conditional on origin (y, k); (b) occupation destination conditional on origin (y, k); (c) non-employment to employment destination distribution across (y, k) pairs.

A2.3.3.8 Firm Classification

Table A2.30 reports firm classification accuracy by firm size. Accuracy is defined as the share of firms whose assigned type under the full likelihood coincides with their best-fit type. The log-likelihood gap is the mean difference in log-likelihood between the best-fit and second-best type, normalized per spell.

Table A2.31. Simulation vs. data: transition correlations.

Transition level	Correlation	Cells
$y \to y'$	0.981	143
$k \to k'$	0.988	64
$(y, k) \to y'$	0.962	593
$(y, k) \to k'$	0.898	536
$p(k \mid y)$	0.997	96
NE $\to$ E destination	0.996	92
Duration by (y, k)	0.778	96

A2.3.3.9 Transition Correlations Summary

Table A2.31 summarizes the correlations between actual and simulated transition shares across all levels of aggregation. Figure A2.7 displays the conditional transition scatters and the non-employment exit destinations.

A2.4 Relation to the AKM Literature

The two-way fixed-effect model of Abowd et al. (1999)—hereafter AKM—has become the standard framework for decomposing wage variation in matched employer-employee data. AKM estimates

$$w_{it} = \alpha_i + \psi_{j(i,t)} + X_{it}\beta + \varepsilon_{it}, \tag{A2.4.1}$$

assigning one fixed effect per worker (α_i) and one per firm (ψ_j), using only wage levels and firm-to-firm transitions for identification. The CEM classification developed in this paper takes a different approach: it assigns workers and firms to a small number of latent types using wages, separation hazards, and occupation transitions jointly. In this section we estimate the AKM model on the same underlying data and examine what the comparison reveals about the objects each framework captures.

A2.4.1 Scope and Objects

The two frameworks differ in three fundamental respects. First, AKM identifies worker and firm effects from wages and employer changes alone. The CEM additionally exploits separation hazards $h(x, y, k,$ age), within-firm occupation transitions $P^{C\neg M}$, and cross-firm occupation transitions P^{CM}—information that is invisible to the AKM estimator. A worker who switches from occupation $k = 6$ to $k = 3$ at the same firm j is a stayer in AKM (neither α_i nor ψ_j changes) but an informative transition in the CEM. In the data, we observe 4.78 million occupation changes over the panel, of which 3.07 million (64%) occur within the same firm. These within-firm occupation switches—which identify the $P^{C\neg M}$ matrices and contribute to the separation of worker types from occupation types—are entirely absent from AKM's information set.

Second, AKM imposes additivity: the expected wage in match (i, j) is $\alpha_i + \psi_j$, ruling out complementarities between worker and firm characteristics. The CEM allows unrestricted cell means $\mu_{x,y,k}$ that can

capture arbitrary interactions between worker type, firm type, and occupation. Third, AKM estimates individual-level effects (17 million α_i's, 757,000 ψ_j's), while the CEM estimates group-level parameters (3 worker types × 12 firm types × 8 occupations = 288 wage cells). The AKM decomposition is therefore far more granular but structurally less informative: it can quantify how much of wage variation is "between workers" versus "between firms," but cannot distinguish occupation from other worker characteristics within the worker effect, nor allow the firm premium to vary by occupation. This is non-trivial since we observe 4.78 million changes in occupation over the 5-year panel, of which 3.07 million occur within the same firm.

The comparison is informative precisely because the two frameworks decompose the same wage variation through different lenses. If AKM's worker effects α_i are well-predicted by the CEM's latent types, this validates the structural classification. If they are not, the discrepancy reveals what each framework captures that the other cannot.

A2.4.2 Data Structure

A practical difference concerns the unit of observation. AKM requires one observation per worker-year, typically constructed by retaining only full-year spells—the standard practice in the French AKM literature (e.g., Card et al., 2013; Bonhomme et al., 2019). We depart from this convention because the CEM requires the within-year transition information that full-year restrictions discard. Our data tracks employment spells at daily frequency; a worker who transitions from firm A to firm B after 120 days in a given year contributes two spells to the CEM estimation, each with its own duration, wage, occupation, and destination.

To estimate AKM on the same population, we construct a worker-year panel by retaining the dominant employer per year (the firm at which the worker accumulated the most days). This preserves all workers in the sample—including those with multiple within-year spells—rather than dropping them. The resulting AKM sample contains 67 million worker-year observations in the largest connected set (92% of the full sample), covering 17 million workers and 757,000 firms.

This data construction differs from the standard approach in two ways that affect the AKM estimates. First, workers with within-year transitions are retained rather than excluded, increasing the representation of high-mobility workers. Second, the dominant-employer wage is the daily wage from the longer spell within the year, after removing within-year non-employment episodes. This retains more workers than the full-year restriction but assigns each worker-year a wage that reflects one employer rather than the full year of earnings. Both features tend to increase residual wage variance and may affect the covariance between worker and firm effects. We note these differences not as limitations but as consequences of maintaining comparability with the CEM's estimation sample: the same workers, the same firms, the same years.

Table A2.32. AKM variance decomposition.

	Spec 1 (age)	Spec 2 (+gender)	KSS (leave-out)	3-way ($+\kappa_k$)	Between CEM types
$\text{Var}(w)$	0.331	0.317	—	0.317	0.317
$\text{Var}(\alpha)/\text{Var}(w)$	84.4%	85.9%	74.2%	78.9%	26.9%
$\text{Var}(\psi)/\text{Var}(w)$	26.5%	27.7%	27.3%	27.3%	5.7%
$2\text{Cov}/\text{Var}(w)$	−23.6%	−26.7%	−26.8%	−28.2%	+4.1%
$\rho(\alpha, \psi)$	−0.249	−0.274	−0.298	−0.304	+0.166
$\text{Var}(\varepsilon)/\text{Var}(w)$	12.6%	13.2%	—	13.1%	—

Notes: Spec 1 residualizes on age and age^2. Spec 2 adds gender. KSS applies the Kline et al. (2020) leave-out correction to Spec 2 (heteroscedastic, $P_{ii} < 0.95$; 8.8% of observations excluded). The 3-way model adds occupation fixed effects κ_k. "Between CEM types" projects $\hat{\alpha}_i$ onto $E[\hat{\alpha} \mid x, k]$ and $\hat{\psi}_j$ onto $E[\hat{\psi} \mid y]$, where (x, k) are the CEM worker-type/occupation cells (24 groups) and y is the CEM firm type (12 groups). Connected set: 67,049,088 spells, 16,960,776 workers, 756,741 firms (92.1% of full sample).

A2.4.3 Variance Decomposition

We report results for two wage residualizations: Spec 1 removes age and age squared; Spec 2 additionally residualizes on gender. Table A2.32 reports the AKM variance decomposition across specifications. Several features are noteworthy.

Bias correction. The KSS leave-out correction (Kline et al., 2020) reduces $\text{Var}(\hat{\alpha})$ by 13.6% (from 85.9% to 74.2% of $\text{Var}(w)$) but barely affects $\text{Var}(\hat{\psi})$ (27.7% to 27.3%). This asymmetry is consistent with the leverage diagnostics: the mean worker leverage is $\bar{P}_{DD} = 0.445$, indicating that each observation is nearly half-determined by its own worker effect, while mean firm leverage is $\bar{P}_{FF} = 0.011$. Workers are estimated from approximately 4 observations on average; firms from approximately 90. The limited-mobility bias is therefore concentrated entirely in the worker dimension.

Negative sorting. The sorting correlation $\rho(\hat{\alpha}, \hat{\psi})$ is negative across all individual-level specifications, ranging from −0.25 (plug-in, age only) to −0.30 (three-way). This is more negative than typically reported in the French AKM literature, where recent studies using full-year spells and bias corrections find correlations near zero or mildly positive (Babet et al., 2023). As discussed above, the difference reflects our data construction: retaining within-year job changers via dominant-employer assignment, rather than restricting to full-year spells, increases the representation of high-mobility workers whose assigned wage reflects only their dominant employer within the year. This is a consequence of maintaining the same estimation sample as the CEM.

Three-way model. Adding occupation fixed effects κ_k to the AKM equation reduces $\text{Var}(\hat{\alpha})$ by 7.0% (the contamination bound), but the sorting correlation $\rho(\hat{\alpha}, \hat{\psi})$ becomes *more* negative ($-0.274 \rightarrow -0.304$). The cross-correlation $\rho(\hat{\alpha}, \hat{\kappa}) = +0.45$ reveals why: the occupation component of the worker effect is positively correlated with firm effects—workers in high-κ occupations tend to work at high-ψ firms—so extracting it removes a positive component from $\text{Cov}(\hat{\alpha}, \hat{\psi})$, leaving the residual

Table A2.33. Projections of AKM effects onto CEM types.

Dependent variable	Regressors	R^2	d.f.
Panel A: Worker effects $\hat{\alpha}_i$ (16,960,776 workers)			
	Worker type x	0.012	2
	Occupation k	0.267	7
	$x + k$ (additive)	0.276	9
	$x \times k$ (full interaction)	0.297	23
	Marginal $k \mid x$	0.264	
	Marginal $x \mid k$	0.009	
Panel B: Firm effects $\hat{\psi}_j$ (756,741 firms)			
	Firm type y	0.129	11
	Modal occupation k	0.013	7
	Occupation shares	0.021	7
	$y + k$ (additive)	0.131	18
	y + occupation shares	0.132	18
	$y \times k$ (full interaction)	0.138	~95
	Marginal $k \mid y$	0.002	
	Marginal occ shares $\mid y$	0.003	
	Marginal $y \mid k$	0.119	

Notes: Each row reports the R^2 from an OLS regression of the estimated AKM effect on the indicated regressors (with an intercept). Worker type $x \in \{1, 2, 3\}$ and firm type $y \in \{1, \ldots, 12\}$ are CEM latent types. Occupation $k \in \{1, \ldots, 8\}$ is the worker's (Panel A) or firm's (Panel B) modal PCS category. Occupation shares are the firm's workforce composition across PCS categories (continuous). "Marginal $a \mid b$" is $R^2(a + b) - R^2(b)$.

more negative. The occupation fixed effects themselves are modest ($\text{Var}(\hat{\kappa}) = 0.8\%$ of $\text{Var}(w)$; range $[-0.065, +0.110]$ in log points), but their role in mediating the worker-firm covariance is substantial.

Between-type sorting. The final column of Table A2.32 projects AKM effects onto CEM types: $E[\hat{\alpha} \mid x, k]$ (24 worker-type/occupation cells) and $E[\hat{\psi} \mid y]$ (12 firm types). At this level of aggregation, the sorting correlation reverses sign to $\rho = +0.17$. Between CEM types, sorting is positive—consistent with the mutual information analysis in Section **??**. The negative AKM sorting is entirely a within-type phenomenon. Within each (x, k) worker cell and each y firm type, individual $\hat{\alpha}_i$ and $\hat{\psi}_j$ co-vary negatively. This within-type covariation likely reflects compensating differentials—workers accepting lower ψ firms in exchange for non-wage amenities such as job stability or occupation match quality—as well as the mechanical consequences of the dominant-employer construction, which assigns high-mobility workers (who tend to have volatile $\hat{\alpha}_i$) to whichever firm they spent the most days at. Neither channel is captured by the CEM, which classifies workers and firms into types using the full mobility and wage structure (Section 2.2) and measures sorting at the type level (Section **??**).

A2.4.4 What AKM Effects Capture

To understand what the AKM effects represent in terms of the CEM's structural parameters, we project $\hat{\alpha}_i$ and $\hat{\psi}_j$ onto the CEM classification variables. Table A2.33 reports R-squared values from these regressions.

Worker effects are primarily occupation. Seven occupation dummies explain 26.7% of the variance of $\hat{\alpha}_i$ across 17 million workers, while the CEM worker type x (2 dummies) explains only 1.2%. The marginal contribution of x given k is 0.9%—worker type adds almost nothing once occupation is known. The full (x, k) interaction achieves $R^2 = 0.30$ with 23 degrees of freedom. To benchmark this, note that the CEM worker type x—which captures time-invariant worker characteristics through the latent classification—explains only $R^2 = 0.012$ of $\hat{\alpha}_i$ with 2 degrees of freedom. Seven occupation dummies alone explain 22 times more of the AKM worker effect than the CEM's own worker type. For further context, Card et al. (2013) report that adding industry and occupation controls to a human capital wage equation raises the R^2 from 0.35 to 0.50 on German data; our finding that occupation dummies explain 27% of the individual-level worker *fixed effects*—after netting out age—indicates that a substantial share of what AKM attributes to persistent individual heterogeneity is, in fact, occupation.

The remaining 70% of $\text{Var}(\hat{\alpha})$ is within-(x, k) variation that includes genuine within-occupation ability differences, but also match-specific wage premiums, hours heterogeneity, transitory shocks, and estimation noise. The KSS diagnostics indicate that the last of these is substantial: with mean worker leverage $\bar{P}_{DD} = 0.445$, each $\hat{\alpha}_i$ is estimated from very few effective observations, and the KSS leave-out correction reduces $\text{Var}(\hat{\alpha})$ from 0.273 to 0.236—a 13.6% reduction attributable to estimation variance (Table A2.32, Spec 2 vs. KSS column). Adjusting the denominator for this noise yields an implied R^2 on the signal component of approximately $0.30/(1 - 0.136) \approx 0.35$. Furthermore, the 3.07 million within-firm occupation changes—which the CEM models explicitly through $P^{C\neg M}$ and which shift wages from one $\mu_{x,y,k}$ cell to another—are stayer observations in AKM that contribute to ε_{it} or distort $\hat{\alpha}_i$. The "unexplained" within-(x, k) variance is therefore not exclusively worker ability that the CEM fails to capture; it is a mixture of sample construction (for comparability) and components that AKM structurally cannot separate and the CEM models through distinct channels.

Firm effects are genuinely firm-specific. The pattern for $\hat{\psi}_j$ is the mirror image. Twelve CEM firm-type dummies explain 12.9% of firm effect variance, and this is almost entirely orthogonal to occupation: adding the firm's modal occupation (+0.2%) or its full workforce occupation composition (+0.3%) contributes negligibly beyond firm type. The marginal R^2 of y given k is 11.9%, confirming that CEM firm types capture genuine firm-level wage variation that is unrelated to the occupational composition of the workforce. This connects to the Shapley decomposition in Section **??**, where firm type contributes to wage variance primarily through direct effects on $\mu_{x,y,k}$ rather than through its correlation with occupation.

A2.4.5 Non-Additivity and the Event Study

The AKM model imposes that the expected wage in match (i, j) is $\alpha_i + \psi_j$—additively separable. The CEM allows arbitrary cell means $\mu_{x,y,k}$. We test additivity directly through two exercises.

Table A2.34. Mover event study: mean wage change by origin and destination $\hat{\psi}$-quartile.

	Destination quartile			
Origin quartile	Q1	Q2	Q3	Q4
Q1	+0.020	+0.234	+0.437	+0.714
Q2	−0.246	+0.008	+0.063	+0.265
Q3	−0.473	−0.086	−0.003	+0.121
Q4	−0.811	−0.281	−0.145	−0.002

Notes: Each cell reports the mean change in log daily wages for workers moving between firms in the indicated $\hat{\psi}$-quartiles. Quartiles are defined over the full distribution of $\hat{\psi}_j$ weighted by employment. Total movers: 3,680,891.

Cell means versus additive prediction. For each (x, y, k) cell, we compute the mean AKM prediction $E[\hat{\alpha} \mid x, y, k] + E[\hat{\psi} \mid x, y, k]$ and compare it to the actual cell mean $E[w \mid x, y, k]$. The difference is the mean AKM residual within the cell, $E[\hat{\varepsilon} \mid x, y, k]$, which under correct specification should be zero. Across 288 populated (x, y, k) cells, the correlation is 0.998, but the residuals have a standard deviation of 0.040 log points and range from −0.46 to +0.06. This confirms that while AKM's additive structure approximates the cell means well on average, specific (x, y, k) combinations deviate meaningfully—particularly cells where occupation interacts with firm type to depress wages below the additive benchmark.

In terms of explanatory power, AKM's 17.7 million individual parameters achieve $R^2 = 0.87$, while the CEM's 36 (x, y) cells achieve $R^2 = 0.22$ and its 288 (x, y, k) cells achieve $R^2 = 0.44$. The gap of 0.43 reflects the granularity difference: AKM fits individual-level wage variation that the CEM's group structure does not target. The addition of occupation to the CEM cells doubles the R^2 (from 0.22 to 0.44), consistent with occupation being a primary axis of wage variation that the AKM worker effect absorbs but the CEM explicitly separates.

Mover event study. Following Card et al. (2013), we classify movers by the $\hat{\psi}$-quartile of their origin and destination firms and compute the mean wage change Δw for each origin-destination pair (Table A2.34). Under AKM's additive structure, the wage gain from moving Q1→Q4 should equal (in absolute value) the wage loss from moving Q4→Q1. We find a systematic asymmetry: downward movers lose more than upward movers gain. The sum $\Delta w^{\text{up}} + \Delta w^{\text{down}}$ is −0.097 for Q1↔Q4, −0.036 for Q1↔Q3, −0.016 for Q2↔Q4, and −0.024 for Q2↔Q3. The asymmetry is consistent in sign across all pairs, indicating mild but systematic non-additivity.

Furthermore, the wage gain from Q1→Q4 moves varies across occupations, ranging from +0.63 (occupation $k = 3$) to +0.80 (occupation $k = 8$), with a standard deviation of 0.060 across the 8 occupation categories. This occupation-dependence of the firm premium is precisely the interaction that the CEM's $\mu_{x,y,k}$ parameterization captures and AKM's additive $\alpha_i + \psi_j$ structure cannot accommodate.

A2.4.6 Discussion

The AKM decomposition and the CEM classification offer complementary perspectives on wage inequality. AKM provides a comprehensive accounting at the individual level—every worker and every firm receives an estimated effect—but cannot distinguish occupation from other worker characteristics within the worker effect, nor allow the firm premium to vary by occupation. The CEM operates at a coarser resolution but exploits a richer information set: separation hazards, occupation transitions (including 3.07 million within-firm occupation changes), and non-employment dynamics that are invisible to the AKM estimator.

The projections reveal that these frameworks decompose the same wage variation into different objects. What AKM calls worker heterogeneity is predominantly occupation ($R^2 = 0.27$ from 7 dummies); what it calls firm heterogeneity genuinely captures firm-level variation ($R^2 = 0.13$ from 12 type dummies, with occupation adding nothing). Between CEM type, sorting is positive ($\rho = +0.17$), consistent with the mutual information analysis. The negative AKM sorting reflects within-type individual variation—including estimation noise, match-specific effects, and the consequences of the dominant-employer data construction—that the CEM's structural approach does not target and cannot use.

These findings do not invalidate the AKM framework; rather, they clarify what its decomposition measures in the presence of occupational heterogeneity. The CEM's three-dimensional classification (x, y, k) separates latent worker heterogeneity from occupation and allows firm effects to vary by occupation—distinctions that the AKM framework, by construction, cannot make.

A2.5 Validation using CEM on Simulated Data

A2.5.1 Simulation Design

We assess the stability of the CEM's variance decomposition through a Monte Carlo exercise. Starting from the estimated parameters on the French DADS data ($X = 3$ worker types, $Y = 12$ firm types, $K = 8$ occupations, 5 age bins, 2 genders), we simulate 200,000 career trajectories and re-estimate the CEM on the simulated data. The comparison between the input and recovered decompositions tests whether the Shapley variance shares and sorting correlations are structural features of the model or artifacts of a particular classification.

Career simulation. Each simulated worker is initialized by drawing a worker type x and gender jointly from the estimated marginal $\pi(x, f)$, an age bin uniformly from $\{1, \ldots, 5\}$ with starting age equal to the bin mean, and an initial firm type y and occupation k from the estimated initial match distribution $P_1(y, k \mid x, f,$ age bin$)$. The worker then evolves through the estimated mobility structure: each employment spell has a duration drawn from an exponential distribution with rate $h(x, y, k, f) \cdot \exp(\gamma_1(A - \bar{A}) + \gamma_2(A - \bar{A})^2)$, where A is the worker's current age. Upon separation, a destination code

is drawn from the transition matrix $PM(x, y, k, f,$ age bin): the worker may experience a within-firm occupation change (drawn from $P^{C\neg M}$), exit to non-employment, or move to a new firm type y' with a new occupation drawn from P^{CM}. Wages are drawn from $w = \mu_{x,y,k,f} + \beta_1 A + \beta_2 A^2 + \sigma\varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 1)$. The simulation uses the same 5-year panel structure as the data (1,800 days).

Firm identities. The CEM classifies individual firms, so the simulated data requires firm identities, not just firm types. We generate two versions of the simulated panel, differing only in how firm identities are assigned within each type:

- **Uniform:** Each firm type y has N_y firms, scaled proportionally from the actual data to match the simulated sample size (total $\approx$ 18,000 firms). Workers are assigned to firms uniformly at random within their type. This produces approximately equal-sized firms within each type.
- **Zipf:** The same number of firms per type, but workers are assigned with probability proportional to $1/j$ (Zipf's law with $\alpha = 1.0$), producing a realistic heavy-tailed firm size distribution. The resulting distribution has median firm size 2, $p_{99} = 96$, and maximum 13,001 spells—close to the actual data.

Career trajectories are identical across the two versions; only the mapping from firm types to firm identities differs. This isolates the effect of firm size heterogeneity on the CEM's classification performance.

Re-estimation. We run the CEM algorithm on each simulated dataset with the same settings as the main estimation ($X = 3$, $Y = 12$, $K = 8$, 5 age bins, 300 iterations, annealing from $T = 0.5$ to $T = 0.01$). The re-estimated parameters are saved and compared to the original inputs.

A2.5.2 Scalar recovery.

The global parameters—estimated from the full sample without depending on cell-level classification—are recovered precisely in both simulations (Table A2.35).

Table A2.35. Scalar parameter recovery.

Parameter	Original	Zipf	Uniform
σ (wage SD)	0.365	0.375	0.378
β_1 (age effect)	0.468	0.466	0.480
β_2 (age^2 effect)	−0.046	−0.045	−0.047
γ_1 (hazard age)	−0.181	−0.205	−0.184
γ_2 (hazard age^2)	0.032	0.039	0.024

A2.5.3 Variance Decomposition

Table A2.36 compares the four-factor additive wage variance decomposition between the original CASD estimation and the two simulated re-estimations. The decomposition partitions $\mathrm{Var}(\tilde{w})$ into pure heterogeneity terms $\mathrm{Var}(\hat{\alpha}_x)$, $\mathrm{Var}(\hat{\psi}_y)$, $\mathrm{Var}(\hat{\gamma}_k)$, $\mathrm{Var}(\hat{\delta}_f)$; pairwise sorting covariances; a non-additive interaction; and a within-cell residual.

Table A2.36. Monte Carlo validation: four-factor additive variance decomposition.

	% of Var($\tilde{w}$)			Sorting correlation ρ		
Component	Original	Uniform	Zipf	Original	Uniform	Zipf
Pure heterogeneity						
Var($\hat{\gamma}_k$) — occupation	17.8	16.4	17.5			
Var($\hat{\psi}_y$) — firm type	17.4	17.7	18.5			
Var($\hat{\delta}_f$) — gender	1.5	1.1	1.2			
Var($\hat{\alpha}_x$) — worker type	1.4	1.8	0.0			
Pairwise sorting						
2Cov($\hat{\psi}$, $\hat{\gamma}$) — firm–occ.	7.1	7.4	6.9	+0.201	+0.216	+0.191
2Cov($\hat{\psi}$, $\hat{\delta}$) — firm–gender	2.3	1.8	1.9	+0.221	+0.205	+0.202
2Cov($\hat{\alpha}$, $\hat{\gamma}$) — worker–occ.	0.6	0.9	−0.0	+0.056	+0.078	−0.027
2Cov($\hat{\alpha}$, $\hat{\delta}$) — worker–gender	−0.7	−0.3	−0.2	−0.245	−0.103	−0.435
2Cov($\hat{\alpha}$, $\hat{\psi}$) — worker–firm	−0.7	−0.1	−0.1	−0.065	−0.008	−0.080
2Cov($\hat{\gamma}$, $\hat{\delta}$) — occ.–gender	−0.0	0.2	−0.1	−0.002	+0.018	−0.012
Non-additive components						
Var(interaction)	16.9	16.7	18.5			
Var(within-cell residual)	36.3	36.4	36.0			
Var(additive)	46.7	46.8	45.5			
Total between-cell	63.7	63.6	64.0			

Notes: "Original" is the CEM estimation on 27 million spells from the DADS (2015–2019). "Uniform" and "Zipf" are re-estimations on 200,000 simulated workers (≈315,000 employed spells) generated from the original parameters, differing only in within-type firm size distribution. The decomposition partitions Var($\tilde{w}$) into additive effects, pairwise covariances, a non-additive interaction (cell-specific premia beyond the additive structure), and a within-cell residual. Sorting correlations are computed from the estimated fixed effects.

The dominant components of the variance decomposition are recovered precisely. Firm-type heterogeneity (17.4% original vs. 17.7% uniform, 18.5% Zipf), occupation heterogeneity (17.8% vs. 16.4%, 17.5%), and firm–occupation sorting ($\rho = 0.201$ vs. 0.216, 0.191) are all within 1.5 percentage points and 0.02 correlation units of the original. The within-cell residual (36.3% vs. 36.4%, 36.0%) and the non-additive interaction (16.9% vs. 16.7%, 18.5%) are similarly stable. The total between-cell variance share is virtually identical across all three estimates (63.7%, 63.6%, 64.0%). These are the paper's central findings, and they are robust to re-estimation on independent data at one-ninetieth of the original sample size.

Worker-type heterogeneity is the least stable dimension. Var($\hat{\alpha}_x$) contributes 1.4% in the original, 1.8% in the uniform simulation, and effectively zero in the Zipf simulation. The worker-related sorting correlations are correspondingly noisy: $\rho(\hat{\alpha}, \hat{\delta})$ ranges from −0.10 to −0.44 across the three estimates. This instability reflects the difficulty of separating $X = 3$ latent worker types that contribute only 1.4% of wage variance—the types are primarily distinguished through mobility and occupation patterns rather than wage levels, making them sensitive to sample size and firm classification quality. Crucially, this instability does not affect the structural conclusions: even if the worker-type classification shifts, the dominant variance components (firm, occupation, their interaction) are preserved.

Firm size heterogeneity affects worker-type recovery but not the aggregate decomposition. The uniform simulation, where all firms within a type have equal size, outperforms the Zipf simulation on worker-type recovery (1.8% vs. 0.0%) and worker-related correlations. This is expected: with Zipf-distributed sizes, many firms have only 1–2 spells, making the C-step classification noisy and propagating uncertainty into the worker-type posteriors. The firm-type and occupation components, which are identified from larger firms with many observations, are unaffected. For the main CASD estimation with 27 million spells, data density per cell (≈47,000 spells per cell on average) is far above the threshold where these finite-sample effects arise.

A2.5.4 Mutual Information Decomposition

Table A2.37 compares the four-factor Shapley mutual information decomposition. The total MI, $I(x, y, k, f; w)$, measures the share of wage entropy explained by the cell classification. The Shapley values partition the total MI into the marginal contributions of each factor, accounting for all interaction orders.

Table A2.37. Monte Carlo validation: Shapley mutual information decomposition.

	% of $H(\tilde{w})$			Error (pp)	
Component	Original	Uniform	Zipf	Uniform	Zipf
Shapley values					
Occupation (k)	8.24	6.60	6.63	−1.6	−1.6
Firm type (y)	5.23	4.95	5.02	−0.3	−0.2
Worker type (x)	2.14	2.35	2.02	+0.2	−0.1
Gender (f)	1.61	1.30	1.74	−0.3	+0.1
Unconditional MI					
$I(k; w)$	8.77	6.87	6.88	−1.9	−1.9
$I(y; w)$	5.79	5.31	5.39	−0.5	−0.4
$I(x; w)$	0.48	0.92	0.35	+0.4	−0.1
$I(f; w)$	0.71	0.59	0.59	−0.1	−0.1
Conditional MI					
$I(k; w \mid x, y, f)$	7.20	6.31	6.50	−0.9	−0.7
$I(y; w \mid x, k, f)$	4.66	4.86	4.98	+0.2	+0.3
$I(x; w \mid y, k, f)$	3.32	3.77	3.87	+0.5	+0.6
$I(f; w \mid x, y, k)$	2.29	2.11	3.11	−0.2	+0.8
Total $I(x, y, k, f; w)$	17.22	15.19	15.42	−2.0	−1.8
Net synergy	1.47	1.50	2.21	+0.0	+0.7

Notes: All values are expressed as percentages of the target entropy $H(\tilde{w})$. The Shapley value for factor j is $\phi_j = \frac{1}{4!} \sum_\pi [I(S_j^\pi \cup \{j\}; w) - I(S_j^\pi; w)]$, where the sum is over all orderings of the four factors. The total MI equals the sum of Shapley values. The systematic ∼2pp underestimation of total MI relative to the original reflects the smaller sample (200,000 vs. 18 million workers): fewer observations per cell reduce the measured information content.

The Shapley ranking is preserved identically across all three estimates:

$$\phi_k \text{ (occupation)} > \phi_y \text{ (firm type)} > \phi_x \text{ (worker type)} > \phi_f \text{ (gender)}.$$

The firm-type and worker-type Shapley values are recovered within 0.3 percentage points. The occupation Shapley is systematically lower in both simulations (6.6% vs. 8.2%), tracking the ~2pp reduction in total MI. When expressed as ratios relative to the occupation Shapley—which nets out the level effect—the relative contributions are stable: the firm-to-occupation ratio is $\phi_y/\phi_k = 0.63$ in the original versus 0.75–0.76 in the simulations.

The conditional MI $I(y; w \mid x, k, f) = 4.7$–$5.0\%$ confirms that firm types carry substantial wage information even after conditioning on worker type, occupation, and gender—a result that is stable across all three estimates. The total MI is systematically lower by ~2 percentage points in both simulations, reflecting the smaller sample rather than a structural difference: fewer observations per cell mechanically reduce the measured information content. Taken together with the variance decomposition, the MI results confirm that the CEM's structural findings—the dominant role of occupation and firm type, the firm–occupation sorting channel, and the relatively minor contribution of latent worker types—are robust features of the estimated model.

A2.5.5 AKM decomposition on simulated vs original data

Table A2.38 reports the AKM variance decomposition — worker fixed effects $\hat{\alpha}_i$, firm fixed effects $\hat{\psi}_j$, and their covariance — estimated on the original DADS data and on each simulated panel, for both the standard plug-in estimator and the KSS leave-one-out bias correction. All figures are expressed as percentages of $\text{Var}(w)$.

Table A2.38. AKM variance decomposition: original vs simulated data.

	Plug-in			KSS-corrected		
Component	Original	Uniform	Zipf	Original	Uniform	Zipf
$\text{Var}(\hat{\alpha})$ — worker	85.9%	96.1%	83.8%	74.2%	90.2%	80.2%
$\text{Var}(\hat{\psi})$ — firm	27.7%	46.6%	32.8%	27.3%	45.2%	32.2%
$2\text{Cov}(\hat{\alpha}, \hat{\psi})$	−26.7%	−48.9%	−23.4%	−26.8%	−49.6%	−23.7%
$\rho(\hat{\alpha}, \hat{\psi})$	−0.274	−0.365	−0.223	−0.298	−0.389	−0.233
$\text{Var}(\hat{\varepsilon})$	13.2%	6.1%	6.8%	13.2%	6.1%	6.8%

Notes: Specification 2, controlling for age polynomials. The original estimation covers 16.96 million workers and 757K firms (connected component). Each simulated panel contains 200,000 workers; "Uniform" has 17,654 connected firms (98.1%); "Zipf" has 12,407 (76.4%). The KSS correction applies the leave-one-out estimator to remove limited mobility bias from $\text{Var}(\hat{\alpha})$ and $\text{Var}(\hat{\psi})$.

The Zipf simulation closely replicates the dominant feature of the original AKM decomposition: the large worker effect. The plug-in $\text{Var}(\hat{\alpha})$ is 83.8% in the Zipf simulation versus 85.9% in the original — a gap of just 2.1 percentage points. The uniform simulation substantially overshoots at 96.1%.

The mechanism runs through the firm effect. In the uniform simulation, every firm within a type has the same number of spells, so the AKM cleanly identifies firm fixed effects that map almost exactly onto the 12 CEM type means. This inflates $\text{Var}(\hat{\psi})$ to 46.6% (vs. 27.7% in the original) and produces a correspondingly large negative covariance (−48.9%), which mechanically pushes $\text{Var}(\hat{\alpha})$ upward.

The Zipf distribution breaks this clean mapping: many firms have only one or two spells and thus noisy fixed effect estimates, while a handful of mega-firms dominate the connected component. The resulting within-type dispersion in $\hat{\psi}$ better mimics the genuine within-type heterogeneity present among 757K real firms, attenuating $\text{Var}(\hat{\psi})$ toward its true level and keeping the worker effect in the right range.

The KSS correction is informative about the limited mobility bias present in each dataset. In the original data, the correction reduces $\text{Var}(\hat{\alpha})$ substantially — from 85.9% to 74.2%, a reduction of 11.7 percentage points — reflecting genuine limited mobility bias that arises when workers are observed at only a handful of firms over the panel. The correction is much smaller in both simulations: −5.9pp for uniform and −3.6pp for Zipf. This asymmetry is expected: simulated workers visit a smaller, more evenly connected set of firms, so the movers used to identify firm effects are less selected and less sparse than in the full DADS. The Zipf simulation, with its heavier concentration of workers at large firms, comes closest to replicating the original correction magnitude. After the KSS correction, the Zipf worker effect (80.2%) remains within 6 percentage points of the original (74.2%), while the uniform overshoots by 16 percentage points.

The residual $\text{Var}(\hat{\varepsilon})$ is roughly half the original level in both simulations (6–7% vs. 13.2%). This gap is structural: wages in the simulation are generated with a single noise draw σ_ε per spell, whereas real wages contain additional within-spell variation (hours fluctuations, bonuses, measurement error) that the model abstracts from. The KSS correction leaves the residual unchanged, as expected.

A2.5.6 Supporting plots

We now present some graphs comparing the mean wages, firm count per type, and mean firm size in each of the original and simulated datasets. These graphs illustrate the quality of the re-estimation and the simulation.

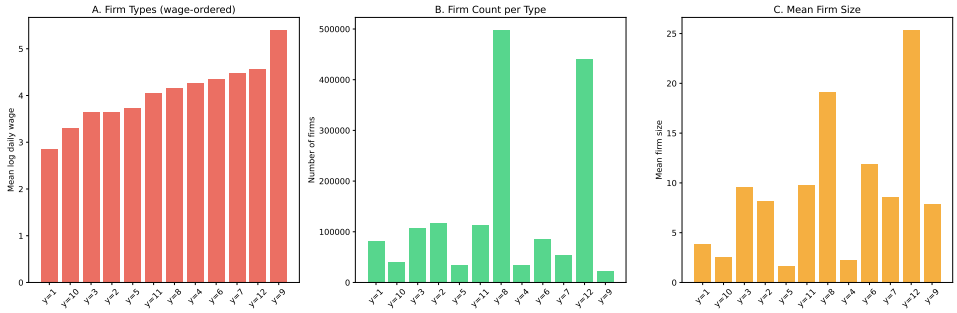


Figure A2.8. On original dataset.

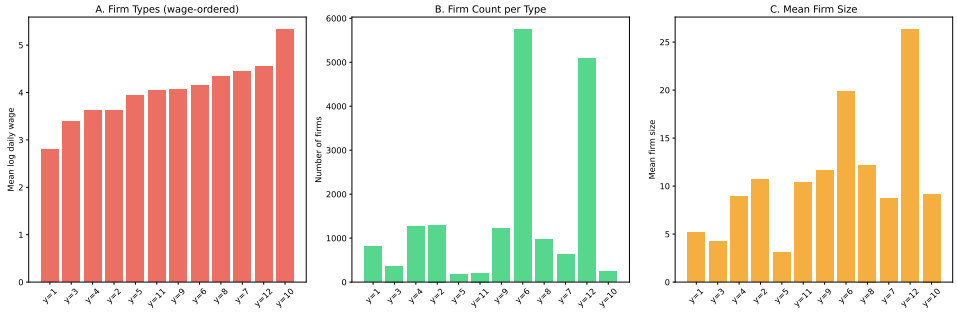


Figure A2.9. On uniform simulated dataset.

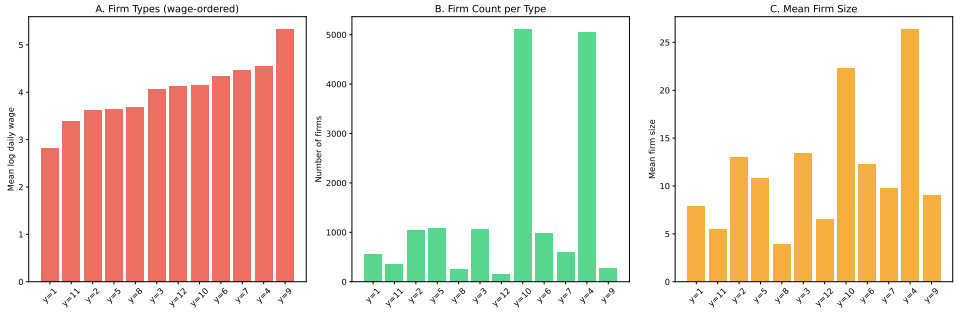


Figure A2.10. On Zipf simulated dataset.

Bibliography

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

Damien Babet, Olivier Godechot, and Marco Palladino. In the land of akm: The performance of the akm estimator in finite samples. *Labour Economics*, 83:102388, 2023.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer-employee data. *Econometrica*, 83(3):699–739, 2019.

David Card, Jorg Heining, and Patrick Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015, 2013.

Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Leave-Out Estimation of Variance Components. *Econometrica*, 88(5):1859–1898, September 2020.

Appendix to Chapter 3

A3.1 Relevance of the AKM Methodology

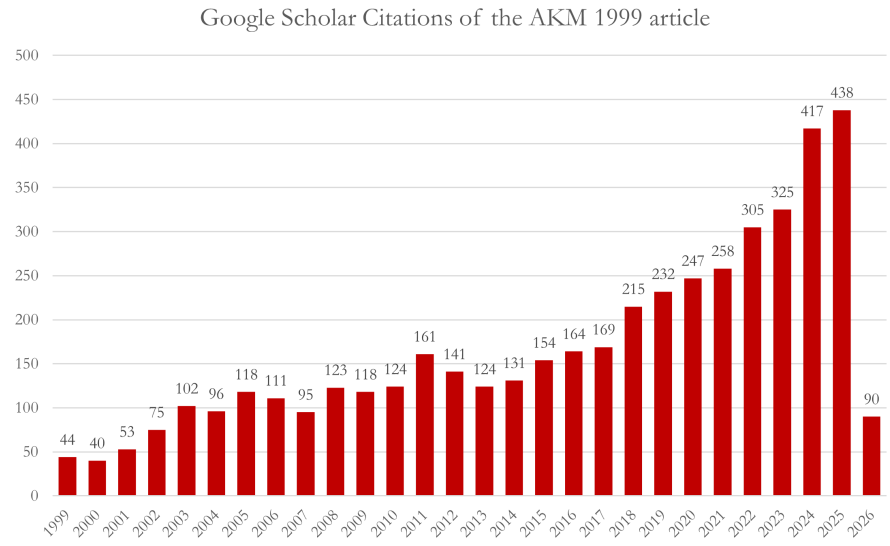


Figure A3.1. Google Scholar citations of the Abowd et al. (1999) article

A3.2 Proofs

A3.2.1 Proof of Proposition 1

I use the notation of Section 3.2.1. The data consist of $n = NT$ stacked observations. The worker indicator matrix $\mathbf{D} \in \{0,1\}^{n \times N}$, the firm indicator matrix $\mathbf{F} \in \{0,1\}^{n \times J}$, and the within-worker projector $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ are as defined in Section 3.2.1. We assume each worker appears at least once ($T_i > 0$ for all i), so $\mathbf{D}'\mathbf{D}$ is invertible.

Positive semidefiniteness of $\mathbf{S}$. Define $\mathbf{S} = \mathbf{F}'\mathbf{M}_D\mathbf{F} \in \mathbb{R}^{J \times J}$. Because $\mathbf{M}_D$ is a symmetric idempotent matrix ($\mathbf{M}_D = \mathbf{M}_D' = \mathbf{M}_D^2$), for any $\mathbf{x} \in \mathbb{R}^J$:

$$\mathbf{x}'\mathbf{S}\mathbf{x} = \mathbf{x}'\mathbf{F}'\mathbf{M}_D\mathbf{F}\mathbf{x} = (\mathbf{F}\mathbf{x})'\mathbf{M}_D(\mathbf{F}\mathbf{x}) = (\mathbf{F}\mathbf{x})'\mathbf{M}_D'\mathbf{M}_D(\mathbf{F}\mathbf{x}) = \|\mathbf{M}_D\mathbf{F}\mathbf{x}\|_2^2 \geq 0.$$

Therefore $\mathbf{S}$ is symmetric and positive semidefinite. □

Entrywise form and Laplacian representation. Expand the concentrated form:

$$\mathbf{S} = \mathbf{F}'\mathbf{M}_D\mathbf{F} = \mathbf{F}'\mathbf{F} - \mathbf{F}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{F}.$$

Index observations by (i,t) with $i \in \{1,\dots,N\}$ and $t \in \{1,\dots,T_i\}$. Let $n_{ij} = \#\{t : j(i,t) = j\}$ denote the number of periods worker i spends at firm j, and $T_i = \sum_{j=1}^{J} n_{ij}$ the total periods for worker i.

First term. Since $(\mathbf{F}'\mathbf{F})_{jk} = \sum_i \sum_{t=1}^{T_i} \mathbf{1}\{j(i,t) = j\}\,\mathbf{1}\{j(i,t) = k\}$, a given observation can only be at one firm, so:

$$(\mathbf{F}'\mathbf{F})_{jk} = \begin{cases} \sum_{i=1}^{N} n_{ij}, & j = k, \\ 0, & j \neq k. \end{cases}$$

Second term. Since $\mathbf{D}'\mathbf{D}$ is diagonal with $(\mathbf{D}'\mathbf{D})_{ii} = T_i$:

$$\big(\mathbf{F}'\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'\mathbf{F}\big)_{jk} = \sum_{i=1}^{N} \frac{1}{T_i}\Big(\sum_{t=1}^{T_i} \mathbf{1}\{j(i,t) = j\}\Big)\Big(\sum_{t=1}^{T_i} \mathbf{1}\{j(i,t) = k\}\Big) = \sum_{i=1}^{N} \frac{n_{ij}\,n_{ik}}{T_i}.$$

Combining. Subtracting gives, for all $j, k \in \{1,\dots,J\}$:

$$S_{jk} = \sum_{i=1}^{N}\Big(n_{ij}\,\mathbf{1}\{j = k\} - \frac{n_{ij}\,n_{ik}}{T_i}\Big),$$

yielding the diagonal and off-diagonal entries:

$$S_{jj} = \sum_{i=1}^{N} n_{ij}\Big(1 - \frac{n_{ij}}{T_i}\Big), \qquad S_{jk} = -\sum_{i=1}^{N} \frac{n_{ij}\,n_{ik}}{T_i} \quad (j \neq k).$$

Laplacian form. Define the symmetric weight matrix $\mathbf{W} \in \mathbb{R}^{J\times J}$ by $W_{jk} = \sum_{i=1}^{N} n_{ij}n_{ik}/T_i$ for $j \neq k$ and $W_{jj} = 0$, and let $d_j = \sum_{k\neq j} W_{jk}$. Then:

$$d_j = \sum_{k\neq j}\sum_{i=1}^{N} \frac{n_{ij}n_{ik}}{T_i} = \sum_{i=1}^{N} \frac{n_{ij}}{T_i}\sum_{k\neq j} n_{ik} = \sum_{i=1}^{N} n_{ij}\Big(1 - \frac{n_{ij}}{T_i}\Big) = S_{jj},$$

so $\mathbf{S} = \operatorname{diag}(d_1,\dots,d_J) - \mathbf{W}$, the standard form of a weighted graph Laplacian on the firm graph with edge weights W_{jk}. The off-diagonal entries are nonpositive and each row sums to zero because for each fixed j and each worker i:

$$\sum_{k=1}^{J}\Big(n_{ij}\,\mathbf{1}\{j = k\} - \frac{n_{ij}n_{ik}}{T_i}\Big) = n_{ij} - \frac{n_{ij}}{T_i}\sum_{k=1}^{J} n_{ik} = n_{ij} - n_{ij} = 0.$$

□

Null space, rank, and uniqueness up to additive constants. From $\mathbf{F}\mathbf{1}_J = \mathbf{1}_n$ (each observation belongs to exactly one firm) and $\mathbf{M}_D\mathbf{1}_n = \mathbf{0}$ (demeaning removes within-worker constants):

$$\mathbf{S}\mathbf{1}_J = \mathbf{F}'\mathbf{M}_D\mathbf{F}\mathbf{1}_J = \mathbf{F}'\mathbf{M}_D\mathbf{1}_n = \mathbf{0},$$

so $\mathbf{1}_J \in \ker(\mathbf{S})$.

More generally, if the firm graph has C connected components with indicator vectors $\mathbf{1}_{c_\ell}$ for $\ell = 1, \ldots, C$, then for any worker i whose firms all belong to component ℓ, the vector $\mathbf{F}\mathbf{1}_{c_\ell}$ is constant across all of worker i's observations. Hence $\mathbf{M}_D\mathbf{F}\mathbf{1}_{c_\ell} = \mathbf{0}$ for such workers, giving $\mathbf{S}\mathbf{1}_{c_\ell} = \mathbf{0}$ for $\ell = 1, \ldots, C$. Therefore $\dim\ker(\mathbf{S}) \geq C$.

Conversely, the Laplacian identity $\mathbf{x}'\mathbf{S}\mathbf{x} = \frac{1}{2}\sum_{j,k} W_{jk}(x_j - x_k)^2$ implies that if $\mathbf{x}'\mathbf{S}\mathbf{x} = 0$ then $x_j = x_k$ whenever $W_{jk} > 0$. By connectivity within each component, $\mathbf{x}$ must be constant on each connected component. Hence $\dim\ker(\mathbf{S}) = C$ and $\operatorname{rank}(\mathbf{S}) = J - C$.

When the firm graph is connected ($C = 1$), imposing $\sum_{j=1}^{J} \hat{\psi}_j = 0$ or equivalently using $\mathbf{S}^+$ restricted to $\ker(\mathbf{S})^\perp = \{\mathbf{1}_J\}^\perp$ yields the unique minimum-norm solution. □

A3.2.2 Proof of Proposition 2

Part (a): Expected node degree under general Markov mobility. Consider a balanced panel of N workers over T periods at J firms ($T_i = T$ for all i). Let $X_{i,t} \in \{1, \ldots, J\}$ denote the firm employing worker i at time t. Workers move independently according to a common, time-homogeneous Markov chain with transition matrix $\mathbf{Q} \in [0,1]^{J \times J}$, where $Q_{k,\ell} = \Pr(X_{i,t+1} = \ell \mid X_{i,t} = k)$. The initial allocation at $t = 1$ is deterministic and given by $\{n_k^{\text{start}}\}_{k=1}^{J}$, where $\sum_k n_k^{\text{start}} = N$.

For a given firm j, the node degree is $d_j = \sum_{i=1}^{N} n_{ij}(1 - n_{ij}/T)$. The average node degree is $\bar{d} = \frac{1}{J}\sum_{j=1}^{J} d_j = \operatorname{tr}(\mathbf{S})/J$.

By linearity of expectation and the fact that workers are independent conditional on their starting firm, the expected degree of firm j is

$$\begin{aligned}
\mathrm{E}[d_j] &= \sum_{i=1}^{N} \mathrm{E}\left[n_{ij} - \frac{n_{ij}^2}{T}\right] \\
&= \sum_{k=1}^{J} \sum_{i:X_{i,1}=k} \mathrm{E}\left[n_{ij} - \frac{n_{ij}^2}{T} \,\middle|\, X_{i,1} = k\right] \\
&= \sum_{k=1}^{J} n_k^{\text{start}}\, C_{k\to j},
\end{aligned}$$

where $C_{k\to j} \equiv \mathrm{E}[n_{ij} \mid X_{i,1} = k] - \frac{1}{T}\mathrm{E}[n_{ij}^2 \mid X_{i,1} = k]$ is the expected contribution of a worker starting at firm k.

Step 1: First moment of occupancy. Let $\mathbf{G} = \sum_{t=0}^{T-1} \mathbf{Q}^t$ be the finite-horizon fundamental matrix. The probability of worker i being at firm j at time t, given they started at k, is $(\mathbf{e}_k \mathbf{Q}^{t-1})_j$, where $\mathbf{e}_k$ is the k-th standard basis (row) vector. Summing over t:

$$\mu_{k \to j} \equiv \mathrm{E}[n_{ij} \mid X_{i,1} = k] = \sum_{t=1}^{T} (\mathbf{e}_k \mathbf{Q}^{t-1})_j = G_{kj}.$$

Step 2: Second moment with finite-horizon correction. We expand the square of the occupancy sum:

$$\mathrm{E}[n_{ij}^2 \mid X_{i,1} = k] = \mathrm{E}\left[\sum_{s=1}^{T} \mathbf{1}_{\{X_{i,s}=j\}} + 2 \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \mathbf{1}_{\{X_{i,s}=j, X_{i,t}=j\}} \;\middle|\; X_{i,1} = k \right].$$

By the Markov property, the joint probability in the double sum is $\Pr(X_{i,s} = j \mid X_{i,1} = k) \Pr(X_{i,t} = j \mid X_{i,s} = j)$. Let $u = t - s$:

$$\mathrm{E}[n_{ij}^2 \mid X_{i,1} = k] = G_{kj} + 2 \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} \sum_{u=1}^{T-s} (\mathbf{Q}^u)_{jj}.$$

Define $f_j(H) = \sum_{u=1}^{H} (\mathbf{Q}^u)_{jj}$ as the expected return occupancy to firm j within a remaining horizon of H periods. Then:

$$\mathrm{E}[n_{ij}^2 \mid X_{i,1} = k] = G_{kj} + 2 \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} f_j(T - s).$$

Step 3: Expected contribution $C_{k \to j}$. Substituting these moments into the definition of $C_{k \to j}$:

$$C_{k \to j} = G_{kj} - \frac{1}{T} \left[G_{kj} + 2 \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} f_j(T - s) \right].$$

Simplifying the G_{kj} terms yields:

$$C_{k \to j} = \frac{T-1}{T} G_{kj} - \frac{2}{T} \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} f_j(T - s).$$

Step 4: Expected degree and expected average degree. Summing over the initial allocation n_k^{start} gives the exact expression:

$$\boxed{\mathrm{E}[d_j] = \frac{T-1}{T} \sum_{k=1}^{J} n_k^{\text{start}} G_{kj} - \frac{2}{T} \sum_{k=1}^{J} n_k^{\text{start}} \sum_{s=1}^{T-1} (\mathbf{Q}^{s-1})_{kj} f_j(T - s).} \tag{A3.2.1}$$

Equation (A3.2.1) is exact for any $T \geq 2$, any transition matrix $\mathbf{Q}$, and any initial allocation. Finally, the expected average node degree is

$$\mathrm{E}[\bar{d}] = \frac{1}{J}\sum_{j=1}^{J}\mathrm{E}[d_j] = \frac{1}{J}\sum_{j=1}^{J}\sum_{k=1}^{J} n_k^{\text{start}} C_{k\to j}, \tag{A3.2.2}$$

which is exactly the statement in part (a). Part (b) follows immediately by setting $n_k^{\text{start}} = N/J$.

□

Part (c): Inverse proportionality. Recall from Section 3.2.2 that $\mathbf{N}_F = \mathrm{diag}(n_1, \ldots, n_J)$ and $\mathbf{n}_F = (n_1, \ldots, n_J)'$, where $n_j = \sum_i n_{ij}$ is the total number of observations at firm j. Under the regular network assumption that every firm has the same number of observations, $n_j = NT/J$ for all j (because each of the N/J workers contributes T periods). Hence $\mathbf{N}_F = (NT/J)\mathbf{I}_J$ and $\mathbf{n}_F = (NT/J)\mathbf{1}_J$.

The centering correction in (3.2.5) vanishes:

$$\mathbf{n}_F'\mathbf{S}^+\mathbf{n}_F = (NT/J)^2\mathbf{1}_J'\mathbf{S}^+\mathbf{1}_J = 0,$$

since $\mathbf{S}^+\mathbf{1}_J = \mathbf{0}$. Therefore

$$B_\psi = \frac{1}{n}\,\mathrm{tr}(\mathbf{S}^+\mathbf{N}_F) = \frac{1}{NT}\,\mathrm{tr}\left(\mathbf{S}^+ \cdot \frac{NT}{J}\mathbf{I}_J\right) = \frac{1}{J}\,\mathrm{tr}(\mathbf{S}^+).$$

Using the spectral decomposition $\mathbf{S}^+ = \sum_{k=2}^{J}\lambda_k^{-1}\mathbf{v}_k\mathbf{v}_k'$, $\mathrm{tr}(\mathbf{S}^+) = \sum_{k=2}^{J}\lambda_k^{-1}$, so $B_\psi = \frac{1}{J}\sum_{k=2}^{J}\lambda_k^{-1}$.

The arithmetic-harmonic mean inequality applied to the $J-1$ positive eigenvalues $\lambda_2, \ldots, \lambda_J$ gives

$$\frac{1}{J-1}\sum_{k=2}^{J}\frac{1}{\lambda_k} \;\geq\; \frac{J-1}{\sum_{k=2}^{J}\lambda_k} = \frac{J-1}{\mathrm{tr}(\mathbf{S})},$$

because $\lambda_1 = 0$. Hence

$$B_\psi = \frac{1}{J}\sum_{k=2}^{J}\lambda_k^{-1} \geq \frac{(J-1)^2}{J\,\mathrm{tr}(\mathbf{S})} = \frac{(J-1)^2}{J^2\bar{d}},$$

where $\bar{d} = \mathrm{tr}(\mathbf{S})/J$ is the average node degree (Definition 1). □

Remark (Asymptotics and Skewness). *The bound in Proposition 3.2.11 provides two key insights. First, as the average connectivity $\bar{d} \to \infty$, the bias vanishes. Second, while the bound is tightest for regular networks, departures from this benchmark (such as Zipf-distributed firm sizes) introduce a strictly positive centering correction $\frac{1}{n^2}\mathbf{n}_F'\mathbf{S}^+\mathbf{n}_F$, mechanically reducing the bias B_ψ.*

Part (d): Mohar bound. This is a corollary from Mohar (1991, Proposition 2.1). Recall from Proposition 1 that for $j \neq k$, $S_{jk} = -W_{jk}$, where W_{jk} is the edge weight in the projected graph. Define the test vector $\mathbf{x} \in \mathbb{R}^J$ by $x_v = 1$ at the vertex v attaining $d_{\min} = \min_j S_{jj}$, and $x_j = -1/(J-1)$ for all

$j \neq v$. Then $\sum_j x_j = 1 - (J-1)/(J-1) = 0$, so $\mathbf{x} \perp \mathbf{1}_J$. The Rayleigh quotient gives

$$\lambda_2 \;\leq\; \frac{\mathbf{x}'\mathbf{S}\mathbf{x}}{\mathbf{x}'\mathbf{x}}.$$

The numerator is $\mathbf{x}'\mathbf{S}\mathbf{x} = \sum_{j<k} W_{jk}(x_j - x_k)^2$. Only terms involving v contribute a factor $(1 + 1/(J-1))^2 = (J/(J-1))^2$; all other pairs contribute $(1/(J-1) - 1/(J-1))^2 = 0$ (since their x-values are identical). Hence

$$\mathbf{x}'\mathbf{S}\mathbf{x} = \frac{J^2}{(J-1)^2} \sum_{k \neq v} W_{vk} = \frac{J^2}{(J-1)^2}\, d_v = \frac{J^2}{(J-1)^2}\, d_{\min}.$$

The denominator is $\mathbf{x}'\mathbf{x} = 1 + (J-1)/(J-1)^2 = J/(J-1)$. Dividing gives $\lambda_2 \leq (J/(J-1))\, d_{\min}$. □

A3.3 Additional Monte Carlo Tables

Table A3.1. Variance-share distortions at 10% mobility (Zipf firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	Andr.	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	0.80	0.35	4.58	15	-1	2	-12	24	23	1	-56	-57	14
10	0.84	0.66	7.62	14	3	5	4	17	17	-6	-42	-43	-7
15	0.84	0.70	11.31	12	5	5	5	15	15	-5	-30	-31	-6
30	2.60	2.50	22.10	1	-2	-2	-1	13	13	-1	-8	-8	1
50	4.80	4.67	36.96	2	0	0	-1	12	12	0	-5	-5	1
100	12.84	12.56	73.75	2	1	1	1	10	10	0	-4	-4	-1
500	77.04	77.29	369.10	0	0	0	0	10	10	0	-1	-1	0
Panel B: $J = 100$													
5	0.80	0.32	4.82	14	-2	2	-11	23	22	-5	-59	-60	4
10	0.80	0.42	8.00	8	-2	0	-5	18	18	-3	-30	-31	1
15	0.80	0.69	11.40	7	0	1	-2	15	15	-3	-19	-20	3
30	1.40	1.34	22.18	3	0	0	0	12	12	-1	-10	-10	1
50	2.80	2.70	36.97	0	-1	-1	-1	11	11	-1	-4	-4	1
100	9.52	9.34	73.83	0	0	0	0	11	11	0	-3	-3	0
500	64.28	64.13	369.17	0	0	0	0	10	10	0	0	0	0
Panel C: $J = 500$													
5	0.80	0.17	5.14	13	-1	2	-6	20	20	-2	-45	-46	-2
10	0.80	0.25	8.19	8	0	2	0	16	16	-2	-28	-28	0
15	0.80	0.36	11.68	6	0	1	0	15	15	-2	-20	-21	0
30	0.80	0.71	22.25	3	0	0	0	12	12	-1	-10	-10	0
50	1.12	1.09	37.05	2	0	0	0	12	11	-1	-6	-6	0
100	4.08	4.03	74.06	1	0	0	0	11	11	0	-3	-3	0
500	41.36	41.13	370.30	0	0	0	0	10	10	0	0	0	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_\psi = \sigma^2_\alpha = 0.50$, $\sigma^2_\varepsilon = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct $\widehat{\mathrm{Var}}(\alpha)$ by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.*

Table A3.2. Variance-share distortions at 20% mobility (Zipf firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	Andr.	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	0.80	0.61	7.19	15	4	5	-1	18	18	-5	-54	-56	-7
10	1.16	1.03	13.69	7	0	0	-2	15	15	0	-18	-19	-4
15	2.16	2.09	20.53	2	-1	-1	-1	11	11	-2	-9	-10	-1
30	6.44	6.28	40.92	1	0	0	0	11	11	-1	-4	-5	0
50	11.80	11.64	68.19	1	0	0	0	11	11	0	-2	-2	1
100	25.20	25.20	135.98	1	1	1	1	10	10	0	-2	-2	-1
500	154.28	154.81	680.55	0	0	0	0	10	10	0	0	0	0
Panel B: $J = 100$													
5	0.80	0.45	7.33	8	-2	0	-2	18	17	-3	-38	-40	0
10	0.84	0.73	13.81	3	-3	-2	-2	12	12	-3	-16	-17	-3
15	1.44	1.43	20.56	2	-1	-1	-1	12	12	-3	-12	-12	1
30	4.84	4.67	40.92	1	-1	-1	-1	11	11	-1	-3	-4	2
50	9.08	8.96	68.30	1	0	0	0	11	11	0	-2	-2	1
100	20.68	20.56	136.75	0	0	0	0	11	11	0	-1	-1	1
500	125.64	125.86	683.18	1	0	0	0	10	10	0	-1	-1	-1
Panel C: $J = 500$													
5	0.80	0.25	7.65	7	-1	1	-1	16	15	-4	-28	-29	1
10	0.80	0.49	13.92	4	0	0	-1	14	13	-3	-18	-18	-2
15	0.80	0.61	20.64	3	0	0	0	12	12	-2	-11	-12	-1
30	1.52	1.50	41.15	1	0	0	0	11	11	-1	-4	-5	0
50	4.64	4.55	68.55	1	0	0	0	10	10	-1	-3	-3	0
100	13.12	13.01	137.15	0	0	0	0	10	10	0	-1	-1	0
500	88.56	88.37	685.64	0	0	0	0	10	10	0	0	0	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_{\psi} = \sigma^2_{\alpha} = 0.50$, $\sigma^2_{\varepsilon} = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct* $\widehat{\mathrm{Var}}(\alpha)$ *by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.*

Table A3.3. Variance-share distortions at 40% mobility (Zipf firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	Andr.	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	0.96	0.81	11.71	8	2	2	3	15	15	-5	-36	-39	-8
10	2.84	2.62	23.28	3	1	1	1	12	12	-3	-9	-11	4
15	5.20	5.07	34.86	2	0	0	0	10	10	-3	-6	-7	2
30	13.20	13.07	69.81	1	0	0	0	10	10	-2	-7	-7	-4
50	22.84	22.78	116.40	0	-1	-1	-1	10	10	-1	0	-1	2
100	50.40	50.62	232.70	0	0	0	0	10	10	0	0	0	1
500	272.36	274.23	1163.78	0	0	0	0	10	10	0	-1	-1	0
Panel B: $J = 100$													
5	0.80	0.67	11.82	2	-3	-3	-2	12	12	-7	-18	-20	3
10	1.88	1.77	23.35	5	2	2	1	11	10	-3	-11	-13	-2
15	3.76	3.56	35.07	1	0	0	0	12	12	-1	-5	-6	0
30	9.36	9.27	70.38	1	0	0	0	10	10	-2	-3	-4	0
50	19.20	19.11	117.20	1	0	0	0	10	10	0	-3	-3	-1
100	41.44	41.37	234.29	0	0	0	0	10	10	0	-1	-2	-1
500	228.28	228.61	1171.26	0	0	0	0	10	10	0	0	0	0
Panel C: $J = 500$													
5	0.80	0.44	11.94	7	1	1	0	12	11	-7	-23	-25	-3
10	0.84	0.80	23.61	2	-1	-1	-1	11	11	-3	-8	-10	2
15	1.28	1.26	35.41	2	0	0	0	11	11	-2	-7	-8	0
30	5.56	5.48	70.76	1	0	0	0	10	10	-1	-2	-3	1
50	10.56	10.49	117.96	0	0	0	0	10	10	0	-1	-2	0
100	25.68	25.59	235.86	0	0	0	0	10	10	0	0	-1	0
500	159.24	159.17	1179.62	0	0	0	0	10	10	0	0	0	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_{\psi} = \sigma^2_{\alpha} = 0.50$, $\sigma^2_{\varepsilon} = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct* $\widehat{\mathrm{Var}}(\alpha)$ *by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.*

Table A3.4. Variance-share distortions at 5% mobility (Uniform firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	Andr.	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	0.80	0.08	2.21	76	-12	20	§	75	73	-22	-269	-269	-23
10	0.84	0.42	3.89	23	1	1	-19	33	32	-4	-79	-80	41
15	1.44	1.01	5.87	16	3	3	-7	20	20	-7	-48	-48	13
30	4.64	3.84	11.56	9	4	4	5	14	14	-1	-21	-22	-8
50	10.52	9.21	19.21	3	0	0	-1	12	12	-1	-8	-8	3
100	25.28	23.19	38.48	-1	-2	-2	-2	13	13	1	-2	-2	3
500	163.04	159.82	192.39	0	0	0	0	10	10	0	-1	-1	0
Panel B: $J = 100$													
5	0.80	0.06	2.33	78	-1	12	-81	76	74	21	-280	-281	-42
10	0.80	0.36	3.91	26	0	2	-30	34	34	-8	-90	-91	25
15	1.00	0.75	5.79	12	-2	-2	-7	23	22	-4	-42	-42	13
30	4.00	3.30	11.53	3	-2	-2	-1	15	15	-1	-15	-15	2
50	9.36	8.14	19.29	3	0	0	0	12	12	0	-11	-11	-2
100	23.64	21.94	38.47	1	0	0	0	11	11	0	-5	-5	0
500	156.64	153.28	192.51	1	0	0	0	10	10	0	-2	-2	-1
Panel C: $J = 500$													
5	0.80	0.03	2.31	88	3	17	-17	89	86	21	-303	-304	59
10	0.80	0.21	3.93	25	0	3	-18	35	34	-9	-89	-89	15
15	0.80	0.50	5.79	13	0	0	-5	23	22	-6	-47	-47	8
30	2.68	2.27	11.55	5	0	0	0	15	15	-1	-19	-19	-2
50	7.52	6.91	19.27	3	0	0	-1	13	13	-1	-9	-9	2
100	21.72	20.18	38.52	1	0	0	0	11	11	0	-4	-5	0
500	154.92	150.12	192.62	0	0	0	0	10	10	0	-1	-1	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_{\psi} = \sigma^2_{\alpha} = 0.50$, $\sigma^2_{\varepsilon} = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct* $\widehat{\mathrm{Var}}(\alpha)$ *by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.* § *not computed because the random halving of workers leaves too few movers in each split to maintain a common Largest Connected Set across both estimation samples.*

Table A3.5. Variance-share distortions at 10% mobility (Uniform firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	Andr.	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	0.80	0.42	3.81	34	9	11	-40	38	37	-14	-114	-116	20
10	2.44	1.87	7.40	12	3	3	4	16	16	-7	-36	-37	3
15	4.52	3.67	11.09	5	-1	-1	-1	15	15	-1	-20	-21	-1
30	12.64	11.34	22.19	2	-1	-1	0	13	13	0	-7	-7	0
50	25.80	23.17	37.22	1	-1	-1	-1	11	11	0	-4	-5	1
100	56.20	53.42	73.85	1	0	0	0	10	10	-1	-3	-3	-1
500	331.20	327.76	370.69	0	0	0	0	10	10	0	-1	-1	0
Panel B: $J = 100$													
5	0.80	0.32	3.81	28	4	6	-22	31	30	-8	-98	-99	31
10	1.96	1.55	7.41	9	-1	-1	-1	20	20	-3	-37	-38	1
15	3.88	3.32	11.09	5	-1	-1	0	15	15	-3	-21	-21	-1
30	12.68	10.89	22.20	2	-1	-1	-1	12	12	-1	-6	-6	3
50	22.64	20.96	37.03	1	0	0	0	11	11	-1	-5	-5	1
100	55.04	51.92	74.16	1	1	1	0	11	11	0	-3	-3	0
500	328.68	322.79	370.93	0	0	0	0	10	10	0	0	0	0
Panel C: $J = 500$													
5	0.80	0.23	3.78	24	-1	2	-12	35	34	-24	-94	-95	3
10	1.12	0.95	7.45	7	-2	-2	-3	19	19	-4	-30	-31	3
15	2.76	2.43	11.14	5	-1	-1	-1	15	15	-3	-19	-20	1
30	9.68	8.85	22.25	2	0	0	0	12	12	-2	-8	-8	1
50	21.04	19.52	37.08	1	0	0	0	11	11	-1	-5	-5	1
100	50.84	48.33	74.28	1	0	0	0	11	11	0	-2	-2	1
500	320.88	314.16	371.05	0	0	0	0	10	10	0	0	0	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_{\psi} = \sigma^2_{\alpha} = 0.50$, $\sigma^2_{\varepsilon} = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct* $\widehat{\mathrm{Var}}(\alpha)$ *by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.*

Table A3.6. Variance-share distortions at 20% mobility (Uniform firm sizes, size-weighted mobility)

	Diagnostic			$\widehat{\mathrm{Var}}(\psi)$ distortion (%)				$\widehat{\mathrm{Var}}(\alpha)$ distortion (%)			$2\widehat{\mathrm{Cov}}$ distortion (%)		
N/J	$d_{\min}\mathbb{S}$	λ_2	$\bar{d}$	Plug-in	Andr.	KSS	Split	Plug-in†	KSS	Split	Plug-in‡	KSS	Split
Panel A: $J = 50$													
5	2.28	1.68	6.89	13	3	3	3	17	16	-12	-39	-41	2
10	6.44	5.61	13.76	6	1	1	0	11	11	-6	-17	-18	3
15	12.68	11.03	20.64	3	0	0	0	12	12	-2	-10	-11	-1
30	28.68	26.61	41.19	2	0	0	1	11	11	-1	-1	-2	4
50	53.04	50.79	68.71	1	0	0	0	11	11	0	-3	-3	0
100	115.16	112.23	137.44	1	1	1	1	10	10	0	-2	-2	-1
500	638.84	636.97	687.44	0	0	0	0	10	10	0	0	0	0
Panel B: $J = 100$													
5	2.00	1.55	6.91	10	0	0	1	17	17	-10	-43	-45	-2
10	6.52	5.59	13.84	5	1	1	1	13	13	-4	-16	-17	-1
15	11.72	10.46	20.67	3	1	1	1	11	11	-2	-14	-15	-4
30	27.24	25.59	41.22	2	0	0	0	10	10	-2	-6	-6	-1
50	49.20	47.10	68.73	1	0	0	0	10	10	-1	-3	-3	0
100	109.44	106.43	137.59	1	0	0	0	10	10	0	-2	-2	0
500	630.00	624.70	687.88	0	0	0	0	10	10	0	0	0	0
Panel C: $J = 500$													
5	1.32	1.07	6.89	10	1	1	0	18	18	-9	-39	-41	-5
10	4.48	4.02	13.80	3	-1	-1	-1	13	13	-3	-14	-15	1
15	9.76	8.86	20.72	2	0	0	0	12	12	-2	-10	-11	0
30	24.08	22.70	41.34	1	0	0	0	11	11	-1	-4	-5	0
50	47.72	45.56	68.91	1	0	0	0	11	11	0	-3	-3	0
100	106.96	103.79	137.73	0	0	0	0	10	10	0	-1	-2	0
500	620.48	612.81	689.29	0	0	0	0	10	10	0	0	0	0

Notes. Number of workers were adjusted for each design. Each cell reports the average distortion over 10 replications of the estimated variance component from the true variance component in percentage. DGP: sorting with $\rho = 0.3$, $\sigma^2_\psi = \sigma^2_\alpha = 0.50$, $\sigma^2_\varepsilon = 0.25$, $T = 5$. †*Homoskedastic (HO) worker-effect columns omitted: method does not correct* $\widehat{\mathrm{Var}}(\alpha)$ *by construction.* ‡*HO covariance columns omitted: corrections are identical to plug-in at reported precision.*

Bibliography

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

Bojan Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, editors, *Graph Theory, Combinatorics, and Applications*, volume 2, pages 871–898. Wiley, 1991.

Institut d'études politiques de Paris
École Doctorale de Sciences Po
Programme doctoral en économie
Département d'économie
Doctorat en Sciences économiques

Analyse quantitative de la dynamique du marché du travail

Riddhi KALSI

Thèse dirigée par
Jean-Marc ROBIN, Professeur des Universités, Sciences Po

soutenance le juin 11, 2026

Jury:
Clément de CHAISEMARTIN, Professeur des Universités, Sciences Po
Gregory JOLIVET, Professor of Economics, University of Bristol *(reviewer)*
Costas MEGHIR, Professor of Economics, Yale University
Hélène PÉRIVIER, Directrice de recherche OFCE -Sciences Po
Fabien POSTEL-VINAY, Professor of Economics, University College London *(reviewer)*
Jean-Marc ROBIN, Professeur des Universités, Sciences Po

Résumé en Français

Les inégalités salariales demeurent l'un des phénomènes les plus persistants et les plus étudiés en économie du travail. Démêler les sources de la dispersion des salaires nécessite de comprendre une interaction complexe entre les cadres institutionnels, les politiques salariales des entreprises, les caractéristiques des travailleurs et l'appariement assortatif (assortative matching) entre les différentes composantes du marché du travail. Historiquement, la littérature a abordé les écarts de salaires sous l'angle de la théorie du capital humain, attribuant la variation principalement à des traits observables tels que l'éducation, l'expérience et les compétences (Mincer, 1974; Becker, 1964). Cependant, les données empiriques démontrent systématiquement que les caractéristiques observables n'expliquent qu'une fraction de la variance totale des salaires, laissant un écart résiduel substantiel déterminé par l'hétérogénéité inobservée et les structures institutionnelles.

Cette thèse examine la nature multidimensionnelle des inégalités salariales en traçant un cheminement analytique allant des frontières institutionnelles du marché du travail à l'économétrie structurelle des réseaux travailleurs-entreprises-professions. La recherche présentée ici se déploie comme une progression intellectuelle, passant d'une énigme empirique spécifique concernant les écarts de salaires sectoriels à un modèle statistique de tri multidimensionnel (multidimensional sorting), puis à des questions fondamentales sur les outils économétriques les plus couramment utilisés pour décomposer la variance des salaires. En évaluant l'écart de rémunération public-privé, en estimant un modèle multidimensionnel flexible des revenus sur le cycle de vie, puis en diagnostiquant la fiabilité des estimateurs canoniques de décomposition de la variance, cette thèse contribue à expliquer comment s'opère le tri sur le marché du travail et comment il est mesuré.

Frontières institutionnelles : l'écart de rémunération public-privé

La première interrogation de cette thèse examine comment de larges frontières institutionnelles façonnent les résultats économiques à long terme. La division entre les secteurs public et privé agit comme une dualité du marché du travail, particulièrement dans les économies de marché coordonnées. Le premier chapitre, *Évaluer l'écart de rémunération dans le secteur public selon le genre*, aborde une énigme empirique dans la littérature sur les salaires public-privé : la persistance de conclusions contradictoires quant à savoir si l'emploi public génère une prime salariale ou une pénalité. Alors que les analyses transversales des salaires horaires suggèrent fréquemment des pénalités pour les travailleurs hautement qualifiés, d'autres méthodologies laissent entrevoir des primes institutionnelles, créant une compréhension fragmentée de la rémunération sectorielle.

Résoudre cette énigme nécessite de déplacer la focale analytique des salaires horaires statiques et transversaux vers les revenus dynamiques sur l'ensemble du cycle de vie. En s'appuyant sur de riches données de panel administratives françaises couvrant la période de 2012 à 2019, le premier chapitre établit un ensemble de faits stylisés, inédits mais intuitifs, concernant la dynamique des salaires, la mobilité sectorielle et les disparités de genre. Le résultat central révèle que des écarts de salaires horaires modestes dissimulent systématiquement des disparités substantielles en matière de revenus sur le cycle de vie et de stabilité de l'emploi. L'évaluation de la rémunération sur un horizon de carrière démontre que les primes de revenus sur le cycle de vie diffèrent radicalement en fonction des caractéristiques à la fois observables et inobservées.

De manière cruciale, l'analyse identifie des gagnants et des perdants distincts au sein du cadre sectoriel public-privé. Les femmes et les travailleurs peu diplômés bénéficient d'un avantage significatif en termes de revenus sur le cycle de vie dans le secteur public. Cette prime n'est pas seulement tirée par le salaire de base, mais par des caractéristiques institutionnelles qui favorisent une meilleure rétention, un travail à temps partiel mieux rémunéré et un volume d'heures annuelles plus équitable par rapport au secteur privé. Dans le secteur privé, les écarts de genre demeurent nettement plus marqués, en particulier pour les individus possédant des niveaux d'éducation plus élevés. À l'inverse, les hommes hautement qualifiés subissent une pénalité prononcée sur leurs revenus de cycle de vie dans l'emploi public, conséquence directe de structures salariales rigides et centralisées qui compriment les queues supérieures de la distribution des salaires et limitent les rendements du capital humain avancé.

En modélisant dynamiquement les transitions sectorielles, les entrées et sorties de l'emploi, et l'hétérogénéité des revenus à l'aide d'une approche algorithmique Espérance-Maximisation (EM) inspirée par Postel-Vinay and Turon (2007), le premier chapitre démontre que les primes comme les pénalités dépendent systématiquement du genre, de l'éducation et de l'expérience accumulée sur le marché du travail. L'analyse révèle qu'une part importante de l'hétérogénéité inobservée persiste dans la dynamique des salaires, même après avoir contrôlé pour un riche ensemble de variables observables. Ce chapitre unifie des récits jusqu'alors conflictuels en fournissant un compte rendu exhaustif des différences sectorielles, mettant en lumière le double rôle des secteurs publics : atténuer les inégalités de genre par la stabilité de l'emploi tout en imposant simultanément des plafonds à la progression salariale pour certains groupes.

La présence d'une profonde hétérogénéité inobservée dictant les différentiels de revenus sur le cycle de vie invite naturellement à une question plus large. Si des caractéristiques inobservées déterminent des résultats si divergents d'un secteur à l'autre, quelles sont les dimensions structurelles plus profondes — la qualité des travailleurs, les politiques salariales des entreprises et la nature même des emplois — le long desquelles s'opère le tri sur le marché du travail dans son ensemble ?

Tri multidimensionnel et valeur sur le cycle de vie

Répondre à cette question nécessite un cadre conceptuel capable de saisir la pleine dimensionnalité du marché du travail. Depuis la contribution fondatrice de Abowd et al. (1999), la littérature étudiant

la manière dont les salaires individuels varient entre les travailleurs et les entreprises s'est concentrée sur des modèles additifs à effets fixes qui décomposent la variance des salaires en composantes liées au travailleur et à l'entreprise. Cette approche, prolongée par des méthodes de correction de biais (Andrews et al., 2008; Kline et al., 2020), s'est révélée extrêmement féconde pour comprendre la structure bipartite de la détermination des salaires. Cependant, par construction, les modèles additifs travailleurs-entreprises traitent le marché du travail comme un problème d'appariement bilatéral. En pratique, un emploi n'est pas simplement une paire travailleur-entreprise : c'est un triplet comprenant un travailleur, une entreprise et une profession. Les travailleurs changent de profession — tant au sein qu'entre les entreprises — et ces transitions véhiculent une information riche sur les trajectoires de carrière et la valeur sur le cycle de vie. Dans les cadres bipartites, cette dimension professionnelle est nécessairement absorbée dans les effets travailleurs, les effets entreprises ou le résidu, plutôt que d'être isolée comme une source distincte de tri.

Le deuxième chapitre, *Mobilité professionnelle et valeurs de carrière*, co-écrit avec Jean-Marc Robin, comble cette lacune en estimant conjointement un tri tripartite entre les travailleurs, les entreprises et les professions. Pour y parvenir sans imposer les hypothèses comportementales caractéristiques des modèles de recherche structurels, le chapitre implémente un algorithme de Classification Espérance-Maximisation (CEM) flexible et de forme réduite (Lentz et al., 2023) sur des données administratives françaises. En définissant un emploi comme un triplet composé d'un type de travailleur, d'un type d'entreprise et d'une profession, l'analyse identifie trois types latents de travailleurs discrets et douze types latents d'entreprises. Cette discrétisation contourne intrinsèquement le problème des paramètres incidents associé à l'estimation de millions d'effets fixes, tout en autorisant les interactions non linéaires entre la qualité du travailleur, la politique salariale de l'entreprise et le choix professionnel, interactions que les modèles additifs excluent.

S'appuyant sur l'approche de mélange fini de Bonhomme et al. (2019) et le modèle structurel de mobilité de Lentz et al. (2023), l'algorithme CEM classifie les entreprises en utilisant l'information complète issue à la fois des distributions salariales et des profils de mobilité, plutôt que des seuls salaires. L'hétérogénéité professionnelle est introduite via huit catégories de professions basées sur les codes PCS (Professions et Catégories Socioprofessionnelles), croisées avec l'âge et le genre, permettant au modèle de capturer la riche variation de la manière dont différents types de travailleurs se répartissent dans différents types d'emplois à différentes étapes de leur carrière. De plus, le chapitre introduit les décompositions de la valeur de Shapley dans la littérature sur la variance des salaires. Cette innovation méthodologique permet une attribution exacte et indépendante de l'ordre de la contribution de chaque dimension du tri, offrant une solution de principe au problème de corrélation qui se pose chaque fois que les types de travailleurs, les types d'entreprises et les professions sont déterminés conjointement.

Trois résultats majeurs émergent de ce cadre multidimensionnel. Premièrement, l'analyse identifie la profession comme le principal facteur contributif à la fois à la variation des salaires et à la variation de la valeur sur le cycle de vie — une dimension largement absorbée ou occultée dans les modèles traditionnels travailleurs-entreprises. L'interaction complète des professions, du genre, et des types

latents de travailleurs et d'entreprises explique près de 64 % de la variance des logarithmes des salaires ajustée selon l'âge. Au sein de cette variance expliquée, la profession représente la part la plus importante (26,6 %), suivie de près par les types d'entreprises (23,0 %), tandis que les types latents de travailleurs (7,8 %) et le genre (6,3 %) fournissent des contributions marginales plus faibles.

Deuxièmement, le chapitre étend l'analyse des salaires actuels à la valeur sur le cycle de vie en calculant la valeur actuelle nette (VAN) de chaque emploi à l'aide d'une induction rétrospective exacte sur la chaîne de Markov estimée des transitions des travailleurs. Les résultats révèlent une divergence frappante : alors que le triplet travailleur-entreprise-profession explique 71 % de l'entropie de la VAN, il n'explique que 17 % de l'entropie des salaires ajustés selon l'âge. Cela indique que la VAN intègre efficacement une structure de transition prospective fortement dictée par les types latents, capturant une dimension de la progression de carrière invisible dans la variance transversale des salaires.

Enfin, le chapitre évalue les moteurs de la mobilité des travailleurs. L'analyse démontre que les choix de destination des travailleurs réagissent fortement à la composante capitalisée du salaire actuel de la valeur d'un emploi, mais ne montrent presque aucune réaction à la composante de valeur d'option. Les transitions sont massivement orientées vers l'augmentation des salaires, fournissant des preuves convaincantes que la valeur de construction de carrière ou de « tremplin » d'un changement d'employeur est nettement moins importante pour la mobilité des travailleurs que les gains salariaux immédiats. Ces résultats soulignent l'immense contenu informationnel de l'hétérogénéité professionnelle et redéfinissent la compréhension des raisons pour lesquelles les travailleurs se répartissent à travers le marché du travail.

La fiabilité de la décomposition additive de la variance

Ayant établi qu'une approche multidimensionnelle à types discrets révèle que la profession est le contributeur dominant à la variance des salaires et met à jour des interactions non linéaires que les modèles additifs ne peuvent capturer, une question complémentaire naturelle se pose : sous quelles conditions les décompositions additives de la variance — l'outil de base de la littérature sur les données appariées employeurs-employés — recouvrent-elles de manière fiable leurs paramètres cibles ? Le cadre de Abowd et al. (1999) fournit une décomposition élégante et parcimonieuse de la variance des log-salaires en effets travailleurs, effets entreprises et une covariance de tri. Pour la question bipartite à laquelle il était censé répondre — quelle part de la variance des salaires est attribuable aux travailleurs par rapport aux entreprises — il demeure l'outil canonique. Cependant, à mesure que la recherche s'appuyant sur des bases de données administratives s'est développée, des preuves croissantes suggèrent que les parts de variance produites par les estimateurs AKM sont sensibles à la structure de l'échantillon, particulièrement en ce qui concerne le degré de mobilité des travailleurs entre les entreprises (Andrews et al., 2008; Kline et al., 2020).

Le troisième chapitre, *Signaux, bruit et tri*, s'attaque directement à cette question. Ce chapitre étudie les conditions structurelles sous lesquelles la décomposition additive de la variance est fiable. Alors

que la littérature existante identifie la mobilité limitée comme une source de biais de paramètres incidents, le chapitre reconceptualise ce problème à travers le prisme de la théorie spectrale des graphes. L'analyse démontre que le biais inhérent aux estimateurs *plug-in* est fondamentalement déterminé par les propriétés spectrales du laplacien de graphe, qui caractérise le réseau entreprise-entreprise de la mobilité des travailleurs.

Une innovation critique de ce chapitre consiste à identifier le degré moyen des nœuds ($\bar{d}$) comme un diagnostic principal et facile à calculer de la fragilité de l'estimation. L'analyse révèle que les distorsions dans la covariance de tri et l'inflation artificielle des parts de variance des travailleurs ne sont pas de simples artefacts liés à des tailles d'échantillon limitées. Elles découlent plutôt de l'interaction entre le bruit idiosyncratique et la structure de pondération du graphe biparti. Lorsque le degré moyen des nœuds est faible — un symptôme mathématique d'une taille moyenne d'entreprise réduite et de transitions clairsemées de travailleurs — le spectre du laplacien s'effondre vers zéro. Dans ces conditions, le bruit idiosyncratique détruit le signal de tri positif, distordant lourdement toutes les composantes de la variance de type *plug-in*.

À travers des simulations de Monte-Carlo approfondies, le chapitre évalue la performance des méthodes de correction de biais les plus récentes, y compris les estimateurs par validation croisée (*leave-out*) et les techniques de séparation d'échantillon (*split-sample*). Les résultats indiquent que, bien que ces corrections réussissent à réduire le biais lié à la mobilité limitée, elles restent très sensibles à l'amplification du bruit lorsque la densité spectrale du réseau est faible. Ce phénomène atténue systématiquement les preuves d'un tri assortatif positif, jetant le doute sur les résultats issus de réseaux clairsemés. L'estimateur sur échantillon divisé émerge comme l'alternative la plus robuste, car il brise de manière unique le canal d'amplification du bruit spectral tout en éliminant le biais grâce à une récupération basée sur la soustraction.

En fin de compte, le troisième chapitre établit que la fiabilité de la décomposition additive de la variance est fondamentalement contrainte par la connectivité structurelle du réseau de mobilité sous-jacent. En identifiant $\bar{d}$ comme une statistique exhaustive de la gravité de l'effondrement spectral, cette recherche fournit un cadre de diagnostic qui permet aux chercheurs utilisant des modèles de type AKM d'évaluer si leur réseau supporte une inférence crédible. La contribution s'adresse donc à la littérature AKM elle-même : renforcer les conditions sous lesquelles sa décomposition peut être considérée comme fiable.

Progression et contributions

Lus de manière séquentielle, les chapitres de cette thèse retracent une progression intellectuelle. Le parcours débute par une question empirique : en quoi les revenus sur le cycle de vie diffèrent-ils entre les secteurs public et privé ? La découverte que ces différences sont hautement hétérogènes et dépendent de la qualité inobservée des travailleurs motive une investigation plus approfondie des dimensions structurelles de la variance des salaires.

Cela mène directement à l'analyse multidimensionnelle du deuxième chapitre. En estimant un modèle CEM flexible à types latents discrets, le chapitre cartographie le tri tripartite des travailleurs, des entreprises et des professions. La réintroduction de la profession comme pilier central de la détermination des salaires — et l'élargissement de l'horizon de la variance transversale à la valeur actuelle nette sur le cycle de vie — révèle que le marché du travail est plus riche que ne le suggère un monde biparti travailleur-entreprise. La mobilité professionnelle, les interactions non linéaires entre la qualité des travailleurs et des entreprises, et la distinction entre salaires actuels et valeurs de carrière prospectives émergent toutes comme des caractéristiques de premier ordre des inégalités.

Le troisième chapitre se tourne ensuite vers les fondements économétriques de l'outil de référence actuel. Comprendre les propriétés spectrales du laplacien de graphe explique pourquoi la littérature a peiné à atteindre un consensus sur le degré d'appariement assortatif : dans les réseaux clairsemés, la structure même que les chercheurs tentent d'analyser peut fragiliser les estimateurs utilisés pour ce faire. En fournissant un cadre de diagnostic basé sur la connectivité du réseau, ce chapitre renforce la boîte à outils disponible pour la décomposition additive de la variance tout en clarifiant ses limites.

Bien que chaque chapitre constitue une contribution indépendante à la littérature, ils forment ensemble un récit exhaustif. Ils démontrent que la compréhension des inégalités salariales exige de prêter attention aux institutions et à la réalité multidimensionnelle des choix des travailleurs et des politiques des entreprises, ainsi que d'évaluer rigoureusement les outils économétriques que nous utilisons pour les mesurer. À travers cette progression, cette thèse contribue à l'effort continu pour cartographier, mesurer et, *in fine*, comprendre la distribution des revenus.

Bibliography

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

Martyn J Andrews, Leonard Gill, Thorsten Schank, and Richard Upward. High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A*, 171(3):673–697, 2008.

Gary S. Becker. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. Number beck-5 in NBER Books. National Bureau of Economic Research, Inc, none edition, August 1964. doi: None. URL https://ideas.repec.org/b/nbr/nberbk/beck-5.html.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer-employee data. *Econometrica*, 83(3):699–739, 2019.

Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898, September 2020.

Rasmus Lentz, Suphanit Piyapromdee, and Jean-Marc Robin. The anatomy of sorting—evidence from danish data. *Econometrica*, 91(6):2409–2455, 2023. doi: https://doi.org/10.3982/ECTA16425.

Jacob A. Mincer. The human capital earnings function. In *Schooling, Experience, and Earnings*, NBER Chapters, pages 83–96. National Bureau of Economic Research, Inc, August 1974. URL https://ideas.repec.org/h/nbr/nberch/1767.html.

Fabien Postel-Vinay and Hélène Turon. The public pay gap in britain: Small differences that (don't?) matter. *Economic Journal*, 117(523):F422–F442, 2007. doi: 10.1111/j.1468-0297.2007.02067.x.