The Worker and Firm Origins of Life-Cycle Wage Inequality*

Xincheng Qiu† Peking University

Jesse Wedewer‡ Duke University

Runling Wu§ Duke University

May 22, 2026

Abstract

Wage inequality widens substantially over the life cycle. We decompose this widening into worker and firm heterogeneity in returns to experience and dynamic sorting. We extend the AKM model to allow wage growth to differ across workers and firms, and develop a two-way clustering algorithm. Using matched employer-employee data from Italy, we document substantial heterogeneity in both worker and firm growth components. Worker growth accounts for two-thirds of the rise in life-cycle wage inequality, while firm growth accounts for the remaining third. Workers with more accumulated skills increasingly sort into high-paying firms over the life cycle.

Keywords: Two-sided heterogeneity, returns to experience, life-cycle wage inequality, grouped fixed effects

JEL Codes: J24, J31, C38

*We thank Peter Arcidiacono, Patrick Bayer, Vince Chen, Michael Dinerstein, Rafael Dix-Carneiro, Victoria Gregory, Daniel Hannwinckel, Gregor Jarosch, Matthias Kehrig, Philipp Kircher, Rasmus Lentz, Laura Pilossoph, Michael Pollmann, and Uta Schönberg for helpful comments. We also thank audiences at CityU, Duke, Cleveland Fed, HKU, Minneapolis Fed, SED Copenhagen, and SOLE Denver.

†Email: xincheng.qiu@gsm.pku.edu.cn.

‡Email: jesse.wedewer@duke.edu.

§Email: runling.wu@duke.edu.

1 Introduction

Wage inequality widens substantially over the life cycle (Heathcote et al., 2010; Huggett et al., 2011). This widening reflects large and persistent differences in wage growth across workers. Some workers experience rapid and sustained wage growth over their careers, while others follow flat or even declining wage profiles. In the United States, for instance, average earnings between ages 25 and 55 rise almost fivefold for workers at the top 5 percent of the lifetime-earnings distribution, but decline for those in the bottom 20 percent (Guvenen et al., 2021).[1]

Several mechanisms could account for this divergence. Workers may differ in their ability to accumulate human capital (Rubinstein and Weiss, 2006; Huggett et al., 2011; Bagger et al., 2014; Ozkan et al., 2023); firms may differ in the training, learning, and promotion opportunities they provide (Gregory, 2026; Arellano-Bover and Saltiel, 2026; Borovičková and Macaluso, 2024; Ma et al., 2024); and workers may climb the job ladder at different speeds through on-the-job search (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002). Separating these mechanisms is important because each locates the origins of life-cycle inequality to a different place—workers, firms, or matches—and therefore points to a different margin for potential intervention.

This paper develops a unified framework for measuring the contribution of each mechanism to life-cycle wage inequality. We extend the Abowd, Kramarz and Margolis (1999) wage equation by allowing returns to experience to vary on both sides of the labor market. The worker growth component captures heterogeneity in how individual workers convert experience into wage growth. The firm growth component captures heterogeneity in how experience accumulated at particular firms generates wage growth. These components separate the role of individual learning ability from the role of firms as environments for wage growth.

Estimating fully unrestricted worker- and firm-specific growth effects would confront a severe dimensionality problem. We theorefore approximate the continuous distributions of worker and firm returns to experience with a finite number of latent types, in the spirit of Bonhomme and Manresa (2015), and extend the group fixed-effect approach to two-sided heterogeneity. The estimator jointly assigns workers and firms to latent growth types and estimates the corresponding type-specific returns. We implement it with a two-way clustering algorithm that alternates between OLS updates of the type-specific parameters, worker reassignment, and firm reassignment, initialized from a first-difference auxiliary problem. Monte Carlo simulations confirm that the procedure recovers both the latent types and the parameters of the wage equation.

[1]The pattern is not unique to the United States. In Austrian administrative data, workers in the first decile of the lifetime earnings distribution experience virtually no wage growth, whereas wages of workers in the ninth decile more than double between ages 25 and 55 (Borovičková and Macaluso, 2024). We find a similar pattern in Veneto, Italy: workers at the top of the wage distribution more than triple their wages over fifteen years, while those at the bottom see no real growth (Figure 1).

Applying our framework to the Veneto Worker History data, we document three main findings. First, worker heterogeneity in returns to experience is the dominant force behind the rise in life-cycle wage inequality, although firm heterogeneity also plays a substantial role. Between ages 25 and 42, the wage gap between workers in the top and bottom terciles of wage distribution rises by about 0.5 log points, of which 65% is accounted for by heterogeneity in worker returns to experience, 20% by heterogeneity in firm returns to experience, and 15% by differential rates of movement up the job ladder. The variance decomposition yields a similar result that the worker growth component accounts for roughly two-thirds of the increase in wage dispersion over the same ages, while the firm growth component accounts for the remaining third.

Second, worker-firm sorting affects life cycle wage inequality in two opposing ways. Sorting on wage levels magnifies inequality. The covariance between the dynamic worker effect, defined as the sum of the worker fixed effect, accumulated worker growth, and accumulated firm-side growth, and firm fixed effects rises from near zero at age 25 to roughly 0.03 by age 42. This suggests that workers with high accumulated productivity increasingly match with firms that pay high wage premia. Sorting on wage growth, however, has the opposite effect. High-growth workers tend to match with low-growth firms, dampening the divergence in wage trajectories. Without this negative sorting across growth types, life-cycle wage inequality would be substantially larger.

Third, we explore the economic content of the latent growth types. On the worker side, managers have both the highest worker fixed effects and the highest worker growth coefficients, and their pay gaps relative to white- and blue-collar workers widen substantially with experience. On the firm side, large firms pay higher wage premia, but small firms generate faster wage growth. Firm fixed effects rise sharply with firm size, while firm growth coefficients decline. Across industries, high pay premia and steep wage growth need not coincide. Credit and insurance industry has the largest firm wage premia but offers only modest returns to experience; the trade sector has the highest firm growth but pays less.

Related literature. We contribute to three literatures. First, we contribute to the AKM literature on the role of workers and firms in wage determination (Abowd et al., 1999; Card et al., 2013, 2016; Song et al., 2019). This literature has shown that worker and firm fixed effects explain a large share of wage dispersion, but it typically treats returns to experience as homogeneous. We extend the AKM framework to allow returns to experience to vary on both sides of the labor market. The most closely related paper is Sørensen and Vejlin (2011), who estimate worker and firm fixed effects in a wage-growth regression on Danish data. We differ from their approach in two ways. First, we replace individual-level fixed effects, which are noisy and subject to limited-mobility bias, with group fixed effects that pool information across similar workers and firms. Second, by embedding the growth components in a levels specification rather than a wage-growth regression, we jointly decompose both the level and the growth of life-cycle wage inequality and characterize the dynamic sorting between worker and firm types, analyses a wage-growth regression alone does not support.

Methodologically, our estimator builds on the grouped fixed-effect approach (Bonhomme and Manresa, 2015), which approximates high-dimensional heterogeneity with a finite number of latent types. We extend this approach to two-sided heterogeneity. Related clustering approaches have been used in matched employer-employee settings by Dauth et al. (2022), Bradley et al. (2023), and Gregory et al. (2024). The closest paper is Arellano-Bover and Saltiel (2026), who group firms using k-means but restrict worker heterogeneity to wage levels only but not wage growth. By allowing worker and firm growth heterogeneity to be estimated jointly, our framework avoids mechanically attributing the wage growth of high-growth workers to their employers. This distinction matters quantitatively: it changes the estimated dispersion of firm growth effects and yields a different decomposition of life-cycle wage inequality.

Second, we contribute to the literature on heterogeneous returns to experience. A large body of work allows experience profiles to differ across observable groups, such as education (Gathmann and Schönberg, 2010; Bagger et al., 2014; Deming, 2023; Helm et al., 2026). We instead allow returns to experience to vary flexibly across latent types. This is motivated by evidence from the earnings dynamics literature showing that workers differ systematically in their earnings growth over the life cycle (Guvenen, 2009; Magnac et al., 2018). We incorporate such heterogeneity into a Mincerian wage framework that allows returns to experience to vary across both workers and firms.

Third, we contribute to the literature on the evolution of wage inequality over the life-cycle of workers (Rubinstein and Weiss, 2006; Huggett et al., 2011; Bagger et al., 2014; Guvenen et al., 2021; Ozkan et al., 2023). Our contribution is to decompose the sources of wage dispersion over the life-cycle of workers into worker, firm, and sorting components within a unified empirical framework. We find that heterogeneity in worker returns to experience accounts for roughly two-thirds of the increase in life-cycle wage dispersion, while heterogeneity in firm returns to experience accounts for the remaining third. These findings complement recent work that uses structural labor-market models to quantify the roles of worker and firm heterogeneity in wage growth (Borovičková and Macaluso, 2024; Gregory, 2026).

Roadmap. Section 2 presents the empirical framework. Section 3 develops the two-way clustering algorithm. Section 4 presents the empirical results on the decomposition of life-cycle wage inequality and dynamic sorting. Section 5 concludes.

2 Framework

This section develops the framework for estimating heterogeneous returns to experience across workers and firms. We first present the statistical model extending Abowd, Kramarz and Margolis (1999) (hereafter AKM) to allow two-sided heterogeneity in wage growth, and then the empirical specification using a group fixed-effects approach that discretizes returns to experience into worker and firm types.

2.1 Empirical Framework

AKM consider a statistical model in which log wages are an additively separable function of worker and firm fixed effects. For a worker i employed at firm $j(i,t)$ at time t with characteristics $X_{i,t}$, log wages satisfy:

$$\ln y_{i,t} = \underbrace{\alpha_i}_{\text{Worker effect}} + \underbrace{\psi_{j(i,t)}}_{\text{Firm effect}} + \underbrace{\xi X_{i,t}}_{\text{Observable factors}} + \underbrace{r_{i,t}}_{\text{Residual}} . \tag{1}$$

The worker effect α_i is a time-invariant component capturing permanent unobserved attributes rewarded uniformly across firms. The firm effect $\psi_{j(i,t)}$ is a firm-specific premium paid to all workers employed at firm j. Observed wage determinants are captured by $\xi X_{i,t}$, and $r_{i,t}$ is the residual.

As workers accumulate experience, their wages increase. We model this evolution through returns to experience accumulated over the worker's employment history:

$$\ln y_{i,t} = \alpha_i + \psi_{j(i,t)} + \xi X_{i,t} + \underbrace{\beta_i \,\text{Exp}_{i,t}}_{\text{worker growth component}} + \underbrace{\sum_{j \in \mathcal{J}} \phi_j \text{Exp}^j_{i,t}}_{\text{firm growth component}} + r_{i,t}. \tag{2}$$

The *worker growth component* is the product of $\text{Exp}_{i,t}$, the cumulative time worker i has spent in the labor market by time t, and β_i, the worker-specific rate at which experience translates into wage growth. The *firm growth component* sums, over all firms j in worker i's employment history, the product of $\text{Exp}^j_{i,t}$, the experience worker i has accumulated at firm j by time t, and ϕ_j, the firm-specific rate at which experience translates into wage growth.

A natural interpretation of these returns to experience is human capital accumulation in the spirit of Ben-Porath (1967). On the worker side, individuals differ in their learning ability, i.e., the speed at which they accumulate human capital (Huggett et al., 2011), and in their willingness to invest in training (Caliendo et al., 2022, 2023). These differences load onto β_i. On the firm side, firms differ in the training opportunities they provide (Acemoglu and Pischke, 1998; Ma et al., 2024; Gregory, 2026; Arellano-Bover and Saltiel, 2026) and in the extent to which workers learn from coworkers and managers (Jarosch et al., 2021; Herkenhoff et al., 2024; Hong, 2022). These differences load onto ϕ_j.

The statistical model is silent on the underlying sources of wage growth, however, and does rule out alternative interpretations. For example, in theories of employer learning, wage growth reflects the gradual revelation of workers' latent productivity. Although the tradition in this literature does not directly model heterogeneity in learning rates across workers or firms (Farber and Gibbons, 1996; Altonji and Pierret, 2001; Lange, 2007), our framework accommodates such heterogeneity: workers whose productivity is revealed more quickly load onto β_i, and firms within which learning is faster load onto ϕ_j. Alternatively, wage growth may also reflect a sequence of renegotiation events over the career—at hire, in response to outside offers, and at periodic reviews—and both sides shape how these events translate into wage changes. Workers differ in their ability or willingness to bargain, to elicit outside offers, and to ask for raises (Biasi and Sarsons, 2021; Roussille, 2024; Card et al., 2016); firms differ in their responsiveness to outside offers, willingness to renegotiate, and rent-sharing policies (Caldwell et al., 2026).

For the parameters of equation (2) to be identified by OLS, the residual must be mean-independent of experience and observables given worker and firm identities. That is, the error term is assumed to satisfy the strict exogeneity assumption:

$$\mathbb{E}[r_{i,t}|i,t,j(i,t),V_{i,t}=\{\text{Exp}_{i,t},\text{Exp}^{j}_{i,t},X_{i,t}\}]=0.$$

This rules out unobserved shocks to wages that are systematically correlated with workers' selection into employment histories.

2.2 Two-Way Group Fixed Effect

While the two-way individual fixed effects specification above is conceptually clean, it has several practical limitations. First, the estimated individual fixed effects can be noisy because of the high dimensionality, especially for workers and firms observed for relatively short periods. Each fixed effect is estimated from the corresponding worker's or firm's own time-series variation, with exploiting information across similar workers or firms.

Second, a central objective of this paper is to decompose life-cycle wage inequality into worker and firm contributions. However, the standard bias correction for such decompositions, developed by Kline, Saggio and Sølvsten (2020) (hereafter KSS), is not directly applicable in our setting. The KSS correction requires collapsing the data to the match level in order to accommodate serial correlation within worker-firm matches. In our application, this transformation generates a system in which the number of parameters exceeds the number of match-level observations. The resulting design matrix is rank deficient and the least-squares problem admits infinitely many solutions.

Third, one could instead estimate individual fixed effects in wage growth using first differences for job stayers. This approach is computationally tractable, but it is identified only from workers who remain with the same employer in consecutive years. It therefore changes the effective estimation sample by excluding short matches and highly mobile workers, both of which are central to life-cycle wage dynamics.

To address these challenges, we adopt a group fixed-effects (GFE) approach that approximates the underlying continuous distribution of (β_i, ϕ_j) by a discrete distribution, in the spirit of Bonhomme and Manresa (2015), Bonhomme, Lamadon and Manresa (2022), and Bonhomme and Denis (2025). By restricting (β_i, ϕ_j) to take a finite number of values, the GFE approach pools information across workers and firms with similar experience effects. This pooling substantially reduces estimation noise while preserving meaningful heterogeneity in returns to experience on both sides of the labor market.

Formally, suppose workers are partitioned into K_w groups indexed by $\ell \in \{1, ..., K_w\}$, with $\beta_i = \beta_\ell$ for every worker i in group ℓ, and firms are partitioned into K_f groups indexed by $k \in \{1, ..., K_f\}$, with $\phi_j = \phi_k$ for every worker j in group k. Let $\rho(\cdot) : \{1, \ldots I\} \to \{1, \ldots, K_w\}$ and $\kappa(\cdot) : \{1, \ldots, J\} \to \{1, \ldots, K_f\}$ denote the worker and firm assignment functions. The two-way group fixed-effects model is then

$$\ln y_{i,t} = \alpha_i + \psi_{j(i,t)} + \underbrace{\beta_{\rho(i)} \operatorname{Exp}_{i,t}}_{\text{worker growth component}} + \underbrace{\sum_j \phi_{\kappa(j)} \operatorname{Exp}^j_{i,t}}_{\text{firm growth component}} + \xi X_{i,t} + r_{i,t}. \tag{3}$$

We refer to $\beta_{\rho(i)}$ and $\phi_{\kappa(j)}$ as group-specific worker and firm returns to experience. The former captures how an additional year of experience translates into wage growth for workers in a given worker group, and the latter captures the corresponding contribution of employment at firms in a given firm group. The worker growth component and firm growth components therefore summarize the contributions of worker-side and firm-side heterogeneity in life-cycle wage growth.

As in Abowd et al. (1999), the worker and firm effects are identified only within a connected set of workers and firms linked by mobility. We restrict attention to the largest connected set. The group-specific effects require an analogous notion of connectedness, but at the group level. Because group memberships are coarser than individual identities, the effective mobility at the group level is substantially higher than at the individual level. This mitigates the limited mobility bias documented by Andrews et al. (2008, 2012) for individual FE estimates. With many workers per group, even modest worker mobility generates dense group-to-group transitions, so the sparseness of worker-firm links that drives the bias at the individual level largely disappears at the group level.

3 Two-Way Clustering

3.1 Iterative Search Algorithm

This section develops the two-way clustering algorithm used to estimate equation (3). The algorithm jointly assigns workers and firms to discrete growth types and estimates the corresponding group-specific returns to experience. Formally, it minimizes minimizes the sum of squared residuals over both the parameters and the assignment functions:

$$\min_{\substack{\rho(i):\,\mathcal{I}\to\{1,\ldots,K_w\},\\ \kappa(j):\,\mathcal{J}\to\{1,\ldots,K_f\},\\ \beta_1,\ldots,\beta_{K_w},\,\phi_1,\ldots,\phi_{K_f}\\ \{\alpha_i\}_{i\in\mathcal{I}},\,\{\psi_{j(i,t)}\}_{j\in\mathcal{J},\xi}}} \sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}_i}\left(\ln y_{it}-\alpha_i-\psi_{j(i,t)}-\xi' X_{it}-\beta_{\rho(i)}\mathrm{Exp}_{it}-\sum_{j\in\mathcal{J}_i}\phi_{\kappa(j)}\mathrm{Exp}^j_{it}\right)^2, \qquad (4)$$

For notational convenience, let $\theta \equiv \{\{\beta_\ell\},\{\phi_k\},\{\alpha_i\},\{\psi_j\},\xi\}$ collects the of parameters.

While the objective is smooth in the parameter vector θ conditional on the assignments, it is discontinuous in the assignment functions themselves. The resulting optimization problem is therefore non-convex and combinatorial. The set of feasible joint assignments has cardinality $K_w^N \times K_f^J$, which grows exponentially with the number of workers and firms, making exhaustive search infeasible even for moderate values of N and J.

To address this computational challenge, we propose an iterative algorithm in the spirit of Mann (2024). Starting from an initial set of group assignments, the algorithm alternates among three conditional minimization steps. First, given the current worker and firm assignments, we update model parameters θ by ordinary least squares. Second, holding firm assignments and θ fixed, we reassign each worker to the group that yields the lowest sum of squared residuals. Third, holding worker assignments and θ fixed, we update firm assignments analogously to minimize the sum of squared residuals. We repeat these steps until converge.

Main Algorithm: Coordinate Descent. Starting from initial assignments $(\rho^{(0)},\kappa^{(0)})$, we iterate the following three steps until convergence:

1. *Parameter update.* Given current assignments $(\rho^{(s)},\kappa^{(s)})$, update parameters by OLS:

$$\theta^{(s+1)} = \operatorname*{argmin}_{\theta}\sum_{i,t}\left(\ln y_{it}-\alpha_i-\psi_{j(i,t)}-\beta_{\rho^{(s)}(i)}\mathrm{Exp}_{it}-\sum_{j}\phi_{\kappa^{(s)}(j)}\mathrm{Exp}^j_{it}-\xi' X_{it}\right)^2$$

We solve this by preconditioned conjugate gradient on the sparse two-way design matrix.

2. *Worker-type update.* Given current firm assignments $\kappa^{(s)}$ and updated parameters $\theta^{(s+1)}$, reassign each worker i:

$$\rho^{(s+1)}(i) = \underset{\ell \in \{1,\dots,K_w\}}{\operatorname{argmin}} \sum_{t \in \mathcal{T}_i} \left(w_{it}^{(s+1)} - \beta_\ell^{(s+1)} \mathrm{Exp}_{it} - \sum_j \phi_{\kappa^{(s)}(j)}^{(s+1)} \mathrm{Exp}_{it}^j \right)^2$$

where the partial residual is defined as $w_{it}^{(s+1)} \equiv \ln y_{it} - \hat{\alpha}_i^{(s+1)} - \hat{\psi}_{j(i,t)}^{(s+1)} - \hat{\xi}^{(s+1)\prime} X_{it}$. Because each worker's optimal assignment is independent of every other worker's given (θ, κ), all workers are reassigned in parallel.

3. *Firm-type update.* Given updated worker assignments $\rho^{(s+1)}$ and parameters $\theta^{(s+1)}$, reassign each firm j:

$$\kappa^{(s+1)}(j) = \underset{k \in \{1,\dots,K_f\}}{\operatorname{argmin}} \sum_{i: j \in \mathcal{J}_i} \sum_{t \in \mathcal{T}_{ij}} \left(w_{it}^{(s+1)} - \beta_{\rho^{(s+1)}(i)}^{(s+1)} \mathrm{Exp}_{it} - \phi_k^{(s+1)} \mathrm{Exp}_{it}^j - \sum_{j' \neq j} \phi_{\kappa^{(s)}(j')}^{(s+1)} \mathrm{Exp}_{it}^{j'} \right)^2$$

Unlike workers, firms are coupled through the firm-group experience regressor: reassigning firm j changes the cumulative experience for every worker who has accumulated tenure at j. We therefore reassign firms sequentially, starting from the largest, updating residuals after each move (Lentz et al., 2023).

The iteration terminates when assignments stabilize and the reduction in the objective falls below a preset threshold. Formal pseudocode for each step is provided in Appendix B.

3.2 Initialization

The estimation problem in equation (4) is non-convex in the assignment functions, so the algorithm's convergence point can depend critically on its starting values. To avoid poor local minima, we precede the main algorithm with two initialization steps that generate a high-quality warm start from a tractable auxiliary problem: the first-differenced wage equation.

For job stayers ($j(i,t) = j(i,t-1)$), first-differencing equation (2) after partialling out observables $X_{i,t}$ eliminates α_i and $\psi_{j(i,t)}$. Because each stayer gains one year of experience in the labor market ($\Delta \, \mathrm{Exp}_{i,t} = 1$) and one year of experience at firm j ($\Delta \, \mathrm{Exp}_{i,t}^j = 1$), the first difference of log wages reduces to

$$\Delta \ln y_{i,t} \; = \; \beta_i \; + \; \phi_j \; + \; \Delta r_{i,t}. \tag{5}$$

In practice, we first residualize wages on observables $X_{i,t}$ before taking first differences; for notational simplicity we continue to write $y_{i,t}$ throughout the main text and defer details to Appendix B. Stayers' wage growth is therefore a direct, observable signal of the latent growth types (β_i, ϕ_j). Specifically, we initialize the worker and firm groupings in two steps. We first apply k-means clustering separately for workers and firms, using two moments of stayers' wage growth: the empirical CDF and the mean. We next refine the resulting groupings by applying coordinate descent to the same first-differenced equation, searching for worker and firm assignments that further reduce the sum of squared residuals. This initialization helps place the algorithm in a favorable basin of attraction for the more difficult levels problem and substantially reduces the need for random restarts (Bonhomme and Manresa, 2015). We summarize the initialization procedure below.

Initialization Step 1: k-means on Stayers' Wage Growth Distributions. Equation (5) motivates using the distribution of stayers' wage growth as a basis for initial classification. For each worker i, we compute the cumulative distribution function and mean of the worker's residualized wage growth across stayer observations; for each firm j, we compute the cumulative distribution function and mean across all stayer observations at that firm. We then apply k-means separately to these worker-level and firm-level distributions, producing initial worker groups $\hat{\rho}_0(\cdot)$ and firm groups $\hat{\kappa}_0(\cdot)$. Workers or firms without stayer observations (e.g., always-movers) are assigned the modal group.

Initialization Step 2: First-Difference Coordinate Descent. We refine $(\hat{\rho}_0, \hat{\kappa}_0)$ using the same coordinate-descent steps as the main algorithm, alternating OLS parameter updates with worker and firm reassignments, but applied to the group-level version of equation (5). The iteration converges to a refined pair $(\hat{\rho}_1, \hat{\kappa}_1)$ that feeds the main algorithm as a warm start.

3.3 Monte Carlo Simulations

We evaluate the performance of the algorithm in Monte Carlo simulations in which workers and firms are drawn from a discrete-types data-generating process that follows the structure in equation (3). First, when idiosyncratic shocks are set to zero, the algorithm recovers all parameters of the wage equation exactly—the worker and firm fixed effects, the worker and firm growth coefficients, and the assignment functions ρ and κ. This is a sanity check that the procedure introduces no bias of its own under perfect identification. Second, as the variance of idiosyncratic shocks rises, the algorithm continues to recover the true parameters with minimal error: correlations between true and estimated parameters remain close to one, and variance ratios cluster tightly around one across all four parameters. The initialization steps are important to this performance. Together, these results show that the full procedure, including the initialization by k-means and refinement on the first-difference equation, and coordinate descent on the levels objective, jointly recovers the latent grouping structure and the parameters of the wage equation. Full simulation details and results are presented in Appendix C.

3.4 Selection of Group Numbers

The choice of number of worker and firm groups, K_w and K_f, involves a trade-off between model flexibility and estimation precision. A larger number of groups allows the model to capture finer heterogeneity in returns to experience, but it also reduces the number of observations within each group and therefore increases the noise in the group-specific estimates. Standard information criteria, such as AIC or BIC, are not directly applicable in this setting because the effective number of parameters depends on the estimated assignment functions. This dependence complicates the use of conventional penalization approaches and motivates our reliance on cluster-quality diagnostics, robustness checks, and computational feasibility.[2]

In our baseline specification, we set $K_w = K_f = 20$ to balance rich heterogeneity against sufficient observations within each group. To assess sensitivity to this choice, we re-estimate the model under alternative group structures $(K_w, K_f) \in \{(20, 50), (50, 20), (50, 50)\}$. The results are robust across these alternatives. We report these robustness checks in Appendix D.

3.5 Relationship to Existing Methods

Our approach is most closely related to two strands of the recent literature. The first is the grouped fixed-effects estimator of Bonhomme and Manresa (2015). In the one-sided special case, say, $K_f = 1$, so that all firms share a common growth effect, our estimator reduces to their GFE applied to worker-side growth heterogeneity, recoverable by k-means. The extension to two-sided classification introduces a joint problem in which group memberships and type-specific coefficients must be determined simultaneously. This is not solved by k-means; our coordinate descent strategy, alternating between worker and firm assignments, provides a computationally tractable solution that scales to large administrative datasets.

The second is a recent literature that uses finite-mixture models to estimate latent worker and firm types in matched employer-employee data (Bonhomme et al., 2019; Lentz et al., 2023; Mann, 2024). Our approach differs three aspects. First, we treat memberships as fixed parameters to be estimated rather than as realizations from a distribution. This interpretation of ρ and κ as fixed features of workers and firms is analogous to individual fixed effects in AKM. Second, we estimate by minimizing the sum of squared residuals directly, imposing no

[2] Lentz et al. (2023) discuss the practical difficulty of using the elbow method to select the number of clusters.

parametric assumption on the error term, in contrast to the likelihood-based estimation of mixture models. Third, the EM machinery commonly used for mixture estimation does not apply here; we estimate assignments by iterative coordinate descent. The trade-off is that we cannot exploit the well-developed convergence theory of EM, but we avoid imposing a parametric error distribution and remain in the fixed-effects tradition of panel econometrics.

4 Empirical Results

In this section, we first describe the data source and present the empirical finding of substantial heterogeneity in returns to experience across worker and firm types. We then quantify the contributions of worker heterogeneity, firm heterogeneity, and their covariance to life-cycle wage inequality. Next, we examine dynamic sorting patterns over the life cycle. Finally, we characterize the estimated growth types in terms of observable worker and firm attributes.

4.1 Data

We use the Veneto Worker History (VWH) dataset, a social security administrative panel that covers the entire working population and private firms in the region of Veneto, Northeast Italy, from 1975 to 2001. Veneto is the fourth most populous region in Italy, with a population of about five million in 2012.[3] The data include unique worker and firm identifiers and employment histories. Worker-level information includes annual earnings, the number of days worked per year for each job spell, contract type, occupation, gender, age, birthplace, and seniority within the firm.[4] Firm-level information includes location, start and closure dates, and industry classification.

We restrict the analysis to 1984–2001, since data quality is lower in earlier years (Kline et al., 2020). We then follow standard sample selection procedures in the AKM literature. First, we restrict the sample to workers aged 25 or older. Specifically, we consider workers born in 1959 or later in the Veneto region, which allows us to observe their labor market outcomes from age 25 through their mid-forties. We drop workers with missing gender information and employment spells without firm identifiers. We also exclude part-time workers, apprentices, and public sector workers, whose wages are not directly comparable to those of regular full-time employees. To obtain a unique worker–firm match in each year, we retain only the worker's primary job when a worker holds multiple jobs in a given year. We define the primary job as the job with the

[3] Previous papers that use the VWH data include Card et al. (2014), Battisti (2017), Bartolucci et al. (2018), Serafinelli (2019), Kline et al. (2020), and Arellano-Bover and Saltiel (2026).

[4] Annual earnings are recorded without top-coding. Occupation is categorized as white-collar, blue-collar, manager, or apprentice.

highest annual earnings. Annual earnings are adjusted for inflation using 2003 as the base year. For each worker, we observe both days worked and total earnings from their primary job. We construct average daily wages by dividing total earnings by days worked. Following Kline et al. (2020), we drop observations with log wage changes greater than 1 or less than −1. Finally, we restrict the sample to the largest connected set to satisfy the identification requirement in the Abowd et al. (1999) framework.

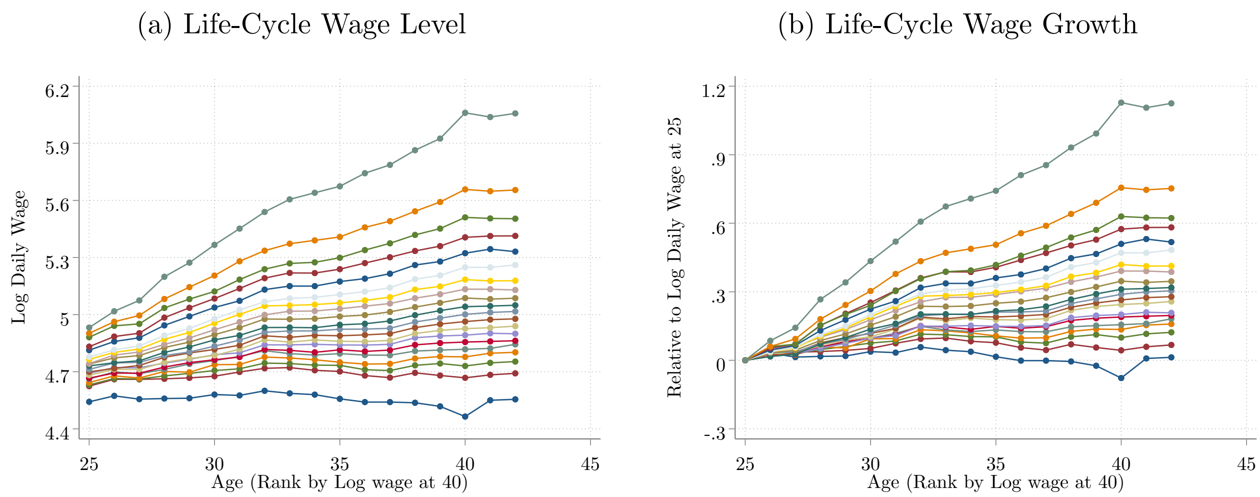
Table 1: Summary Statistics

	(1) Mean	(2) SD
Worker Characteristics		
Fraction of male	0.65	0.48
Worker's age	30.29	4.18
Tenure since age 25	3.48	3.00
Experience since age 25	5.27	3.85
Fraction of blue-collar	0.65	0.48
Fraction of white-collar	0.34	0.47
Fraction of manager	0.01	0.09
Job, Wage Characteristics		
Primary job daily wages	130.56	164.60
Annual paid days at primary job	237.26	104.19
Annual wage growth	0.02	0.12
Within firm wage growth	0.02	0.11
Between firm wage growth	0.01	0.21
Firm Characteristics		
Firm size (no. employees)	373.13	1274.72
Number of person-year observations:		6,910,126
Number of unique workers:		1,136,248
Number of unique firms:		397,603

Notes: This table reports summary statistics for our analysis sample from the Veneto Worker History Panel. All statistics are calculated across person-year observations. Firm size is the firm's primary-job headcount calculated before sample restrictions. Wage-growth statistics are computed only for worker-years whose primary job lasted at least 156 days in both the current and the previous year; sample restrictions are described in Section 4.1.

Table 1 reports summary statistics for our analysis sample. The sample comprises 6.9 million person-year observations, covering 1.1 million workers and roughly 400,000 firms over 1984–2001. The workforce is 65% male. Most workers hold blue-collar positions (65%), with the remainder primarily in white-collar roles (34%). Mean daily earnings at the primary job are 131 in 2003 prices, and workers are paid for 237 days per year on average. Annual wage growth averages 0.02 log points, with within-firm growth of 0.02 and between-firm growth of 0.01 log points.

Figure 1: Life-Cycle Wage Profile



Notes: This figure compares life-cycle wage trajectories for male workers in Veneto, grouped into 20 bins according to their wage rank at age 40. Panel (a) plots log daily wage levels by age. Panel (b) plots cumulative wage growth, measured as log daily wages at each age relative to log daily wages at age 25.

Figure 1 documents a stark divergence in life-cycle wage profiles among male workers in Veneto, Italy. We group workers into 20 bins based on their wage rank at age 40 and track their wage trajectories from age 25 onward. Panel (a) plots the wage level and Panel (b) plots the cumulative wage growth. Their trajectories diverge sharply. Workers in the top wage group experience cumulative wage growth exceeding 1.2 log points, while workers in the bottom group see essentially no real wage growth.

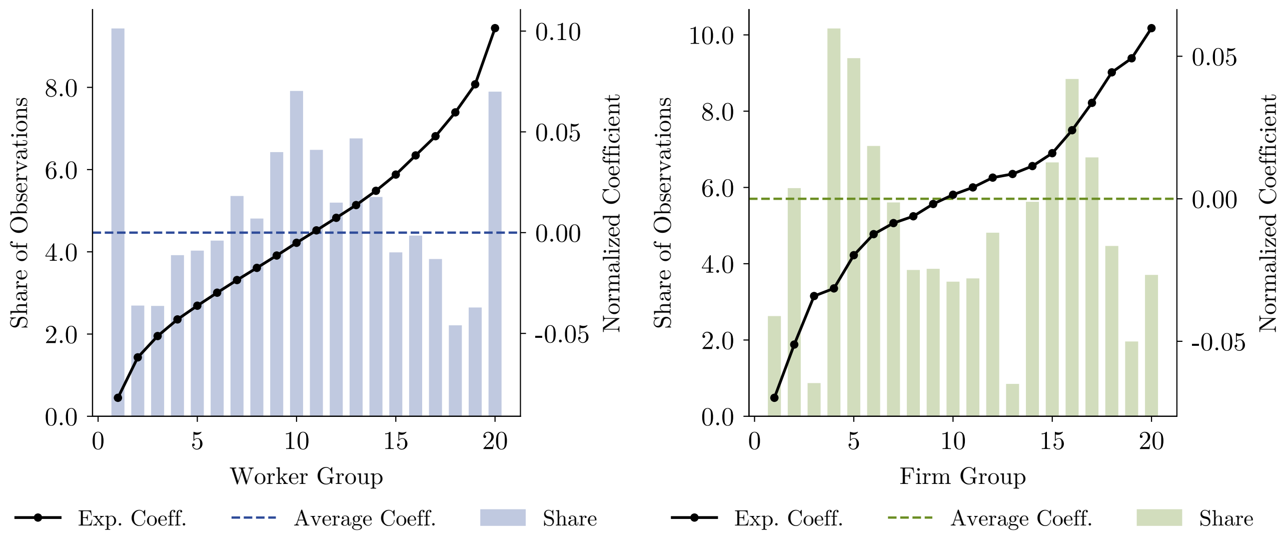
4.2 Heterogeneity in Returns to Experience

Figure 2 displays the estimated returns to experience across worker and firm groups. The black dotted lines plot the coefficient for each group on the right axis, normalized so that the average equals zero (horizontal dashed line).[5] The bars show the share of person-year observations in each group on the left axis.

Our main finding is that returns to experience exhibit substantial heterogeneity across both worker and firm groups. The returns to experience specific to each worker group span a wide range from approximately -0.06 for group 1 (6 log points below average) to 0.10 for group 20 (10 log points above average), a spread of 16 log points per year of experience. Workers

[5]We normalize coefficients by subtracting the average. This facilitates comparison across groups while preserving the dispersion. Negative coefficients indicate that, compared with workers of the same observables, this group accumulates returns to experience below the average.

Figure 2: Returns to Experience by Worker and Firm Groups



Notes: This figure shows the estimated returns to experience by worker and firm groups. The black dotted lines plot the estimated coefficients for each group (right y-axis), normalized so that the average equals zero. The bars show the share of person-year observations in each group (left y-axis). Both worker and firm groups are ranked by the magnitudes of their coefficients.

observations are relatively balanced across groups, with each group containing 2–8% of person-year observations and somewhat larger concentrations at the extremes (groups 1 and 20). Firm groups display a comparable range of heterogeneity, with estimated coefficients spanning from approximately -0.06 (group 1) to 0.06 for (group 20), a spread of 12 log points. Unlike the worker side, however, the extreme firm groups with very high or very low returns to experience collectively employ a much smaller share of the workforce.

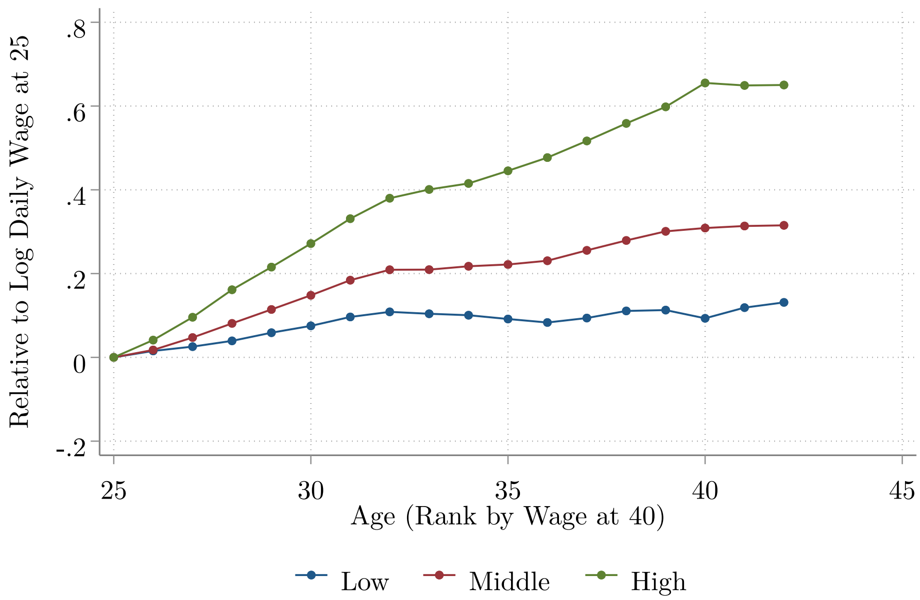
The distributional patterns of estimated returns to experience have important implications for wage inequality. Although the unweighted spread in coefficients is similar across the worker and firm types, the observation-weighted dispersion is substantially larger on the worker side: observations populate all worker growth types with notable mass at the extremes, whereas observations cluster in firm types offering near-average returns, with the extremes accounting for a much smaller share. This suggests that worker heterogeneity may contribute more than firm heterogeneity to life-cycle wage inequality. The following three subsections performs formal assessment of such relative contributions.

4.3 Decomposing Life-Cycle Wage Inequality

Our framework provides a natural decomposition of life-cycle wage inequality into four components: the worker fixed effect, the firm fixed effect, the worker growth component, and the firm growth component. We perform two such decompositions: one of wage levels and one of wage growth.

To implement the decomposition, we first construct a balanced panel of male workers from a single cohort, which holds worker composition fixed and lets us track the same workers over their careers. We then rank workers into three distinct terciles $i \in \{g_{\min}, g_{\text{med}}, g_{\max}\}$ based on their wage at age 40.[6] Figure 3 displays the raw life-cycle wage profiles for each tercile, adjusted for year fixed effects. Within each tercile, we compute the mean of each wage component. The differences in these means across terciles reveal how each component contributes to between-tercile wage inequality over the life cycle.

Figure 3: Life-Cycle Wage Growth



Notes: This figure shows the life-cycle wage growth profiles for workers in three terciles in Veneto, after adjusting for year fixed effects. The profiles are calculated as the difference between log daily wages at later ages and log daily wages at age 25. Workers are classified into three terciles based on their wage rank at age 40: low, middle, and high, corresponding to the 33rd, 66th, and 99th percentiles, respectively.

Life-cycle Wage Level Decomposition. Let $\mathbb{E}_{a,g} \ln y_{i,t}$ denote the average log daily wage for tercile g at age a. We decompose the wage level differences between a tercile g and the lowest

[6] We use age 40 as the reference point because the panel is not long enough to follow workers over the full life cycle, and the wage-age profile begins to flatten around age 40.

tercile g_{min} as follows:

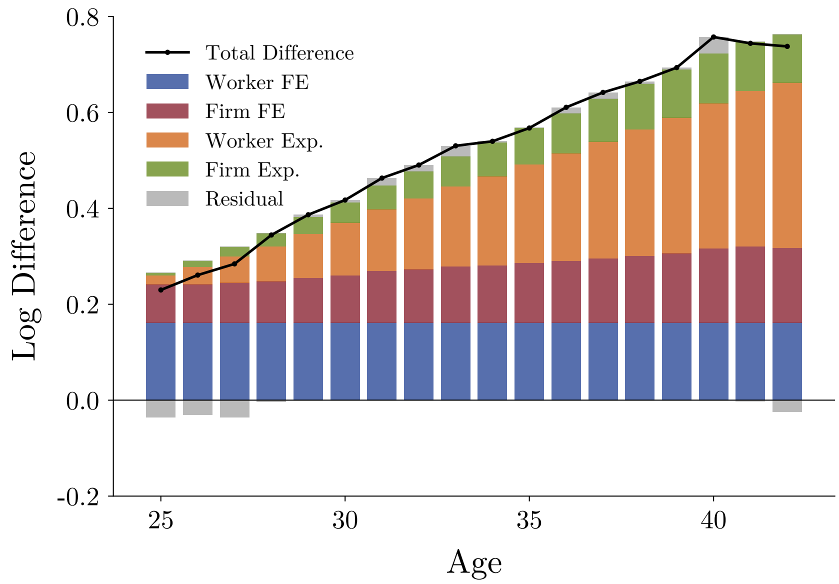
$$\begin{aligned}\mathbb{E}_{a,g}\Big[\ln y_{i,t}\Big] - \mathbb{E}_{a,g_{\min}}\Big[\ln y_{i,t}\Big] = &\underbrace{\Big(\mathbb{E}_{a,g}\big[\hat{\alpha}_i\big] - \mathbb{E}_{a,g_{\min}}\big[\hat{\alpha}_i\big]\Big)}_{\text{Diff. in Worker Effect}} + \underbrace{\Big(\mathbb{E}_{a,g}\big[\hat{\psi}_j\big] - \mathbb{E}_{a,g_{\min}}\big[\hat{\psi}_j\big]\Big)}_{\text{Diff. in Firm Effect}} \\ &+ \underbrace{\Big(\mathbb{E}_{a,g}\big[\hat{\beta}_i \operatorname{Exp}_{i,t}\big] - \mathbb{E}_{a,g_{\min}}\big[\hat{\beta}_i \operatorname{Exp}_{i,t}\big]\Big)}_{\text{Diff. in Worker Growth Component}} \\ &+ \underbrace{\Big(\mathbb{E}_{a,g}\Big[\sum_{\kappa(j)} \hat{\phi}_{\kappa(j)} \operatorname{Exp}_{i,t}^{\kappa(j)}\Big] - \mathbb{E}_{a,g_{\min}}\Big[\sum_{\kappa(j)} \hat{\phi}_{\kappa(j)} \operatorname{Exp}_{i,t}^{\kappa(j)}\Big]\Big)}_{\text{Diff. in Firm Growth Component}} \\ &+ \mathbb{E}_{a,g}\text{Residual}\end{aligned} \tag{6}$$

Equation (6) decomposes the wage difference between terciles into four channels. The worker fixed effect α_i captures permanent unobserved attributes, such as latent ability or skills, rewarded uniformly across firms. The firm fixed effects ψ_j captures firm-specific wage premia; its difference across terciles reveals whether higher-wage workers sort into firms paying higher premia. The worker growth component $\beta_i \text{Exp}_{i,t}$ captures returns to experience accumulated from general labor market experience, and the firm growth component $\sum_{\kappa(j)} \phi_{\kappa(j)} \text{Exp}_{i,t}^{\kappa(j)}$ captures returns to experience accumulated from particular firm types.

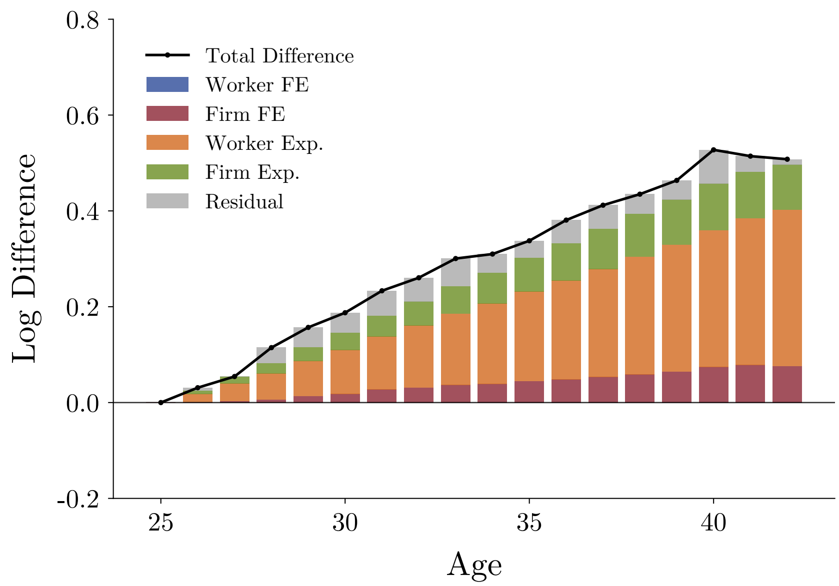
Life-cycle Wage Growth Decomposition. To decompose life-cycle wage growth relative to age 25, we apply the same decomposition to within-tercile differences over age. Define $\Delta_{a,g} \ln y_{i,t} \equiv \mathbb{E}[\ln y_{i,t} \mid \text{age} = a, g] - \mathbb{E}[\ln y_{i,t} \mid \text{age} = 25, g]$. Replacing the expectation operator $\mathbb{E}_{a,g}$ in equation (6) with the difference operator $\Delta_{a,g}$ yields a decomposition of differential life-cycle wage growth between terciles into the same four channels plus a residual.

$$\begin{aligned}\Delta_{a,g}\Big[\ln y_{i,t}\Big] - \Delta_{a,g_{\min}}\Big[\ln y_{i,t}\Big] = &\underbrace{\Big(\Delta_{a,g}\big[\hat{\alpha}_i\big] - \Delta_{a,g_{\min}}\big[\hat{\alpha}_i\big]\Big)}_{\text{Diff in Worker FE}} + \underbrace{\Big(\Delta_{a,g}\big[\hat{\psi}_j\big] - \Delta_{a,g_{\min}}\big[\hat{\psi}_j\big]\Big)}_{\text{Diff in Firm FE}} \\ &+ \underbrace{\Big(\Delta_{a,g}\big[\hat{\beta}_i \operatorname{Exp}_{i,t}\big] - \Delta_{a,g_{\min}}\big[\hat{\beta}_i \operatorname{Exp}_{i,t}\big]\Big)}_{\text{Diff in Worker Growth Component}} \\ &+ \underbrace{\Big(\Delta_{a,g}\Big[\sum_{\kappa(j)} \hat{\phi}_{\kappa(j)} \operatorname{Exp}_{i,t}^{\kappa(j)}\Big] - \Delta_{a,g_{\min}}\Big[\sum_{\kappa(j)} \hat{\phi}_{\kappa(j)} \operatorname{Exp}_{i,t}^{\kappa(j)}\Big]\Big)}_{\text{Diff in Firm Growth Component}} \\ &+ \Delta_{a,g}\text{Residual}\end{aligned} \tag{7}$$

Figure 4: Life-Cycle Wage Decomposition



(a) Life-Cycle Wage Level Decomposition



(b) Life-Cycle Wage Growth Decomposition

Notes: These figures show the decomposition of life-cycle wages and wage growth between workers in the top and bottom terciles. Panel (a) shows the life-cycle wage level decomposition at each age, while Panel (b) decomposes life-cycle wage growth relative to age 25.

Results. Figure 4 presents the decomposition for wage differences between the top and bottom terciles. Panel (a) shows the level decomposition at each age, and Panel (b) decomposes wage growth relative to age 25.

In Panel (a), the wage gap between the top and bottom terciles widens from roughly 0.24 log points at age 25 to 0.74 log points by age 42. The worker fixed effect component remains flat by construction: the balanced panel holds worker composition fixed within each tercile. The firm fixed effect component, by contrast, rises steadily over the life cycle, reflecting progressively stronger sorting of top-tercile workers into firms paying higher wage premia. The remaining widening of the level gap over the life cycle reflects two-sided heterogeneity in returns to experience.

Panel (b) more clearly decomposes this widening of wage inequality. It differences each component in Panel (a) relative to its age-25 value, thereby isolating the differential growth of each component over the life cycle. Top-tercile workers experience steeper returns for each additional year of experience, and this advantage compounds over the career. They also gradually sort into better-paying firms and accumulate experience at firms that offer steeper returns to experience. By age 42, the cumulative wage-growth gap reaches 0.51 log points, with the worker growth component accounting for nearly two-thirds of the increase. Of the remaining third, the firm growth component accounts for about 19%, and increased sorting into higher-paying firms accounts for about 15%.

We report additional decomposition results for the top-versus-middle and the middle-versus-bottom tercile comparisons in Appendix A.1. For the middle-versus-bottom comparison, the firm-side contribution (increasing firm fixed effects and the firm growth component combined) accounts for roughly half of the widening gap, suggesting that firm heterogeneity plays a more prominent role in shaping wage inequality in the lower half of the distribution.

4.4 Variance Decomposition

We next quantify how much of the variance in log wages is accounted for by each component of our decomposition. Building on the AKM literature (e.g., Card et al., 2013; Alvarez et al., 2018), we perform variance decompositions to illuminate the sources of life-cycle wage inequality.

Variance Decomposition. The variance of log wages decomposes as the sum of the variances of each component in equation (3), together with all pairwise covariances and residual variation:

$$\mathrm{Var}\left(\ln y_{i,t}\right) = \underbrace{\mathrm{Var}\left(\hat{\alpha}_i\right)}_{\text{Var(Worker FE)}} + \underbrace{\mathrm{Var}\left(\hat{\psi}_{J(i,t)}\right)}_{\text{Var(Firm FE)}} + \underbrace{\mathrm{Var}\left(\hat{\beta}_{\rho(i)}\,\mathrm{Exp}_{i,t}\right)}_{\text{Var(Worker GC)}} + \underbrace{\mathrm{Var}\left(\sum_{\kappa(j)} \hat{\phi}_{\kappa(j)}\,\mathrm{Exp}_{i,t}^{\kappa(j)}\right)}_{\text{Var(Firm GC)}} \tag{8}$$
$$+ \text{Covariance Terms} + \text{Residuals}$$

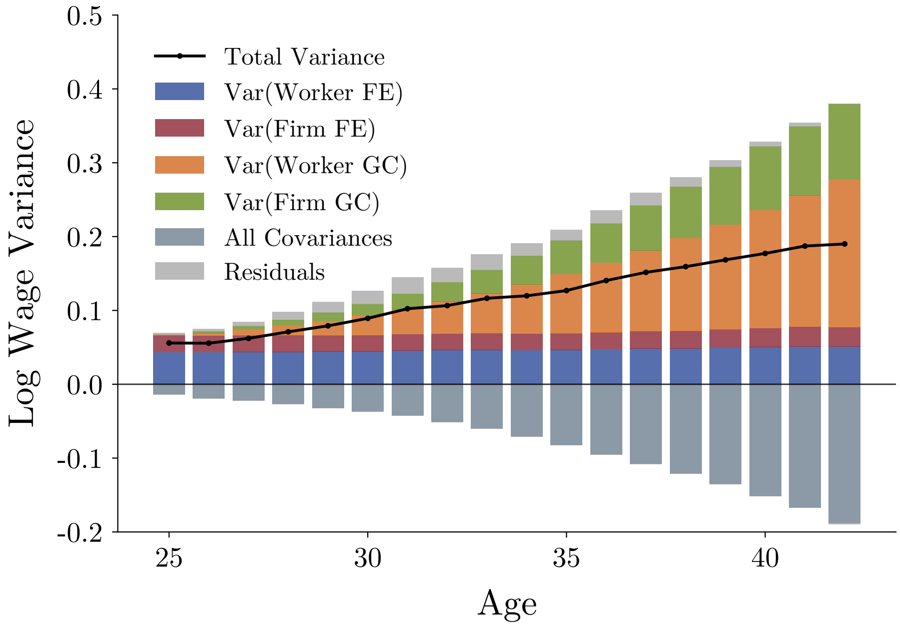
The first four terms capture the independent variance contributions of the worker fixed effect, the firm fixed effect, the worker growth component, and the firm growth component. The covariance terms capture how these components co-vary, including the cross-sectional sorting between worker and firm fixed effects familiar from AKM, together with new covariances involving the worker and firm growth components. We compute the decomposition at each age using the estimated components from the two-way clustering procedure of Section 3. We discuss the variance components first, then the covariance structure.

Figure 5 reports the contribution of each variance component to log wage variance at each age. Several patterns stand out. First, the variance of worker and firm fixed effects is approximately constant over the life cycle, reflecting the time-invariant nature of these components in a balanced panel. Second, the variance of the worker growth component rises sharply with age and accounts for approximately two-thirds of the rise in wage variance between age 25 and age 42, while the firm growth component accounts for the remaining one-third.

The two growth components differ substantially in scale. By age 42, the variance of worker growth component reaches 0.2, roughly twice the variance of the firm growth component. This larger dispersion in the worker growth component is consistent with the wider weighted spread in worker experience effects observed in Section 4.2. Two features moderate the firm growth contribution. First, dispersion in firm experience effects is smaller than dispersion in worker experience effects. Second, worker mobility compresses firm-side variation: most workers in our sample are employed at multiple firms over the observation period, averaging out their exposure to any single firm growth type. Despite these moderating forces, the firm growth component accounts for a non-trivial one-third of rising life-cycle inequality, indicating that where workers accumulate experience matters for long-run wage trajectories.

Covariance Terms. The variance components above quantify independent contributions of worker and firm heterogeneity but do not capture how the two-sided heterogeneity interact. We therefore turn to the covariance structure over the life cycle, which reveals how dynamic sorting between workers and firms amplifies or moderates the rise in inequality.

Figure 5: Variance Decomposition

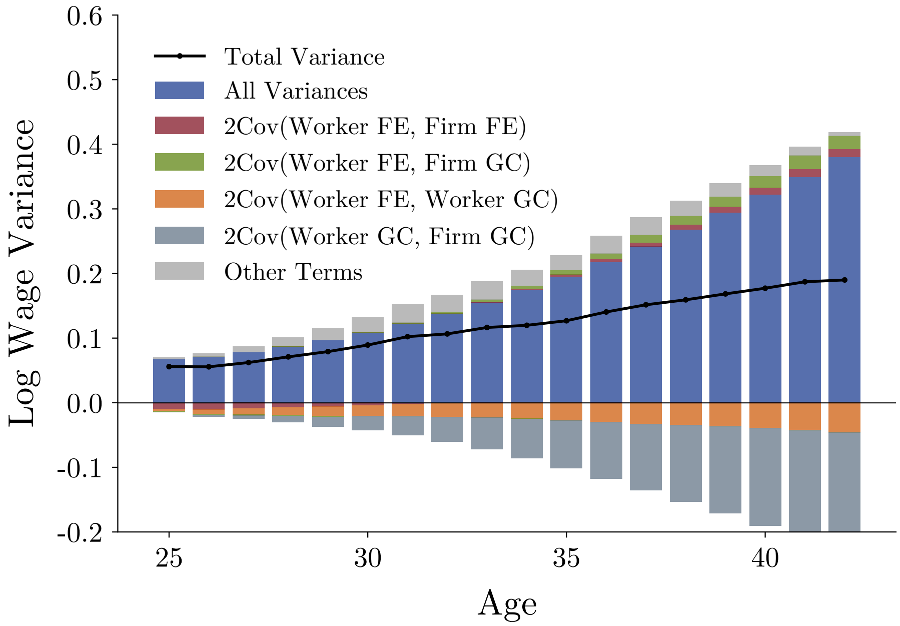


Notes: This figure decomposes the variance of log wages over the life cycle. The black line represents total variance. The stacked bars show $\text{Var}(\hat{\alpha}_i)$ (Worker FE) in blue, $\text{Var}(\hat{\psi}_j)$ (Firm FE) in red, $\text{Var}(\hat{\beta}_{\rho(i)}\,\text{Exp}_{i,t})$ (Worker Growth Component) in orange, $\text{Var}(\sum_{\kappa(j)} \hat{\phi}_{\kappa(j)}\,\text{Exp}_{i,t}^{\kappa(j)})$ (Firm Growth Component) in green. All covariance terms are combined in light blue, and residuals in gray.

The "Covariance Terms" in the variance decomposition above expand into a sum of pairwise covariances among the components of equation (3). Figure 6 focuses on four covariances that capture important dimensions of worker-firm sorting. The covariance between worker and firm fixed effects captures classic AKM-style sorting between permanent worker productivity and firm wage premia. The remaining three covariances are new to our framework: the covariance between worker fixed effects and the firm growth component measures sorting between worker productivity and firm experience returns; the covariance between worker fixed effects and the worker growth component captures the within-worker relationship between initial wages and wage growth; and the covariance between the worker and firm growth components measures sorting between high-growth workers and high-growth firms. The "Other Terms" collect the remaining pairwise covariances, which are smaller in magnitude.

The covariance structure in Figure 6 reveals two competing sorting forces over the life cycle. First, positive sorting between high-wage workers and high-wage firms amplifies inequality. Two covariance components drive this pattern. The covariance between worker fixed effects and firm fixed effects turns from slightly negative early in the career to positive as workers age, indicating that high-wage workers progressively sort into high-wage firms. Furthermore, the covariance between worker fixed effects and the firm growth component also rises with age, indicating that high-wage workers additionally match with firms offering steeper experience profiles. Together,

Figure 6: Covariance Components



Notes: This figure decomposes the total variance of log wages into its covariance components over the life cycle. The black line shows total variance, while the blue bar represents the sum of all variance components (worker FE, firm FE, worker growth, and firm growth variances combined). Covariance terms are shown as $2\,\mathrm{Cov}(\hat{\alpha}_i, \hat{\psi}_j)$ (Worker FE with Firm FE) in red, $2\,\mathrm{Cov}(\hat{\alpha}_i, \sum_{\kappa(j)} \hat{\phi}_{\kappa(j)}\,\mathrm{Exp}_{i,t}^{\kappa(j)})$ (Worker FE with Firm Growth) in green, $2\,\mathrm{Cov}(\hat{\alpha}_i, \hat{\beta}_{\rho(i)}\,\mathrm{Exp}_{i,t})$ (Worker FE with Worker Growth) in orange, $2\,\mathrm{Cov}(\hat{\beta}_{\rho(i)}\,\mathrm{Exp}_{i,t}, \sum_{\kappa(j)} \hat{\phi}_{\kappa(j)}\,\mathrm{Exp}_{i,t}^{\kappa(j)})$ (Worker Growth with Firm Growth) in dark gray, and remaining terms in light gray.

these covariances describe a dynamic sorting process: workers with high initial productivity progressively match with firms that pay both higher levels and steeper growth.

Second, growth-type sorting operates in the opposite direction. The covariance between the worker growth component and firm growth component is negative throughout the life cycle and grows in magnitude with age, indicating that high-growth workers tend to work at low-growth firms. Another negative covariance between worker fixed effects and the worker growth component reinforces the moderating force: high-wage workers tend to have flatter growth profiles. Both negative covariances dampen the rise in inequality that the variance components alone would imply.

Together, the variance and covariance results describe a labor market with two competing sorting forces. Positive assortative matching between high-wage workers and high-wage, high-growth firms amplifies inequality over the life cycle, while negative sorting between worker and firm growth types, reinforced by the negative covariance between worker levels and worker growth, partly offsets it. The net effect is rising but moderated life-cycle wage inequality, with the worker growth component as the largest single source.

4.5 Dynamic AKM Decomposition

We now provide a complementary perspective on life-cycle wage inequality by aggregating the time-varying components of our model into a single time-varying worker effect:

$$\tilde{\alpha}_{i,t} = \hat{\alpha}_i + \hat{\beta}_{\rho(i)} \operatorname{Exp}_{i,t} + \sum_j \hat{\phi}_{\kappa(j)} \operatorname{Exp}^j_{i,t} .$$

This effect collects all portable components of a worker's wage that travel with the worker across employers. The first term is the worker fixed effect, which can be interpreted as the worker's initial human capital. The second term is accumulated through worker-side growth, capturing human capital accumulation due to the worker's own learning ability. The third term is accumulated through firm-side growth. Although these returns are earned at particular firms, once accumulated they become part of the worker's human capital stock and travel with the worker to subsequent employers.

With this dynamic worker effect in hand, the variance of log wages can be decomposed into the variance of the dynamic worker effect, the variance of the firm fixed effect, twice the covariance between the dynamic worker effect and the firm fixed effect, and a residual term:

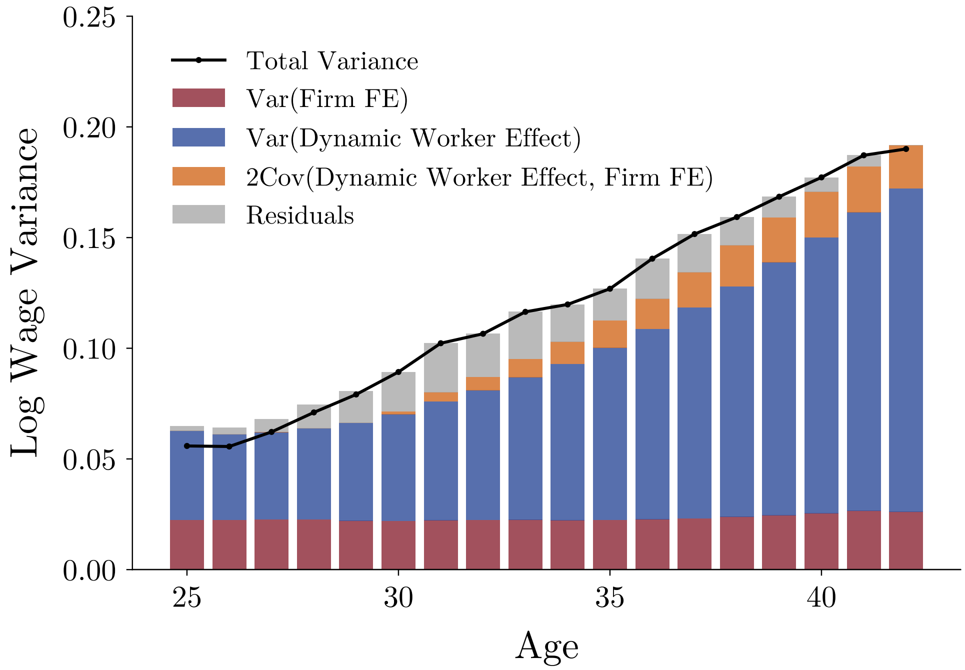
$$\operatorname{Var}(\ln y_{i,t}) = \operatorname{Var}(\tilde{\alpha}_{i,t}) + \operatorname{Var}\left(\hat{\psi}_{J(i,t)}\right) + 2\operatorname{Cov}\left(\tilde{\alpha}_{i,t}, \hat{\psi}_{J(i,t)}\right) + \text{Residuals.}$$

Figure 7 reports the three components of the dynamic AKM decomposition over the life cycle. First, the dynamic worker effect accounts for the largest share of wage variance. By age 42, variance of the dynamic worker effect reaches 0.11, roughly 70% of total wage variance. This result suggests that the main source of life-cycle wage inequality is the worker's accumulated skills. These skills are not only captured by the worker fixed effect, but also include human capital accumulated through both worker-side and firm-side returns to experience. The dominance of the dynamic worker effect thus reinforces the finding of Section 4.4 that life-cycle wage inequality reflects substantial heterogeneity in returns to experience on both worker and firm sides.

Second, the variance of firm fixed effects is small and roughly constant, at about 0.02 throughout the life cycle. This suggests that cross-sectional dispersion in firm wage premia accounts for a much smaller fraction of wage inequality once we properly account for the dynamic worker effect. It also explains little of the rise in life-cycle inequality. This does not mean that firms are unimportant. Rather, their role appears mainly through two dynamic channels: firms differ in the wage growth they provide, and workers increasingly sort into firms with different wage premia over the career.

Third, the covariance between the dynamic worker effect and firm fixed effects rises with

Figure 7: Dynamic AKM Decomposition



Notes: This figure presents the dynamic AKM decomposition of log wage variance over the life cycle. The decomposition shows: $\text{Var}(\hat{\psi}_j)$ (Firm FE) in red, $\text{Var}(\tilde{\alpha}_{i,t})$ (Dynamic Worker Effect) in blue, $2\,\text{Cov}(\tilde{\alpha}_{i,t}, \hat{\psi}_j)$ (covariance between Dynamic Worker Effect and Firm FE) in orange and residuals in gray.

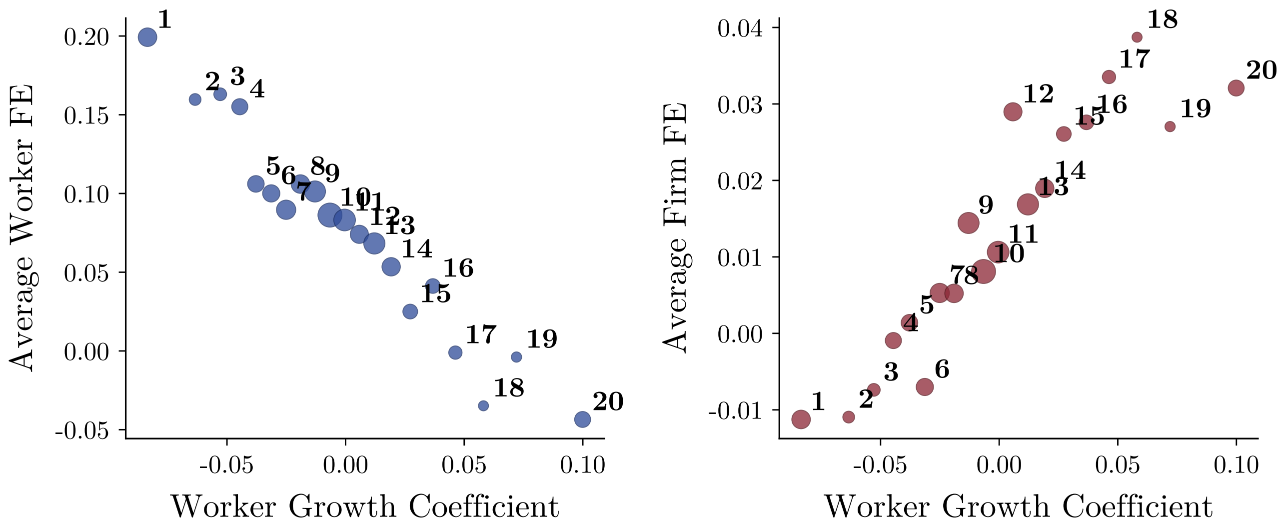
age, from near zero at age 25 to about 0.03 by age 42. This strengthening covariance shows that worker-firm sorting is not fixed at labor market entry, but instead builds over the career. Workers with higher dynamic worker effects progressively sort into firms with higher wage premia. This dynamic sorting amplifies life-cycle inequality: initial differences in worker fixed effects and growth potential are reinforced as these workers move toward better-paying firms. As a result, the widening of wage gaps over the life cycle reflects not only heterogeneous wage growth within workers, but also the increasing alignment between high-productivity workers and high-premium firms.

4.6 Correlations Across Types

We now examine the relationships between worker fixed effects, firm fixed effects, and heterogeneous returns to experience specific to workers and firms. This analysis reveals how wage levels and growth rates are distributed across worker and firm growth types.

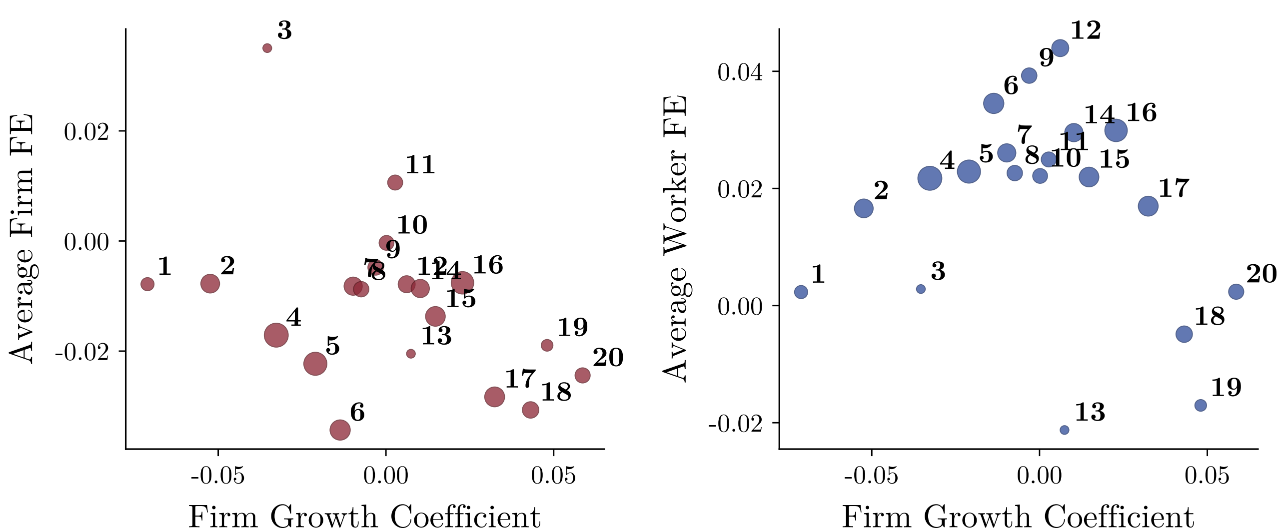
Figure 8 documents two relationships on the worker side. The left panel shows a negative correlation between worker fixed effects and worker growth coefficients, consistent with the negative $\text{Cov}(\hat{\alpha}_i, \hat{\beta}_{\rho(i)}\,\text{Exp}_{i,t})$ documented in the variance decomposition. Workers in higher-

Figure 8: High Growth Workers



Notes: This figure displays the relationship between worker growth coefficients, worker FE, and Firm FE across the 20 worker groups. The left panel plots the average worker fixed effect $\hat{\alpha}_i$ for each worker growth group against the group's worker growth coefficient $\hat{\beta}_{\rho(i)}$. The right panel plots the average firm fixed effect $\hat{\psi}_j$ for workers (at age 25) in each worker growth group against the group's worker growth coefficient. Groups are numbered 1-20, where group 1 has the lowest worker growth coefficient and group 20 has the highest. Circle size reflects the share of workers in each group.

Figure 9: High Growth Firms



Notes: This figure displays the relationship between firm growth coefficients, Firm FE, and worker FE across the 20 firm groups. The left panel plots the average firm fixed effect $\hat{\psi}_j$ for each firm growth group against the group's firm growth coefficient $\hat{\phi}_{\kappa(j)}$. The right panel plots the average worker fixed effect $\hat{\alpha}_i$ for workers employed at firms in each firm growth group against the group's firm growth coefficient. Groups are numbered 1-20, where group 1 has the lowest firm growth coefficient, and group 20 has the highest. Circle size reflects the share of firms in each group.

growth groups tend to have lower worker fixed effects. Specifically, workers in group 1 (lowest growth) have an average worker fixed effect of approximately 0.20, while workers in group 20 (highest growth) have an average worker fixed effect of approximately -0.05. This negative relationship indicates that workers with steeper wage growth tend to start from lower wage levels.

The right panel of Figure 8 shows a contrasting positive relationship between worker growth coefficients and firm fixed effects. Specifically, it reports the average firm fixed effects at age 25. Even at labor market entry, workers with higher growth coefficients are already employed at firms with higher wage premia. Workers in the lowest-growth groups (1–6) are at firms with average firm fixed effects below or near 0, while those in the highest-growth groups (16–20) are at firms with average firm fixed effects of 0.03–0.04. This positive correlation demonstrates that the sorting of high-growth workers into high-wage firms is present from the start of the working life, before the life-cycle dynamics take effect.

In contrast, we find no systematic relationship between firm growth coefficients and either firm fixed effects or worker fixed effects, as shown in Figure 9. The left panel shows that average firm fixed effects scatter around zero across all firm growth groups, and the correlation between these two firm-level components is close to zero. Similarly, the right panel shows no clear relationship between worker fixed effects and firm growth coefficients. Workers employed at firms with different growth profiles do not differ systematically in their average worker fixed effects. High-growth firms are neither high-wage nor low-wage employers on average, and they employ workers across the distribution of worker fixed effects.

Finally, Figure 10 documents strong negative sorting between worker and firm growth types. Workers with the highest returns to experience tend to be employed at firms with lower returns to experience, and vice versa. This negative sorting attenuates life-cycle inequality. If high-growth workers were also concentrated at high-growth firms, the compounding of worker and firm growth effects would generate substantially greater inequality than observed.

Together, these patterns reveal an asymmetry in labor market sorting. Positive sorting of high-growth workers into high-wage firms amplifies inequality by channeling workers with higher returns to experience into firms offering higher wage premia. The negative sorting between worker and firm growth types, however, operates as a countervailing force, preventing workers with high returns to experience from additionally benefiting from high-growth firm environments.

Figure 10: High Growth Workers and High Growth Firms

Notes: This figure displays the relationship between worker growth coefficients and firm growth coefficients across the 20 worker groups and 20 firm groups. Groups are numbered 1-20. Circle size reflects the share of workers in each group.

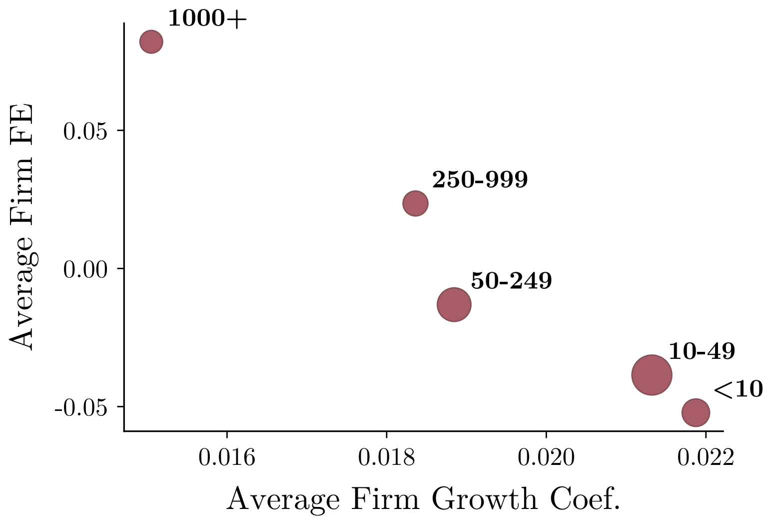
4.7 How Do Firm and Worker Growth Coefficients Relate to Observables?

Lastly, we relate the estimated returns to experience to observable worker and firm characteristics. We examine three dimensions: firm size, worker occupation, and firm industry. In each case, we examine how average fixed effects, $\hat{\alpha}_i$ or $\hat{\psi}_j$, and average growth coefficients,($\hat{\beta}_{\rho(i)}$ or $\hat{\phi}_{\kappa(j)}$, vary with observables.

Firm size: big firms pay more, small firms help workers grow more. Figure 11 plots the average firm fixed effect and firm growth coefficient by firm size. Larger firms pay systematically higher wage premia, but offer flatter returns to experience. This pattern reveals a level–slope trade-off across the firm size distribution. Workers at large firms receive higher pay but accumulate less wage growth from continued tenure, while workers at small firms start from lower wage premia but experience steeper wage growth.

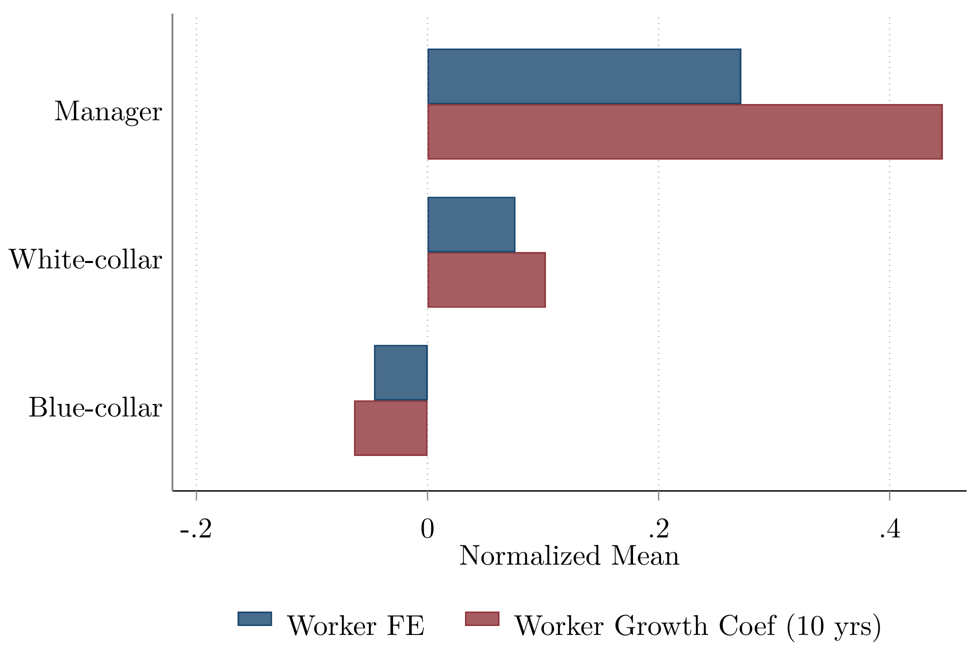
Occupation: manager pay gaps widen with experience. Figure 12 reports average worker fixed effects and worker growth coefficients by occupation. Managers display both higher worker fixed effects and higher worker growth coefficients, while blue-collar lie at the opposite end on both dimensions. As a consequence, the manager–non-manager wage gap widens over the life cycle, as managers' steeper returns to experience compound on top of their higher initial wage.

Figure 11: Firm Pay and Firm Growth by Firm Size



Notes: The figure plots the average firm fixed effect $\hat{\psi}_j$ and the average firm growth coefficient $\hat{\phi}_{\kappa(j)}$ across firm size bins.

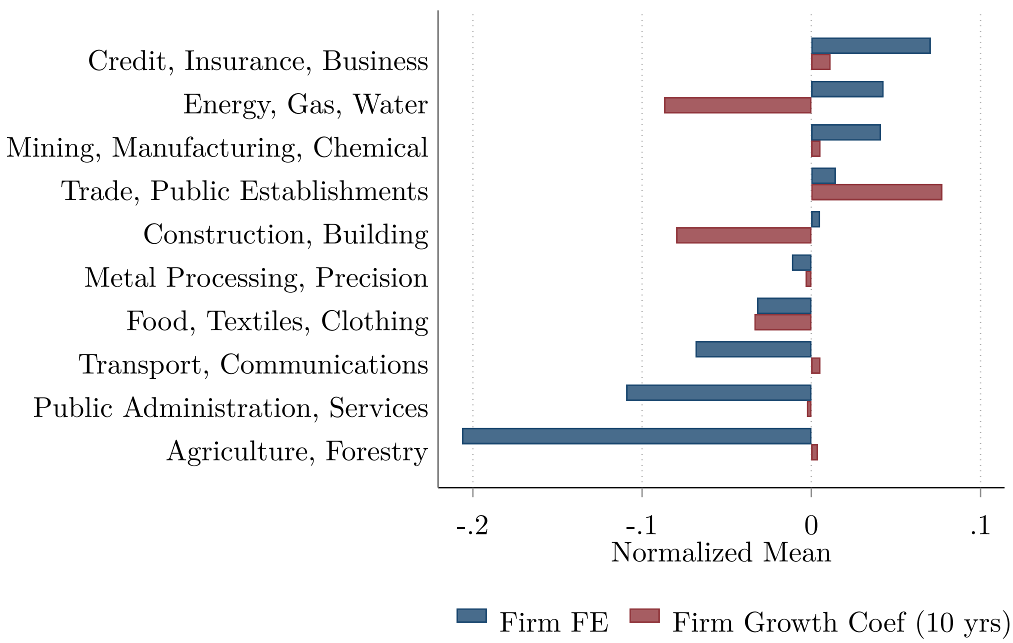
Figure 12: Worker Pay and Worker Growth by Occupation



Notes: The figure plots the average worker fixed effect $\hat{\alpha}_i$ and the average worker growth coefficient $\hat{\beta}_{\rho(i)}$ across occupation categories.

Industry: high pay and steep growth do not necessarily occur in the same sector. Figure 13 plots firm fixed effects and firm growth coefficients across industries. These two dimensions are largely orthogonal, and if anything, mildly negatively related, across sectors. The highest-paying industries, such as finance and utilities, are not necessarily the industries that offer the steepest returns to experience, while several lower-paying sectors deliver above-average firm growth effects. This sectoral pattern reinforces the firm-side finding from Section 4: firm pay premia and firm growth premia are distinct objects, both empirically and across observable margins. Workers therefore cannot simply infer the "best place to grow" from the "best place to start."

Figure 13: Firm Pay and Firm Growth by Industry



Notes: The figure plots the average firm fixed effect $\hat{\psi}_j$ and the average firm growth coefficient $\hat{\phi}_{\kappa(j)}$ across industry classifications.

Together, these three exercises help validate the latent estimates. High worker fixed effects are associated with managerial occupations, high firm fixed effects are associated with large firms, and the weak or negative relationship between firm pay levels and firm growth coefficients appears across both the firm-size and industry distributions.

5 Conclusion

Why do some workers experience rapid wage growth while others see their wages stagnate? This paper provides a framework to decompose the sources of life-cycle wage inequality into three dimensions: differences in how workers accumulate human capital, differences in the firms that employ them, and the evolving sorting of workers to firms over the career.

We extend AKM to incorporate two-sided heterogeneity in returns to experience using a grouped fixed-effects approach. Because estimating individual-level growth coefficients is infeasible with limited observations per worker and firm, we develop a two-way clustering algorithm that jointly classifies workers and firms into discrete growth types while estimating group-specific returns to experience. Monte Carlo simulations confirm that the algorithm recovers the latent grouping structure and the parameters of the wage equation.

Applying this framework to matched employer-employee data from Veneto, Italy, we document three main findings. First, worker heterogeneity in returns to experience dominates firm heterogeneity: the worker growth component accounts for two-thirds of the increase in wage dispersion over the life cycle, with the firm growth component accounting for the remaining one-third. Second, positive sorting of high-wage workers into high-wage and high-growth firms amplifies inequality over the life cycle, but sorting along the growth dimension works in the opposite direction—high-growth workers increasingly sort into low-growth firms, attenuating the divergence in wage trajectories. Third, the latent types have clear economic content: managers exhibit the highest worker growth coefficients, large firms pay more but small firms grow workers faster, and average firm pay premia and firm growth premia are not systematically correlated across industries.

Our framework opens several avenues for future work. One natural application is decomposing the sources of the college wage premium over the life cycle. Does the premium grow because college graduates accumulate human capital faster, work at firms offering steeper returns to experience, or progressively sort into employers that pay more? A parallel question applies to racial and gender wage gaps, where our approach can distinguish between differences in human capital accumulation from workers or firms and differential sorting across firm types. More broadly, the two-way clustering methodology extends to other settings with two-sided heterogeneity, such as student-teacher matching in economics of education and buyer-seller relationships in industrial organization.

References

Abowd, John M, Francis Kramarz, and David N Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Acemoglu, Daron and Jörn-Steffen Pischke, "Why Do Firms Train? Theory and Evidence," *The Quarterly journal of economics*, 1998, *113* (1), 79–119.

Altonji, Joseph G and Charles R Pierret, "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 2001, *116* (1), 313–350.

Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser, "Firms and the Decline in Earnings Inequality in Brazil," *American Economic Journal: Macroeconomics*, 2018, *10* (1), 149–189.

Andrews, Martyn J, Len 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: Statistics in Society*, 2008, *171* (3), 673–697.

_, Leonard Gill, Thorsten Schank, and Richard Upward, "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias," *Economics Letters*, 2012, *117* (3), 824–827.

Arellano-Bover, Jaime and Fernando Saltiel, "Differences in On-the-Job Learning across Firms," *Journal of Labor Economics*, 2026, *44* (1), 149–188.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin, "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 2014, *104* (6), 1551–1596.

Bartolucci, Cristian, Francesco Devicienti, and Ignacio Monz'on, "Identifying Sorting in Practice," *American Economic Journal: Applied Economics*, 2018, *10* (4), 408–438.

Battisti, Michele, "High Wage Workers and High Wage Peers," *Labour Economics*, 2017, *46*, 47–63.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of political economy*, 1967, *75* (4, Part 1), 352–365.

Biasi, Barbara and Heather Sarsons, "Information, Confidence, and the Gender Gap in Bargaining," *AEA Papers and Proceedings*, 2021, *111*, 174–78.

Bonhomme, Stephane and Angela Denis, "Fixed Effects and Beyond. Bias Reduction, Groups, Shrinkage and Factors in Panel Data," Documentos de Trabajo 2526, Banco de España 2025.

Bonhomme, Stéphane and Elena Manresa, "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 2015, *83* (3), 1147–1184.

__ , **Thibaut Lamadon, and Elena Manresa**, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 2019, *87* (3), 699–739.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "Discretizing Unobserved Heterogeneity," *Econometrica*, 2022, *90* (2), 625–643.

Borovičková, Katarína and Claudia Macaluso, "Heterogeneous Job Ladders," *Journal of Monetary Economics*, 2024, p. 103711.

Bradley, James, Bradley Setzler, and Felix Tintelnot, "The Effects of Foreign Multinationals on Workers and Firms in the United States," *The Quarterly Journal of Economics*, 2023, *138* (3), 1499–1542.

Burdett, Kenneth and Dale T. Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, *39* (2), 257–273.

Caldwell, Sydnee, Ingrid Haegele, and Jörg Heining, "Bargaining and Inequality in the Labor Market," *The Quarterly Journal of Economics*, 2026, *141* (1), 315–371.

Caliendo, Marco, Deborah A Cobb-Clark, Cosima Obst, and Arne Uhlendorff, "Risk Preferences and Training Investments," *Journal of Economic Behavior & Organization*, 2023, *205*, 668–686.

__ , __ , __ , **Helke Seitz, and Arne Uhlendorff**, "Locus of Control and Investment in Training," *Journal of Human Resources*, 2022, *57* (4), 1311–1349.

Card, David, 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*, 2016, *131* (2), 633–686.

__ , **Francesco Devicienti, and Agata Maida**, "Rent-Sharing, Holdup, and Wages: Evidence from Matched Panel Data," *Review of Economic Studies*, 2014, *81* (1), 84–111.

__ , **Jörg Heining, and Patrick Kline**, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 2013, *128* (3), 967–1015.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum, "Matching in Cities," *Journal of the European Economic Association*, 2022, *20* (4), 1478–1521.

Deming, David J., "Why Do Wages Grow Faster for Educated Workers?," NBER Working Paper 31373, National Bureau of Economic Research 2023.

Farber, Henry S and Robert Gibbons, "Learning and Wage Dynamics," *The Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

Gathmann, Christina and Uta Schönberg, "How General Is Human Capital? A Task-Based Approach," *Journal of Labor Economics*, 2010, *28* (1), 1–49.

Gregory, Victoria, "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search," Working Paper 2020-036, Federal Reserve Bank of St. Louis 2026.

_ , Guido Menzio, and David Wiczer, "The Alpha Beta Gamma of the Labor Market," *Journal of Monetary Economics*, 2024, p. 103695.

Guvenen, Fatih, "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 2009, *12* (1), 58–79.

_ , Fatih Karahan, Serdar Ozkan, and Jae Song, "What Do Data on Millions of US Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 2021, *89* (5), 2303–2339.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L Violante, "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic dynamics*, 2010, *13* (1), 15–51.

Helm, Ines, Alice Kügler, and Uta Schönberg, "Displacement Effects in Manufacturing and Structural Change," 2026. Working Paper.

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M Phillips, "Production and Learning in Teams," *Econometrica*, 2024, *92* (2), 467–504.

Hong, Long, "Coworker sorting, learning, and inequality," Technical Report, Arizona State University 2022.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Sources of Lifetime Inequality," *American Economic Review*, 2011, *101* (7), 2923–2954.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg, "Learning from Coworkers," *Econometrica*, 2021, *89* (2), 647–676.

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

Lange, Fabian, "The Speed of Employer Learning," *Journal of Labor Economics*, 2007, *25* (1), 1–35.

Lentz, Rasmus, Suphanit Piyapromdee, and Jean-Marc Robin, "The Anatomy of SortingEvidence From Danish Data," *Econometrica*, 2023, *91* (6), 2409–2455.

Ma, Xiao, Alejandro Nakab, and Daniela Vidart, "Human Capital Investment and Development: The Role of On-the-Job Training," *Journal of Political Economy Macroeconomics*, 2024, *2* (1), 107–148.

Magnac, Thierry, Nicolas Pistolesi, and Sébastien Roux, "Post-Schooling Human Capital Investments and the Life Cycle of Earnings," *Journal of Political Economy*, 2018, *126* (3), 1219–1249.

Mann, Lukas, "Spatial Sorting and the Rise of Geographic Inequality," February 2024. Job market paper, Department of Economics, Princeton University.

Ozkan, Serdar, Jae Song, and Fatih Karahan, "Anatomy of Lifetime Earnings Inequality: Heterogeneity in Job-Ladder Risk versus Human Capital," *Journal of Political Economy Macroeconomics*, 2023, *1* (3), 506–550.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Roussille, Nina, "The Role of the Ask Gap in Gender Pay Inequality," *The Quarterly Journal of Economics*, 2024, *139* (3), 1557–1610.

Rubinstein, Yona and Yoram Weiss, "Post Schooling Wage Growth: Investment, Search and Learning," in Eric A. Hanushek and Finis Welch, eds., *Handbook of the Economics of Education*, Vol. 1, Elsevier, 2006, pp. 1–67.

Serafinelli, Michel, ""Good" Firms, Worker Flows, and Local Productivity," *Journal of Labor Economics*, 2019, *37* (3), 747–792.

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

Sørensen, Kenneth Lykke and Rune Vejlin, "Worker and Firm Heterogeneity in Wage Growth: An AKM Approach," *Labour*, 2011, *25* (4), 485–507.

APPENDICES FOR ONLINE PUBLICATION

The Worker and Firm Origins of Life-Cycle Wage Inequality

by Xincheng Qiu, Jesse Wedewer, Runling Wu

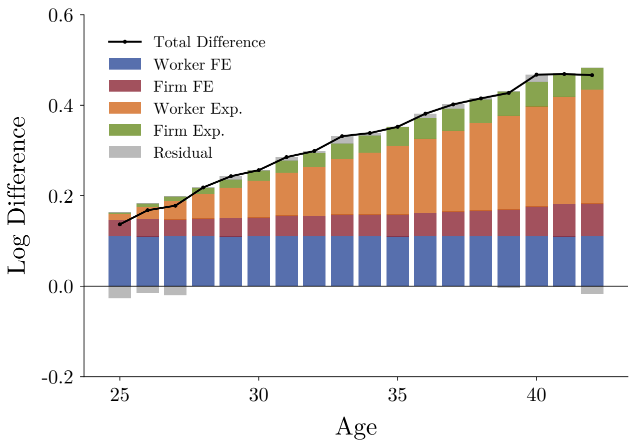
May 22, 2026

A Additional Figures and Tables

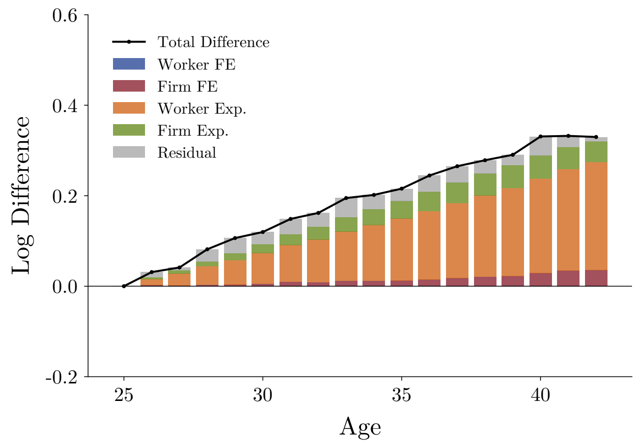
A.1 Additional Life-Cycle Wage Decompositions

This appendix presents additional life-cycle wage decompositions for tercile comparisons not reported in the main text. Figure A-1 reports the wage level and wage growth decompositions for the top versus middle tercile (Panels a–b) and the middle versus bottom tercile (Panels c–d).

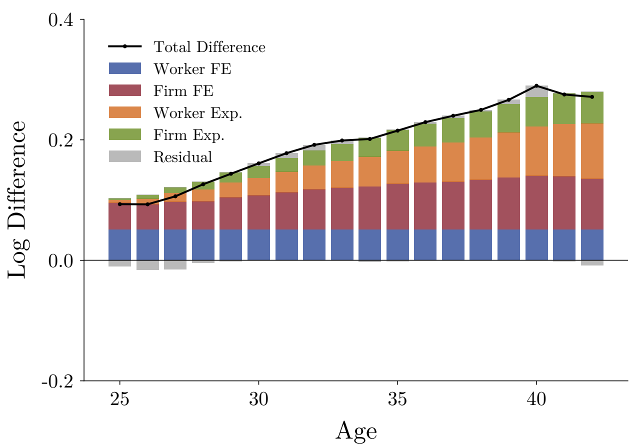
Figure A-1: Life-Cycle Wage Decomposition: Additional Tercile Comparisons



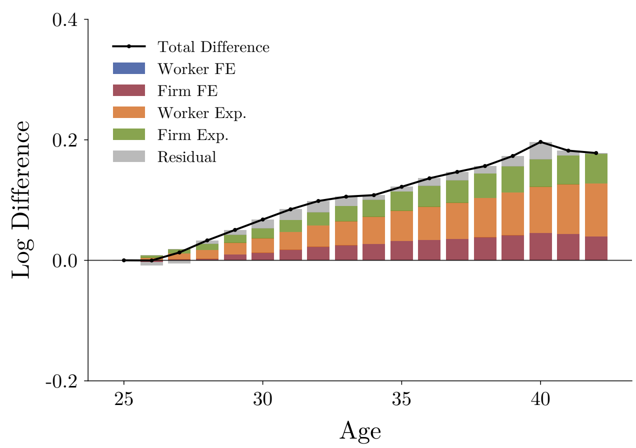
(a) Level: Top vs. Middle Tercile



(b) Growth: Top vs. Middle Tercile



(c) Level: Middle vs. Bottom Tercile



(d) Growth: Middle vs. Bottom Tercile

Notes: This figure presents life-cycle wage decompositions for additional tercile comparisons, where workers are ranked by wage at age 40. Panels (a)–(b) compare the top and middle terciles; Panels (c)–(d) compare the middle and bottom terciles. In each panel, the stacked bars decompose the wage level gap or wage growth gap into contributions from the worker fixed effect, firm fixed effect, worker growth component, firm growth component, and the residual.

B Detailed Three-Layer Algorithm

This appendix gives a formal statement of the iterative algorithm summarized in Section 3.

We describe the stayer first-difference framework that motivates the unit-level estimates $\hat{\beta}_i$ and $\hat{\phi}_j$ feeding the k-means initialization, present descriptive statistics for the stayer sample, and report a validation exercise that compares the group-level estimates from the full three-layer algorithm against the unit-level estimates from the stayer first-difference regression.

Stayers wage growth as a signal of latent growth types. The key observation underlying Layer 1 is that, for job stayers, the first difference of log wages eliminates the level fixed effects and isolates the worker and firm growth components. Let $\tilde{y}_{i,t}$ denote log daily wages first residualized on observables (gender-specific quadratic age effects in our application). For a worker who remains at the same firm in two consecutive periods ($j(i,t) = j(i,t-1)$), the first difference of residualized log wages strips out the level fixed effects α_i and ψ_j and isolates the worker and firm growth components, formally

$$\Delta \ln(\tilde{y}_{i,t}) = \beta_i + \phi_{j(i,t)} + \varepsilon_{i,t} \tag{A1}$$

where $\Delta ln(\tilde{y}_{i,t}) := ln(\tilde{y}_{i,t}) - ln(\tilde{y}_{i,t-1})$ and $\varepsilon_{i,t} = r_{i,t} - r_{i,t-1}$ [7]. Stayers' residualized wage growth $\Delta \ln \tilde{y}_{i,t}$ is therefore a direct, observable signal of the latent growth types $(\rho(i), \kappa(j))$: two workers in the same growth type exhibit similar wage-growth distributions across their stayer spells, and two firms in the same growth type deliver similar wage-growth distributions to the workers they retain. Layer 1 leverages this signal directly, applying k-means to the distributions of stayers' wage growth for workers and firms separately, using both cumulative distribution functions and means.

Stayer sample and summary statistics. The restriction to observations where workers remain at the same employer for two consecutive years substantially reduces the sample size and shifts its composition by removing short-lived jobs. Table A-1 reports descriptive statistics for the stayer sample, which serves as the input to Layer 1.

Let N denote the number of workers, J the number of firms, and N_T the total number of worker-year observations. Let K_w and K_f denote the numbers of worker and firm groups. Let $\text{Exp}_{i,t}$ denote worker i's total labor-market experience at time t, and let $e^{(j)}_{i,t}$ denote her cumulative tenure at firm j through period t. We define the group-level cumulative firm experience

[7] Given that in our context, $X_{i,t}$ are gender-specific age effects (quadratic), we work with residualized log daily wages where $\tilde{y}_{i,t}$ is derived from a regression $\ln(y_{i,t}) = \alpha_0 + \xi X_{i,t} + \ln(\tilde{y}_{i,t})$.

Table A-1: Summary Statistics: Job-Stayer Sample

	(1) Mean	(2) SD
Worker Characteristics		
Fraction of male	0.66	0.47
Worker's age	31.07	3.98
Tenure since age 25	4.69	2.99
Experience since age 25	6.31	3.72
Fraction of blue-collar	0.63	0.48
Fraction of white-collar	0.36	0.48
Fraction of manager	0.01	0.10
Job, Wage Characteristics		
Primary job daily wages	134.37	56.58
Annual paid days at primary job	267.61	83.71
Annual wage growth	0.03	0.16
Firm Characteristics		
Firm size (no. employees)	406.76	1343.60
Number of person-year observations:		4,588,997
Number of unique workers:		879,386
Number of unique firms:		263,104

Notes: This table reports summary statistics from the Veneto Worker History Panel for observations of job stayers (at the same firm) where daily wages are observed in two consecutive periods. All statistics are calculated across person-year observations. Firm size is the firm's primary-job headcount calculated before sample restrictions. Sample restrictions are described in Section 4.1.

as the total time worker i has spent at any firm assigned to group g,

$$\widetilde{E}_{i,t}^{(g)} \equiv \sum_{j:\kappa(j)=g} e_{i,t}^{(j)}. \tag{A2}$$

B.1 Layer 1: k-Means on Stayers' Wage Growth Distributions

Building on the observation above that stayers' residualized wage growth $\Delta \ln \tilde{y}_{i,t}$ is an observable signal of $(\beta_i, \phi_{j(i,t)})$, Layer 1 produces an initial partition of workers and firms by clustering on the empirical distribution of this signal. For each firm j, we compute the cumulative distribution function and mean of $\Delta \ln \tilde{y}_{i,t}$ across the firm's stayer observations and apply k-means with $k = K_f$ to these firm-level distributions, yielding firm groups $\hat{\kappa}^{(0)}$.

For each worker i, we compute the cumulative distribution function and mean across the worker's stayer spells and apply k-means with $k = K_w$ to obtain worker groups $\hat{\rho}^{(0)}$. Clustering on the full distribution rather than a single summary statistic uses the shape of the wage-growth distribution—heavy tails, skewness, dispersion—as additional fingerprints of the underlying growth type, which is particularly valuable for workers or firms observed in few stayer spells. Workers or firms without stayer observations (e.g., always-movers) are assigned the modal group.

B.2 Layer 2: First-Difference Coordinate Descent

Algorithm 1 states Layer 2 formally. Beginning from the Layer-1 groups, the procedure alternates between OLS parameter updates and unit reassignment at the group level on the FD equation.

Algorithm 1 Layer 2: FD Coordinate Descent Refinement

Require: Stayer data; Layer-1 groups $\hat{\rho}^{(1)}, \hat{\kappa}^{(1)}$; tolerance $\tau_2 = 10^{-8}$.
Ensure: Refined groups $\hat{\rho}^{(2)}, \hat{\kappa}^{(2)}$; group-level coefficients $\hat{\beta}_g, \hat{\phi}_g$.

Initialize $\rho \leftarrow \hat{\rho}^{(1)},\ \kappa \leftarrow \hat{\kappa}^{(1)}$.
repeat
  OLS step: build $X = [\, D_w(\rho) \mid D_f(\kappa)[:, : -1] \,]$ with $[D_w(\rho)]_{t,g} = \mathbf{1}[\rho(w_t) = g]$ and solve $\hat{b} = \arg\min_b \|\Delta \ln y - Xb\|^2$ to obtain $(\hat{\beta}_g, \hat{\phi}_g)$; compute $\text{SSR} = \sum_t r_t^2$.
  Worker reassignment: for each worker i,

$$\hat{\rho}(i) \leftarrow \arg \min_{g' \in \{1,\dots,K_w\}} \sum_{t \in \mathcal{S}(i)} \left[\Delta \ln y_t - \hat{\beta}_{g'} - \hat{\phi}_{\kappa(f_t)}\right]^2,$$

  where $\mathcal{S}(i)$ is the set of stayer observations for worker i; workers with $|\mathcal{S}(i)| = 0$ retain their Layer-1 assignment.
  Firm reassignment: analogous minimization over $g' \in \{1, \dots, K_f\}$.
until $|\text{SSR}^{(\text{new})} - \text{SSR}^{(\text{prev})}| < \tau_2$.
return $\hat{\rho}^{(2)}, \hat{\kappa}^{(2)}, \hat{\beta}, \hat{\phi}$.

Layer 3 operates on the full levels equation. Its design matrix is

$$X = \Big[\ \underbrace{D_{\text{worker}}}_{N_T \times N}\ \Big|\ \underbrace{D_{\text{firm}} S}_{N_T \times (J-1)}\ \Big|\ \underbrace{\text{Exp} \odot D_w(\rho)}_{N_T \times K_w}\ \Big|\ \underbrace{\widetilde{E}^{(1)}, \dots, \widetilde{E}^{(K_f)}}_{N_T \times K_f}\ \Big], \tag{A3}$$

where S drops the last firm column for identification, $[\text{Exp} \odot D_w(\rho)]_{t,g} = \text{Exp}_{w_t,t} \cdot \mathbf{1}[\rho(w_t) = g]$, and $\widetilde{E}^{(g)}$ is the column vector defined above.

The parameter vector is $b = (\alpha_1, \dots, \alpha_N, \psi_1, \dots, \psi_{J-1}, \beta_1, \dots, \beta_{K_w}, \phi_1, \dots, \phi_{K_f})$, and the predicted log wage for observation t at worker w_t and firm f_t is $\widehat{\ln y_t} = \hat{\alpha}_{w_t} + \hat{\psi}_{f_t} + \hat{\beta}_{\rho(w_t)} \text{Exp}_{w_t,t} + \sum_g \hat{\phi}_g \widetilde{E}_t^{(g)}$, with residual $r_t = \ln y_t - \widehat{\ln y_t}$.

B.3 Layer 3: Levels-Equation Coordinate Descent

Algorithm 2 states Layer 3 formally. Beginning from the refined groups delivered by Layer 2, the procedure performs a single warm-started coordinate-descent pass on the full levels equation. Worker reassignment is performed in batch because each worker's optimal assignment is independent of every other worker given the current parameters and firm assignments. Firm reassignment, by contrast, is performed sequentially: moving firm j from group old to candidate g' changes $\widetilde{E}_t^{(\text{old})}$ and $\widetilde{E}_t^{(g')}$ for every observation with $e_t^{(j)} > 0$, and this cross-worker spillover makes batch reassignment intractable.

Algorithm 2 Layer 3: Levels-Equation Coordinate Descent (Single Warm-Started Pass)

Require: Panel data; Layer-2 groups $\hat{\rho}^{(2)}, \hat{\kappa}^{(2)}$; tolerance $\tau_3 = 10^{-6}$.
Ensure: Final estimates $\hat{\alpha}_i, \hat{\psi}_j, \hat{\beta}_g, \hat{\phi}_g$ and group assignments.

1: Initialize $\rho \leftarrow \hat{\rho}^{(2)},\ \kappa \leftarrow \hat{\kappa}^{(2)}$.
2: **repeat**
3: **PCG solve:** construct the levels design matrix X and solve $(X^\top X)\, b = X^\top \ln y$ by preconditioned conjugate gradient (PCG) with an incomplete Cholesky preconditioner.
4: Compute residuals $r_t = \ln y_t - \widehat{\ln y_t}$.
5: **Worker reassignment (batch):** for each worker i, the counterfactual residual is $r_t^{(g')} = r_t + (\hat{\beta}_{\rho_{\text{cur}}(w_t)} - \hat{\beta}_{g'})\,\text{Exp}_t$, and $\hat{\rho}(i) \leftarrow \arg\min_{g'} \sum_{t \in \mathcal{T}(i)} [r_t^{(g')}]^2$. All workers are reassigned simultaneously.
6: **Firm reassignment (sequential):** cumulative firm-group experience $\widetilde{E}_{i,t}^{(g)}$ couples firms across workers, so reassign firms *one at a time*, starting from the largest, with residuals and $\widetilde{E}$ updated after each move.
7: **until** $|\text{SSR}^{(\text{new})} - \text{SSR}^{(\text{prev})}| < \tau_3$.
8: Run one final PCG solve at the converged groups to extract $(\hat{\alpha}, \hat{\psi}, \hat{\beta}, \hat{\phi})$.
9: **return** all estimates plus group labels.

C Algorithm Validation and Performance

This section validates our two-way clustering algorithm through Monte Carlo simulations. We first describe the data-generating process, then report results from the three-layer algorithm in Section C.2. We begin by verifying that the algorithm can exactly recover all parameters of the wage equation in the absence of idiosyncratic noise, then evaluate how its performance degrades as the noise level increases. To isolate the contribution of each stage of the algorithm, we compare the full algorithm against two alternatives: bypassing the iterative refinement step or relying only on the main algorithm with random starting points.

C.1 Monte Carlo Experiments

We validate our algorithm through extensive Monte Carlo simulations that demonstrate its ability to recover all parameters of the wage equation accurately. We simulate wage data from the two-way growth fixed effects model of equation (3). In the data-generating process, each worker i draws a growth type $\rho(i) \in \{1, \dots, 20\}$ that determines its return to experience $\beta_{\rho(i)}$, and each firm j draws a growth type $\kappa(j) \in \{1, \dots, 20\}$ that determines its return to experience $\phi_{\kappa(j)}$, so that log wages take the form

$$\ln y_{i,t} = \alpha_i + \psi_{j(i,t)} + \beta_{\rho(i)} \operatorname{Exp}_{i,t} + \sum_{\kappa(j)} \phi_{\kappa(j)} \operatorname{Exp}_{i,t}^{\kappa(j)} + \epsilon_{i,t} \tag{A4}$$

where α_i and $\psi_{j(i,t)}$ are worker and firm fixed effects, $\operatorname{Exp}_{i,t}$ is labor-market experience, $\operatorname{Exp}_{i,t}^{\kappa(j)}$ is experience accumulated at firms of type $\kappa(j)$, and $\epsilon_{i,t}$ is an idiosyncratic shock. The 20 worker growth coefficients $\{\beta_1, \dots, \beta_{20}\}$ and 20 firm growth coefficients $\{\phi_1, \dots, \phi_{20}\}$ are set at the quantile midpoints of underlying normal distributions, so the DGP features exactly the discrete-type structure targeted by our clustering estimator. We then apply the algorithm to jointly classify workers and firms into growth types and estimate the group-specific coefficients $\beta_{\rho(i)}$ and $\phi_{\kappa(j)}$, allowing us to evaluate whether the algorithm correctly recovers the latent grouping structure and whether the estimated variance decompositions accurately reflect the true components over the life cycle.

Table A-2 summarizes our simulation design. The simulated dataset contains 8,000 workers employed at 800 firms observed over 40 periods.

C.2 Monte Carlo Results

Exact recovery without noise. We begin with a transparent check of the algorithm's internal consistency. When we set the variance of the idiosyncratic shocks to zero ($\sigma_\varepsilon = 0$), the algorithm exactly recovers all four parameters of the wage equation: worker fixed effects α_i, firm fixed effects ψ_j, and the worker and firm growth coefficients $\beta_{\rho(i)}$ and $\phi_{\kappa(j)}$, with correlations of one and variance ratios of one across all Monte Carlo replications. This result confirms that the algorithm can identify the true grouping structure and the associated parameters when the only source of variation in wages is the systematic component it is designed to recover.

Robustness to idiosyncratic noise. The more demanding test is whether the algorithm continues to recover the truth as the magnitude of the idiosyncratic shocks increases. Figure A-2 reports the correlation between true and estimated parameters for $\hat{\alpha}_i$, $\hat{\psi}_j$, $\hat{\beta}_{\rho(i)}$, and $\hat{\phi}_{\kappa(j)}$ under

Table A-2: Monte Carlo Simulation Design

Parameter	Value
Number of Workers ($\|\mathcal{I}\|$)	8,000
Number of Firms ($\|\mathcal{J}\|$)	800
Time periods ($\|\mathcal{T}\|$)	40
Number of worker / firm types (K_w, K_f)	20 / 20
Job-switching probability per period	0.10
Worker fixed effects (α_i)	$\alpha_i \sim N(0,\, 0.5^2)$
Firm fixed effects (ψ_j)	$\psi_j \sim N(0,\, 0.5^2)$
Worker growth coefficient ($\beta_{\rho(i)}$)	$N(0.05,\, 0.025^2)$
Firm growth coefficient ($\phi_{\kappa(j)}$)	$N(0.05,\, 0.025^2)$
Idiosyncratic shocks ($\varepsilon_{i,t}$)	$\varepsilon_{i,t} \sim N(0,\, \sigma_\varepsilon^2)$

Notes: Parameters of the discrete-types data-generating process used in the Monte Carlo. Workers and firms are matched at random and workers switch jobs with a fixed per-period probability. Worker and firm growth coefficients are drawn at $K_w = 20$ and $K_f = 20$ quantile midpoints of the indicated normal distributions, so that each worker (firm) belongs to one of 20 discrete growth types. The noise level σ_ε is varied across simulations.

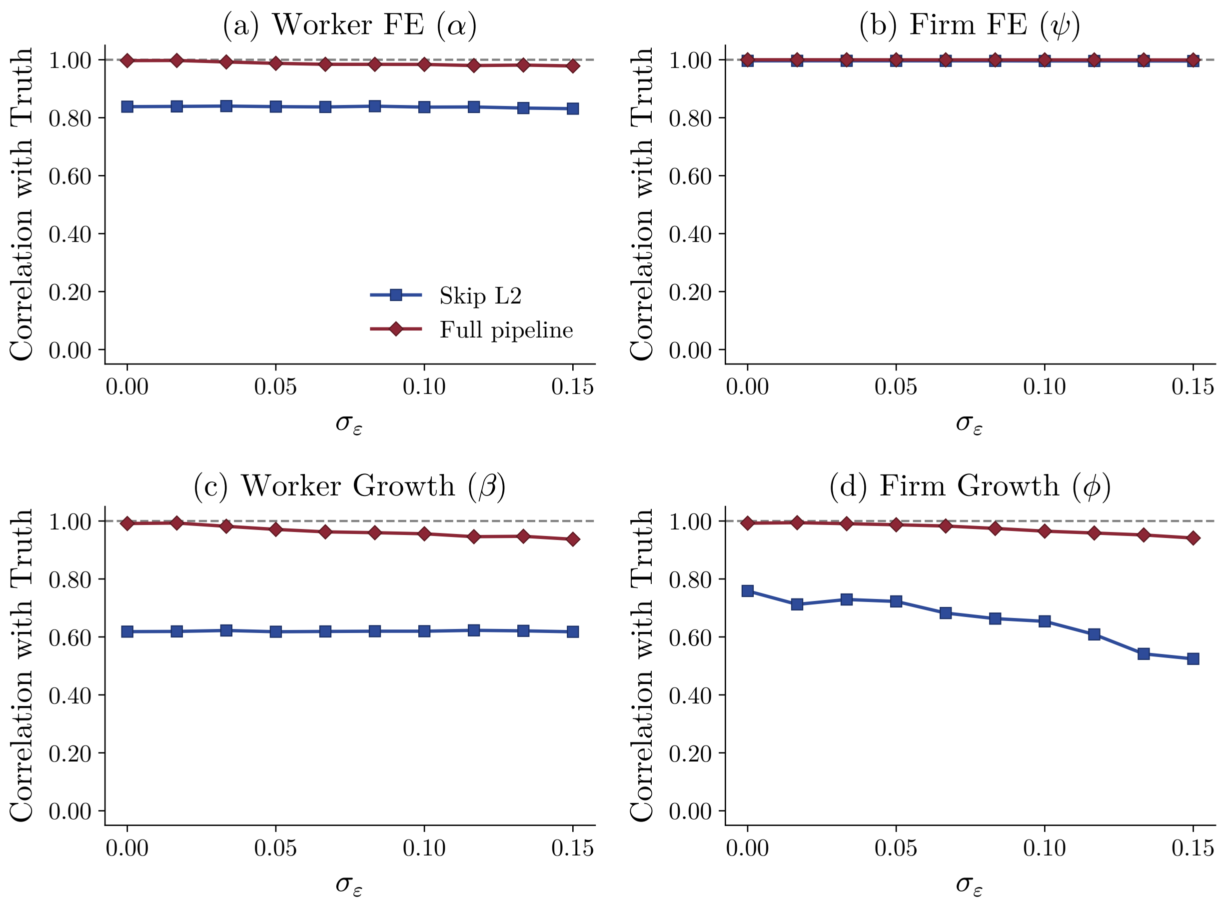
the full pipeline across Monte Carlo replications with increasing σ_ε. Correlations remain close to one for all four parameters. Figure A-3 plots the corresponding ratio of estimated to true variance for each component. Under the full pipeline, variance ratios cluster tightly around one for all four parameters, with only a slight upward drift on the worker side (within a few percent of one) at higher noise levels.

C.2.1 Value of Iterative Refinement

A natural question is whether the iterative reclassification stage of the main algorithm is necessary, or whether the initial grouping from k-means already provides an adequate partition. To isolate this margin, we construct a k-means-only estimator that takes the initial worker and firm groups formed from the distributional k-means clustering of stayers' wage growth in Initialization Step 1, plugs them into the full levels specification, runs OLS, and stops, bypassing iterative refinement entirely.

The k-means-only series, labeled as "skip L2", in Figures A-2 and A-3 reveals the cost of frozen group assignments. The full pipeline dominates uniformly, with the largest gains appearing on $\hat{\beta}_{\rho(i)}$ and $\hat{\phi}_{\kappa(j)}$, precisely the components for which grouping accuracy matters most. When skipping Layer 2, correlations with the truth fall to roughly 0.5–0.6 for the growth coefficients as idiosyncratic noise rises, compared to values close to one under the full pipeline. The variance-ratio gaps are even starker: the k-means-only estimator captures only 10–25% of the true dispersion in firm growth coefficients and 72–85% in worker growth coefficients, and also under-

Figure A-2: Parameter Recovery: Correlations

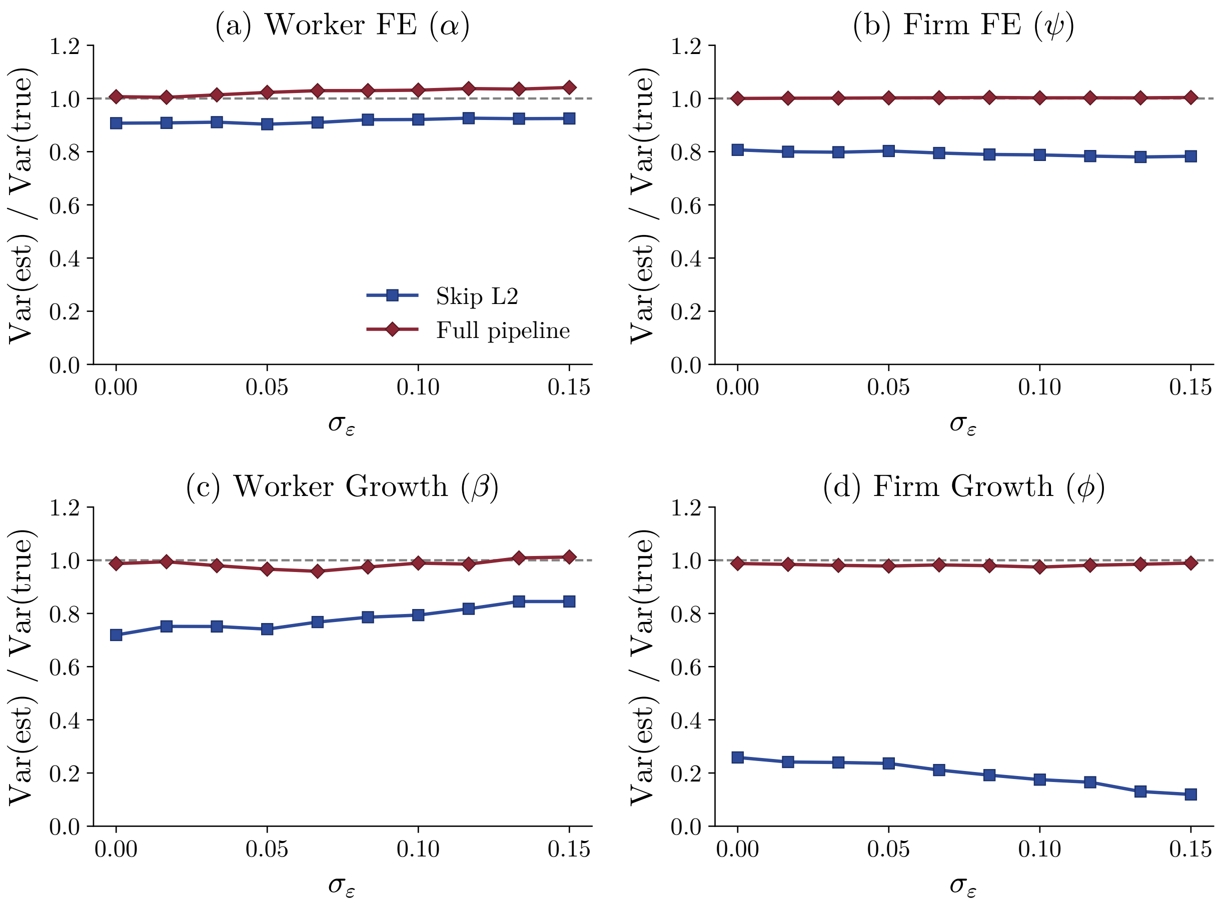


Notes: The figure plots the correlation between true and estimated parameters across Monte Carlo replications, separately for $\hat{\alpha}_i$, $\hat{\psi}_j$, $\hat{\beta}_{\rho(i)}$, and $\hat{\phi}_{\kappa(j)}$, comparing the full pipeline and the algorithm without iterative reclassification.

recovers the variance of both fixed effects. Under the full pipeline, by contrast, variance ratios cluster tightly around one for all four parameters. The pattern reflects the cost of frozen group assignments: the initial k-means partition contains classification errors that propagate into the levels regression when no further refinement is permitted.

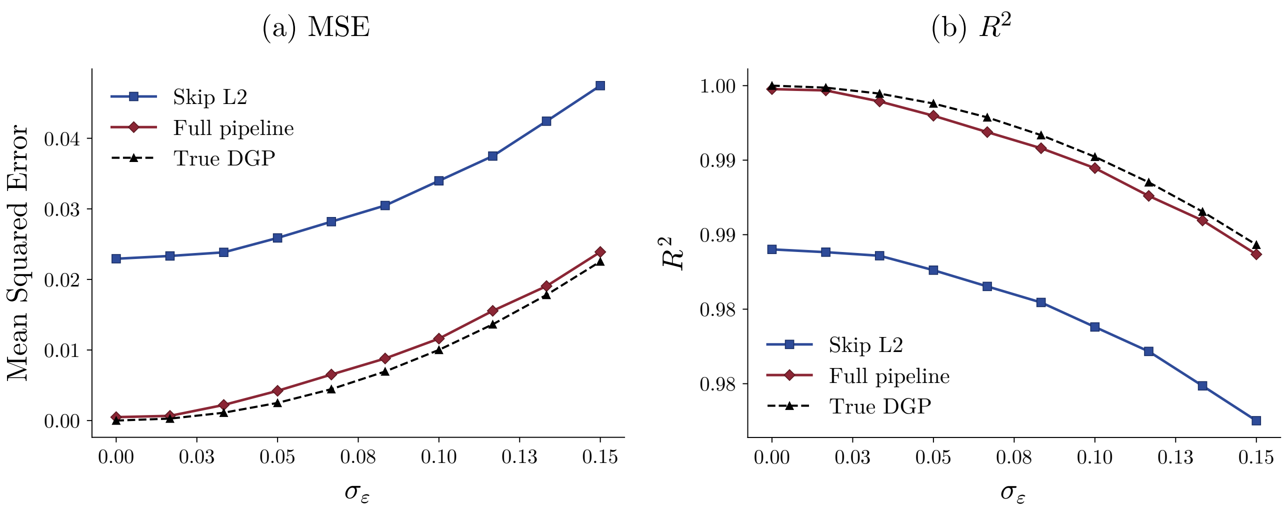
Figure A-4 reports the in-sample mean squared error and the R^2 of the estimated wage equation across Monte Carlo replications. Under the full pipeline, the MSE is concentrated near the lower bound implied by the variance of the idiosyncratic shocks, and the R^2 is concentrated near the share of total variance attributable to systematic components in the data-generating process; skipping iterative refinement, both diagnostics are visibly worse. Together with the correlation and variance-ratio results, this provides the justification for the reclassification step: the full pipeline yields both better fit and parameters that are quantitatively closer to the true DGP.

Figure A-3: Parameter Recovery: Variance Ratios



Notes: The figure plots the ratio of estimated to true variance across Monte Carlo replications for $\hat{\alpha}_i$, $\hat{\psi}_j$, $\hat{\beta}_{\rho(i)}$, and $\hat{\phi}_{\kappa(j)}$, comparing full pipeline and the algorithm without iterative reclassification. A ratio of one indicates that the algorithm recovers the true variance exactly.

Figure A-4: Overall Fit: MSE and R^2



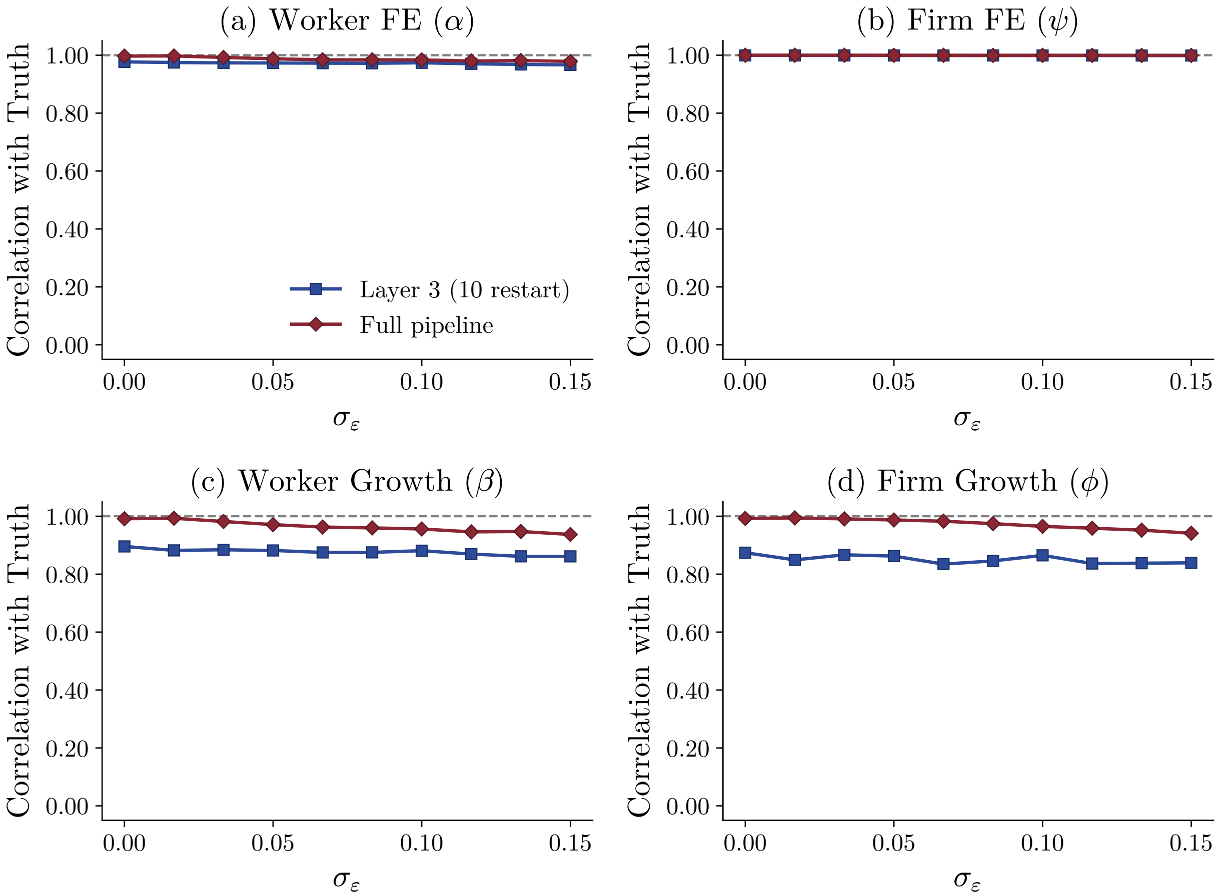
Notes: Panel (a) plots the in-sample mean squared error of the estimated wage equation across Monte Carlo replications, comparing the full pipeline against the k-means-only estimator. Panel (b) plots the corresponding in-sample R^2.

C.2.2 Value of Initialization

Even granting the importance of iterative reclassification, one might ask whether the warm start provided by the two initialization steps matters, or whether running the main algorithm from random starting points would suffice given enough restarts. To test this, we construct a 10-restart benchmark that runs the iterative search algorithm from 10 independent random starting points, bypassing the initialization steps entirely.

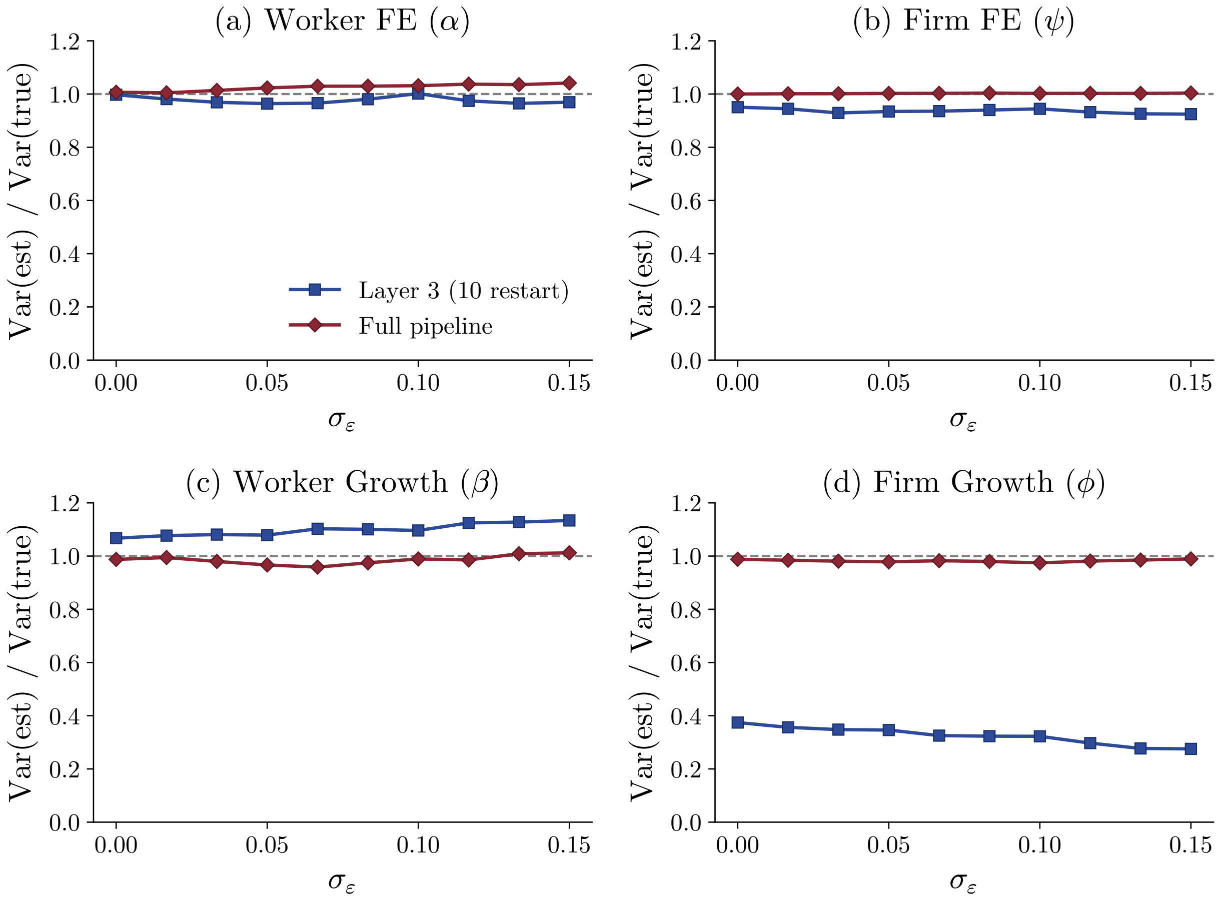
Figures A-5 and A-6 compare the 10-restart benchmark, labeled as "Layer 3 (10 restart)", against the full pipeline. The degradation is qualitatively similar to that of the k-means-only estimator but milder along several margins. Correlations for the fixed effects remain near one, while correlations for the growth components fall modestly, to roughly 0.84–0.90, compared to values close to one under the full pipeline. The variance of the firm growth coefficient is again severely compressed: the 10-restart benchmark captures only 30% of the true dispersion in $\hat{\phi}_{\kappa(j)}$, while over-estimating the dispersion in $\hat{\beta}_{\rho(i)}$ by roughly 5–10%. The 10-restart estimator does iterate, but from uninformed starting points, it converges to local optima that our two-step initialization is designed to avoid.

Figure A-5: Value of Initialization: Correlations



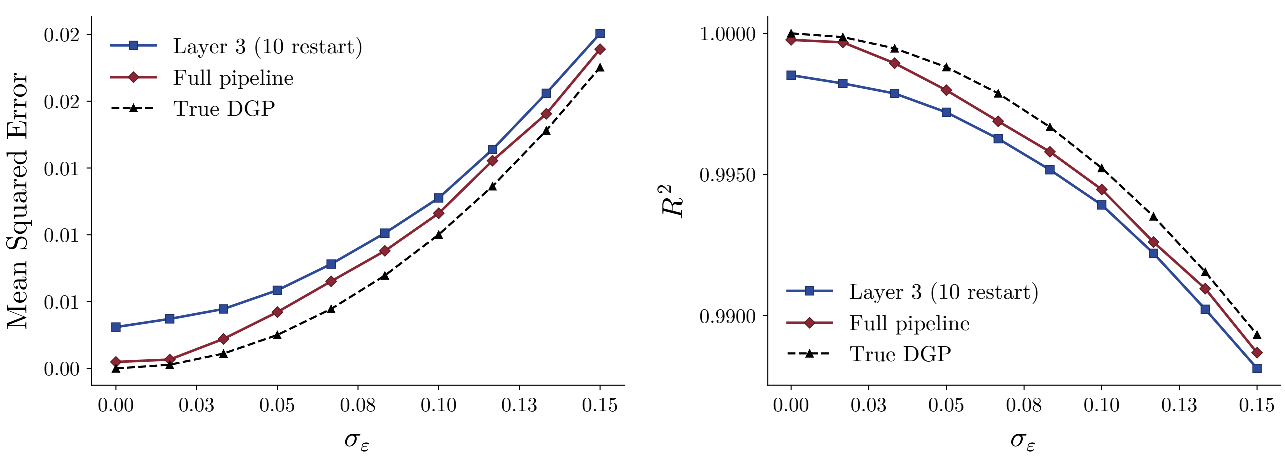
Notes: The figure plots the correlation between true and estimated parameters across Monte Carlo replications, separately for $\hat{\alpha}_i$, $\hat{\psi}_j$, $\hat{\beta}_{\rho(i)}$, and $\hat{\phi}_{\kappa(j)}$, comparing the full pipeline against the 10-restart benchmark.

Figure A-6: Value of Initialization: Variance Ratios



Notes: The figure plots the ratio of estimated to true variance across Monte Carlo replications for $\hat{\alpha}_i$, $\hat{\psi}_j$, $\hat{\beta}_{\rho(i)}$, and $\hat{\phi}_{\kappa(j)}$, comparing the full pipeline against the 10-restart benchmark. A ratio of one indicates that the algorithm recovers the true variance exactly.

Figure A-7: Overall Fit Across Estimators: MSE and R^2



(a) MSE

(b) R^2

Notes: Panel (a) plots the in-sample mean squared error for the full pipeline, and the 10-restart benchmark; the dashed line marks the True DGP benchmark (lower bound set by the variance of the idiosyncratic shocks). Panel (b) plots the corresponding in-sample R^2, with the dashed line marking the True DGP R^2.

Figure A-7 summarizes the in-sample mean squared error and R^2. Under the full pipeline, the MSE tracks the lower bound implied by the variance of the idiosyncratic shocks and the R^2 tracks the share of total variance attributable to systematic components in the DGP; both diagnostics are visibly worse under the two alternatives. Together with the evidence above, these results establish that the k-means initialization from wage growth, refinement by coordinate descent on the first-difference equation, and convergence via the main iterative search algorithm on the full levels equation jointly recover the latent grouping structure and the parameters of the wage equation and that each stage of the pipeline contributes to this recovery.

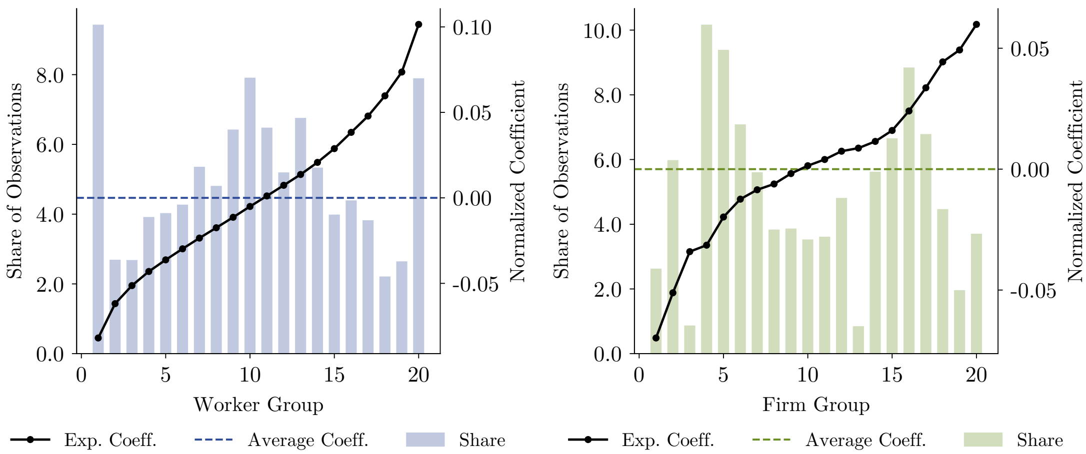
D Robustness to the Number of Worker and Firm Groups

This sections reports robustness of the main empirical results to alternative choices of the number of worker and firm groups, K_w and K_f. Our baseline specification sets $K_w = K_f = 20$. We re-run our algorithm under three alternative specifications: $(K_w, K_f) = (20, 50)$, $(50, 20)$, $(50, 50)$. Our results are qualitatively and quantitatively robust across all specifications.

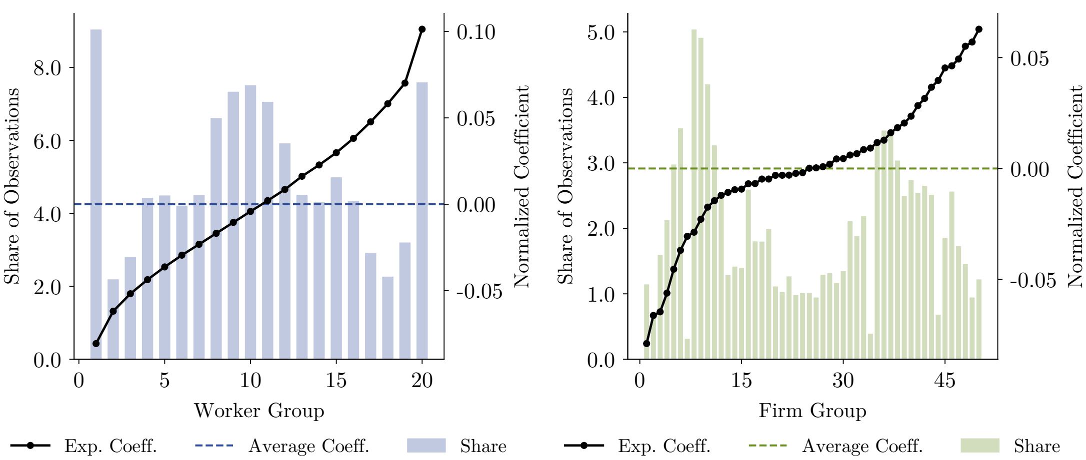
Figure A-8 reports the estimated returns to experience across worker and firm growth groups under all four specifications. The shape, range, and ordering of the worker- and firm-side coefficients are largely preserved as the number of groups varies.

Figure A-9 overlays the life-cycle variance decomposition across all four specifications. The variance contributions of worker and firm fixed effects, worker and firm growth components, the covariance terms, and the residual all track each other closely, indicating that the main finding is not sensitive to the number of worker and firm groups.

Figure A-8: Returns to Experience Across Group Specifications

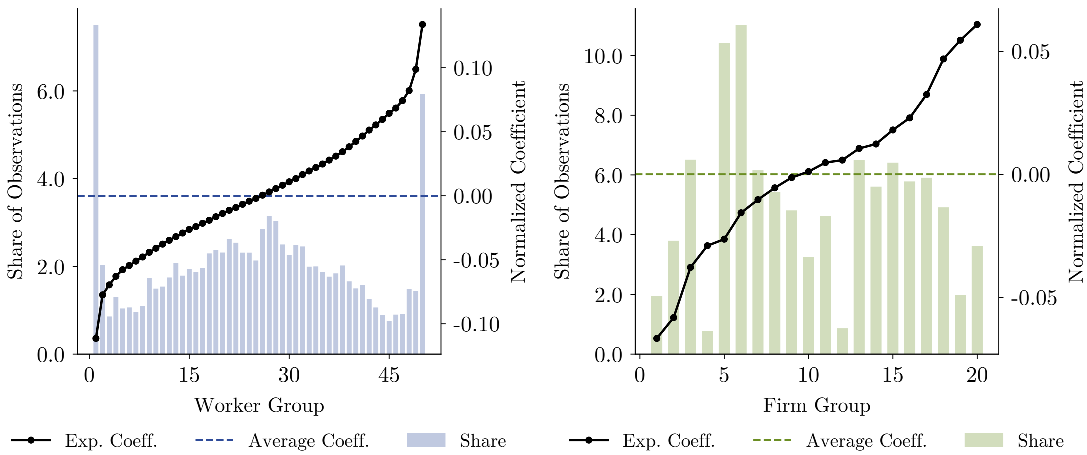


(a) $K_w = 20$, $K_f = 20$ (baseline)

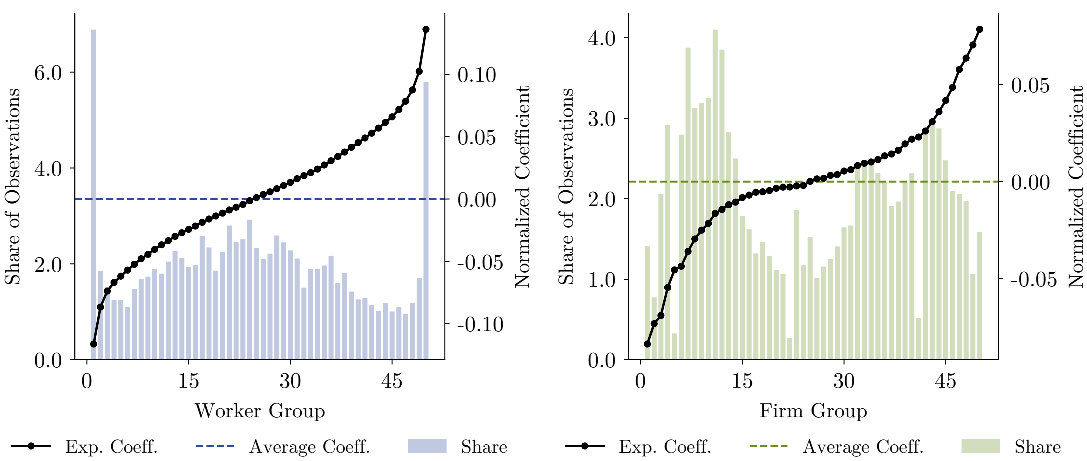


(b) $K_w = 20$, $K_f = 50$

Figure A-8: Returns to Experience Across Group Specifications (continued)



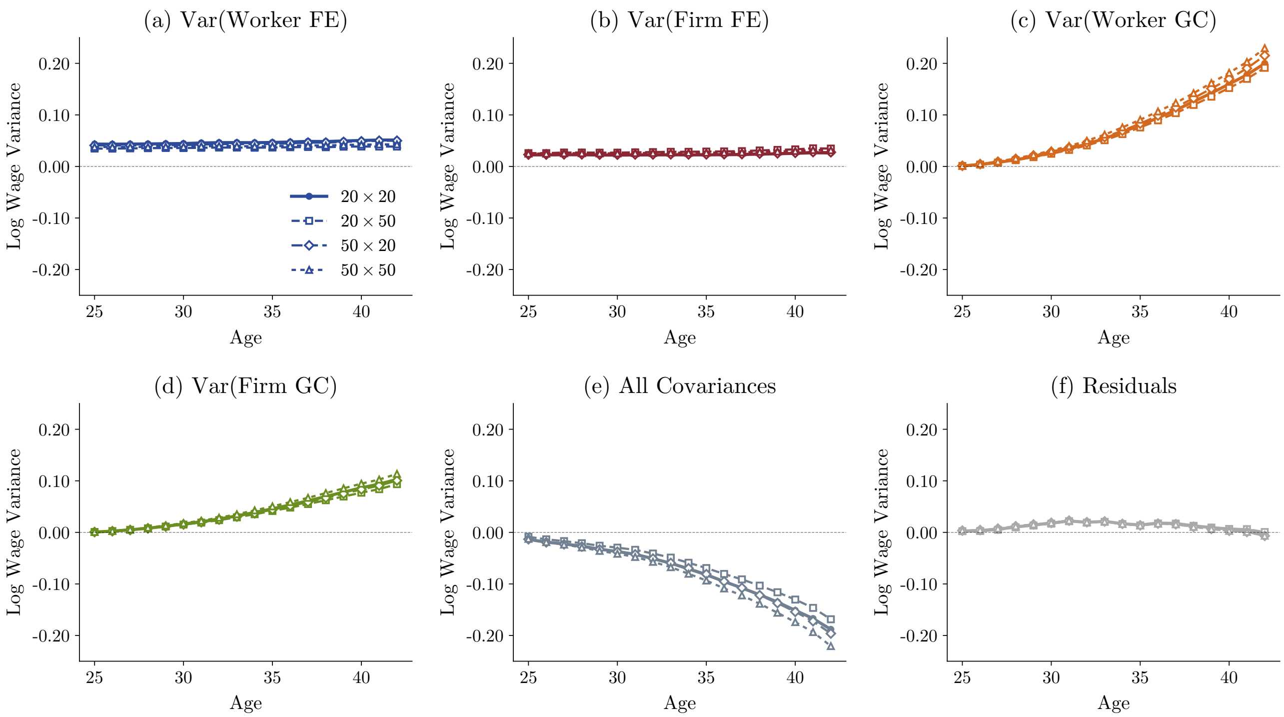
(c) $K_w = 50,\ K_f = 20$



(d) $K_w = 50,\ K_f = 50$

Notes: Estimated returns to experience by worker and firm growth groups under the four specifications $(K_w, K_f) \in \{(20,20), (20,50), (50,20), (50,50)\}$. The baseline specification used in the main text is $(20,20)$.

Figure A-9: Life-Cycle Variance Decomposition Across Group Specifications



Notes: Variance decomposition of log wages over the life cycle, overlaid across the four specifications $(K_w, K_f) \in \{(20, 20), (20, 50), (50, 20), (50, 50)\}$.