

Bias-Corrected Industry Premiums and Wage Floor Effectiveness: Evidence from a Rigid Labour Market

J. Jerónimo

September 3, 2026

1 Introduction

A central theme of empirical labour economics is that wages of observably similar workers vary systematically across firms and industries, creating persistent "pure" wage inequality beyond what human capital or compensating differentials can explain. Advanced methods such as the Abowd, Kramarz, and Margolis (1999; henceforth, AKM) decomposition have shown that firm effects alone explain a significant fraction of wage variance within and across countries. In pioneering work, Card, Heining and Kline (2013) document that in settings with high labour market regulation – such as Portugal and Germany – firm-specific factors remain important despite extensive wage compression from centralized bargaining and collectively set wage floors. However, recent methodological advances have revealed that standard AKM estimates suffer from substantial measurement bias when worker mobility is limited, raising questions about the reliability of existing cross-country and cross-industry comparisons (Andrews et al., 2008; Bonhomme, Lamadon and Manresa, 2023; Kline, Saggio and Sølvsten, 2020).

The magnitude of this limited mobility bias depends critically on two factors: the degree of

worker immobility and the variance in wages across firms (Barth et al., 2016). When workers rarely switch employers, their individual wage trajectories provide limited information to separately identify worker quality from firm wage-setting policies, leading measured firm effects to absorb components of worker heterogeneity. Recent evidence suggests this bias can be substantial: Bonhomme et al. (2023) show that uncorrected AKM estimates overstate firm wage variance by 20–50% in several developed economies, with bias magnitudes varying systematically across industries and institutional contexts. This heterogeneity in bias correction raises a natural question: what institutional features of labour markets moderate the severity of limited mobility bias? We argue that wage floors – legally mandated minimum wages binding at the industry and/or occupation level – represent a particularly important moderating factor. Floors affect both components of the bias mechanism: they compress residual wage variance directly (reducing the scope for bias) while potentially reducing job-to-job mobility (increasing the scope for bias). Which effect dominates remains an open empirical question.

Industry wage premiums provide an ideal setting to examine these issues. Krueger and Summers (1988; henceforth, KS88) pioneered this enquiry by documenting that, once individual characteristics are controlled for, industry affiliation explains a substantial portion of wage variance. Their analysis, based on linear wage equations with individual and industry controls, suggested that industry structure itself contributes to wage inequality, though methodological limitations left questions about causal identification. As summarized by Ferreira (2009), the literature remains divided on whether observed inter-industry wage premia represent true economic rents or whether they mask worker quality, sorting, or compensating differentials imperfectly measured in available longitudinal datasets. KS88 argue that if workers in high-wage industries are truly receiving economic rents, there should be a negative relationship between turnover and inter-industry wage differentials, as higher wages act as rents that incentivize job retention. This provides an empirical test of whether estimated

industry premiums represent "pure" wage differentials.

Recently, Card, Rothstein and Yi (2023; henceforth, CRY23) extended the AKM framework to measure industry wage premiums while accounting for worker and firm heterogeneity. They propose two definitions: (i) an employment-weighted average of firm effects within industries, and (ii) a shrinkage-adjusted measure reflecting the expected wage gain for a worker moving between industries, net of person effects. Both definitions reveal substantial dispersion in pay across sectors even after accounting for worker and firm heterogeneity. Critically, CRY23 emphasize that these industry premiums inherit the biases present in underlying firm effects when estimated via standard AKM. If limited mobility bias inflates measured firm heterogeneity, industry premiums constructed from these biased firm effects will similarly overstate true sectoral wage differences. This bias may vary systematically across industries depending on worker mobility patterns and institutional constraints – precisely the heterogeneity we seek to understand.

The Portuguese labour market offers an interesting setting for examining these issues. First, Portugal exhibits low worker mobility relative to other developed economies: 77.3% of workers never change employers during our sample period (2010-2019), and 22.7% switch firms at least once. Working with the full population of covered workers rather than a sample, however, we find that the resulting mobility network is well connected – the largest connected set covers 96.7% of person-year observations and the leave-one-out set 87.5% – so that limited mobility bias, while present, is not so severe as to preclude credible identification. This combination is itself useful: it allows us to measure the bias precisely rather than merely to acknowledge it. Second, Portugal features a comprehensive system of industry- and occupation-specific wage floors established through collective bargaining agreements and extended via *portarias de extensão*. Binding rates vary substantially across industries, from under 2% in parts of finance and information technology to roughly 37% in the most con-

strained sectors, with a cross-industry mean of 13.7%. This institutional variation is what allows us to examine how wage floor binding moderates the magnitude of limited mobility bias across industries. Third, despite this institutional wage compression, substantial industry wage premiums persist in Portugal (Ferreira, 2009; Card, Cardoso, Heining & Kline, 2018), with higher premiums systematically associated with lower turnover – a pattern consistent with economic rents rather than compensating differentials (Krueger & Summers, 1988). The mechanisms underlying this persistence in a highly regulated environment have not been definitively explained.

We make two contributions to this literature. First, we provide bias-corrected estimates of industry wage premiums for the full population of Portuguese private-sector workers, combining the leave-out estimator of Kline, Saggio and Sølvesten (2020) with the ground-up premium framework of Card, Rothstein and Yi (2024). Estimating the uncorrected, homoskedastic and heteroskedasticity-robust specifications on a single fixed sample, we find that firm effects account for 15.5% of wage variance before correction and 12.6% after, and that worker-firm sorting is positive throughout and strengthens under correction. Second, we document that wage floor binding rates – the share of workers earning within close proximity to the applicable sectoral minimum – systematically predict the severity of limited mobility bias across industries. Industries in which 20-40% of workers are floor-constrained exhibit bias gaps roughly 2.7 times larger than industries in which fewer than 20% are. We show that although floors compress residual wage variance substantially, as mean wage cushions fall by an order of magnitude across the observed range of binding, this compression is outweighed by the reduction in job-to-job mobility that floors induce. Floors therefore systematically amplify limited mobility bias rather than attenuating it.

Our findings carry implications for comparing measured wage inequality across institutional contexts. Because our evidence is cross-sectional, it speaks to comparisons across bargaining

regimes rather than to trends over time; the Portuguese national minimum wage rose in nominal terms throughout our sample period, so we do not address the consequences of floor erosion. What the cross-industry gradient implies is that institutional features of labour markets – wage floors, collective bargaining coverage, employment protection – should be treated as contextual factors affecting the reliability of matched employer-employee data for measuring the sources of wage inequality. Settings with more binding floors will tend to exhibit larger bias corrections even where the underlying economic mechanisms are unchanged, so cross-country differences in measured firm wage heterogeneity may partly reflect differences in the identifying variation that institutions permit. Researchers employing AKM methods across different time periods, regions, or institutional contexts should account for these moderating factors when interpreting bias correction magnitudes and comparing firm or industry wage effects across settings.

The remainder of the paper proceeds as follows. Section 2 describes the methodological background. Section 3 presents and discusses our data. Section 4 reports the results concerning bias-corrected industry premium estimates and wage floor binding dynamics. Section 5 concludes.

2 Methodological Background

The foundational approach to decomposing wages into worker and firm components is the now standard AKM. The canonical two-way fixed effects specification is:

$$w_{ijt} = \alpha_i + \psi_{j(i,t)} + X'_{it}\beta + \epsilon_{ijt} \quad (1)$$

where w_{ijt} denotes the log wage of worker i employed at firm j in year t , α_i is a time-invariant worker fixed effect capturing permanent differences in earnings capacity, $\psi_{j(i,t)}$ is a time-

invariant firm wage premium representing the causal wage effect of employment at firm j , X_{it} is a vector of time-varying covariates (age, calendar year), and ϵ_{ijt} is an idiosyncratic error term. This additive separability assumption implies that a worker’s wage at any firm is the sum of their portable skill α_i and the firm’s wage policy $\psi_{j(i,t)}$, plus observable characteristics and residual shocks.

The residual ϵ_{ijt} incorporates three distinct components that the model does not separately identify: (i) persistent match-specific effects between worker i and firm j (e.g. complementarities between worker skills and firm technology), (ii) transitory person-specific shocks (e.g., health, family circumstances), and (iii) transitory firm-wide shocks affecting all workers (e.g., demand fluctuations, hours adjustments). While extensions of the AKM model explicitly incorporate match effects (Woodcock, 2015; Bonhomme et al., 2023), we follow the standard approach and subsume these into the error term, recognizing this as a maintained assumption that simplifies estimation but requires empirical validation.

Unbiased estimation of $\{\alpha_i, \psi_j\}$ via ordinary least squares (OLS) requires the exogenous mobility assumption: the error term ϵ_{ijt} must be orthogonal to the complete employment history of worker i . Formally,

$$E[\epsilon_{ijt} | \{j(i, s)\}_{s=1}^{T_i}, X_i] = 0 \quad (2)$$

where $\{j(i, s)\}_{s=1}^{T_i}$ denotes the sequence of firms employing worker i over their observed career. This assumption rules out systematic sorting of workers to firms based on anticipated match quality or transitory shocks that trigger mobility decisions. Intuitively, Equation 4.2 requires that workers’ job transitions are unrelated to unobserved components of their wage equation, which is a strong assumption that can fail in multiple scenarios - e.g. Lopes de Melo (2018) documents coworker segregation (i.e. high-skill workers cluster in the same firms) which contradicts the AKM assumption of random assignment conditional on firm

type; Lamadon et al (2022) find that workers with high preference for amenities sort into low-wage, high amenity firms, creating negative correlation between observed firm effects and true wage premia; and Di Addario et al (2022) report that workers hired from unemployment earn different wages than those hired from other firms, implying that the hiring channel (i.e. the source of the worker) predicts wage growth, and mobility is not independent of ϵ_{ijt} .

Violations of exogenous mobility therefore arise when workers strategically move to firms offering higher expected ϵ_{ijt} (positive selection), or when firms selectively hire workers with lower expected ϵ_{ijt} (negative selection). For instance, if high-ability workers systematically sort into firms with complementary production technologies (generating positive match effects), then ϵ_{ijt} will be systematically correlated with the firm identifier $j(i, s)$, biasing estimates of ψ_j upward. Similarly, if workers experiencing positive transitory shocks (e.g., promotion opportunities, improved working conditions) are more likely to remain with their current employer, then observed movers will be negatively selected on ϵ_{ijt} , biasing firm effect estimates. For instance, Abowd, McKinney, and Schmutte (2018) show that match effects are systematically negatively correlated with worker and firm effects in U.S. data, while Borovičková and Shimer (2024) demonstrate that positive assortative matching generates selection bias that renders standard firm fixed effect estimates non-structural. In Portuguese administrative data, Card, Cardoso, and Kline (2016) reject exogenous mobility but conclude that the quantitative bias introduced is modest relative to the substantial firm heterogeneity they document, corroborating broader findings from Kline et al. (2020) that exogenous mobility violations are often economically small despite statistical detectability.

Card, Heining, and Kline (2013, hereafter CHK) propose a comprehensive battery of specification tests to assess the plausibility of (2). First, pre-move wage trends examine whether workers moving to high- ψ_j firms exhibit differential wage growth prior to the move (Ashenfelter’s dip test), which would indicate anticipatory sorting. Second, post-move residual

patterns test whether $\hat{\epsilon}_{ijt}$ systematically varies after job transitions, indicating unmodeled match effects or selective mobility. Third, symmetry of mover flows assesses whether workers moving from high- ψ_j to low- ψ_j firms exhibit mirror-image wage patterns to those moving in reverse, as would be expected under exogenous mobility.

Identification of (α_i, ψ_j) requires sufficient worker mobility to connect firms through chains of movers. The largest connected set (see panel B, Figure 4.4) comprises all firms linked directly or indirectly through worker transitions (Abowd et al., 2002). Workers and firms outside this set cannot be jointly identified - only differences $\alpha_i - \alpha_{i'}$ and $\psi_j - \psi_{j'}$ within connected components are estimable without normalization. In practice, the largest connected set typically includes the vast majority of observations in administrative datasets, but the sparsity of connections within this set governs the precision and bias of estimated effects (Bonhomme et al, 2023).

The structure of the mobility network has profound implications for identification quality. When few workers connect firms - a condition particularly relevant in rigid labour markets with strong employment protection - estimates suffer from *limited mobility bias*, a finite-sample bias that inflates variance estimates and distorts the measured relationship between worker and firm effects (Andrews et al., 2008). This bias arises because the AKM model is overparameterized relative to the information contained in sparse mobility networks: with millions of worker and firm fixed effects but limited job-to-job transitions, the model struggles to separately identify permanent worker quality from firm wage premiums.

Limited mobility bias emerges when the number of workers connecting firms is insufficient relative to the dimensionality of the fixed effects space. Let $G = (V, E)$ denote the mobility network, where V is the set of firms and $(j, j') \in E$ if at least one worker transitions from firm j to firm j' . Identification of $\{\psi_j\}$ requires G to be connected, but the degree of connectivity

– measured by the average number of transition paths linking firms – governs estimation precision. The bias magnitude for firm j with M_j movers is (Andrews et al, 2008; Kline et al, 2020):

$$Bias(\hat{\psi}_j) \approx \frac{\sigma_\alpha^2}{M_j} \times \rho_{jj} \quad (3)$$

where $\sigma_\alpha^2 = var(\alpha_i)$ and ρ_{jj} is the leverage of firm j in the design matrix (analogous to hat-matrix diagonal elements in OLS). Firms with few movers have high ρ_{jj} , amplifying bias. Intuitively, when only a small number of workers connect firm j to the rest of the network, the model has limited information to separately identify worker quality α_i from the firm premium ψ_j – any persistent wage differences could be attributed to either component, leading to incidental parameter bias.

This bias is not asymptotically eliminated as the sample size grows. Adding more person-year observations for existing workers and firms does not resolve the fundamental problem of sparse mobility. The bias vanishes only if the number of movers M_j grows proportionally with the number of firms, a condition rarely satisfied in administrative data where most workers have short job tenure and limited mobility (Kline et al., 2020).

Following Kline et al (2020) and Bonhomme et al (2023), we employ the leave-one-out (LOO) correction to mitigate limited mobility bias. The LOO estimator computes firm effects for firm j excluding observations from worker i :

$$\tilde{\psi}_{j,-i} = \frac{1}{N_j - n_{ij}} \sum_{i' \neq i, t} (w_{i't} - \hat{\alpha}_{i'} - X'_{i't} \hat{\beta}) \quad (4)$$

where $N_j = \sum_{i,t} 1(j(i,t) = j)$ is total employment at firm j and n_{ij} is the number of observations for worker i at firm j . The bias-corrected estimate is:

$$\psi_j^{LOO} = \psi_j - \frac{cov(\alpha_i, \psi_j | j(i, t) = j)}{N_j} \quad (5)$$

The intuition behind the correction is the following: uncorrected $\hat{\psi}_j$ overestimates firm effects when high- α_i workers sort into high- ψ_j firms (positive assortative matching). The spurious covariance $cov(\alpha_i, \psi_j | j)$ arises mechanically from limited mobility – when few workers connect firm j , any correlation between observed wages and worker types is attributed to the firm effect. The LOO correction removes this incidental parameter bias by computing firm effects on a subsample excluding each worker, then averaging across leave-out estimates weighted by sampling variance (the jackknife leverage adjustment of Kline et al., 2020).

The correction is particularly important when average firm tenure is short and mobility rates are low – precisely the conditions characterizing Portugal’s rigid labour market. In our data, only 3.49% of workers switch industries over the sample period, with a clustering coefficient of 0.186 (see the data section for connectivity diagnostics), indicating sparse and fragmented mobility networks. Under these conditions, uncorrected AKM estimates are upward biased and with abnormally large variances, rendering inference about true firm heterogeneity unreliable.

Identification of industry wage premiums depends on the structure of job-to-job transitions connecting firms. While the leave-one-out correction mitigates incidental parameter bias in firm effects, estimation precision varies with connectivity: firms linked through numerous worker transitions receive stable estimates, while isolated firms suffer from scarce identification. We assess identification quality using standard graph-theoretic measures of network connectivity. The connected set coverage, $C = \frac{N_{\text{connected}}}{N_{\text{total}}}$, measures the fraction of observations in the largest connected component; high coverage ensures broad identification but does not guarantee precision in sparsely connected regions. The average path length, $\ell = \frac{1}{|V|(|V|-1)} \sum_{j \neq j'} d(j, j')$, where $d(j, j')$ denotes the shortest path between firms j and j' in

the mobility network, captures overall tightness of connectivity - lower values indicate multiple independent paths linking firms, providing richer information for identification. The clustering coefficient, $CC = \frac{3 \cdot \#\{\Delta\}}{\#\{\text{triples}\}}$, measures the degree to which firms form industry-specific clusters; high clustering indicates that worker transitions concentrate within sectors, potentially limiting cross-sector identification of relative industry premiums. At the firm level, firm-specific mobility, $m_j = \frac{N_j^{\text{unique}}}{N_j}$, captures the proportion of employment spells at firm j that involve worker mobility into or out of the firm, with low m_j indicating stable, long-tenure employment that exacerbates limited mobility bias.

These network-level diagnostics directly inform the identification of industry wage premiums aggregated from corrected firm effects. Industries characterized by high within-industry mobility, $r_s^{\text{within}} = \frac{\# \text{ firm switches within industry } s}{\# \text{ total observations in } s}$ benefit from rich information about wage heterogeneity independent of sector-wide premiums, allowing reliable estimation of hierarchy effects (within-industry firm heterogeneity). Conversely, industries with low cross-industry mobility $r_s^{\text{cross}} = \frac{\# \text{ industry switches from industry } s}{\# \text{ total observations in } s}$ have relative premiums that are weakly identified compared to the reference industry, since few workers provide direct wage comparisons spanning sectors. In rigid labour markets with stringent employment protection and limited worker mobility - such as Portugal - we expect sparse networks with large average path lengths and low cross-sector mobility, conditions that amplify limited mobility bias but make bias-correction through leave-one-out estimation particularly valuable. Empirically, the ratio $r_s^{\text{within}}/r_s^{\text{cross}}$ captures the degree of internal fragmentation in the labour market: a high ratio indicates workers move predominantly within their current industry, creating multiple mini-networks per sector that are loosely connected across sectors. In the following section we provide detailed diagnostics of these network properties in Portuguese data and show that, despite connectivity constraints characteristic of rigid labour markets, the largest connected set comprises sufficient observations with adequate cross-sector linking to enable reliable industry premium identification after bias-correction.

2.1 From Firm Effects to Industry Premiums: The Ground-Up Approach

Let $s(j)$ denote the industry affiliation of firm j . Following Card, Rothstein, and Yi (2023), we define the industry wage premium as the employment-weighted average of firm wage premiums within industry s :

$$\bar{\psi}_s = \frac{\sum_{j \in s} N_j \psi_j}{\sum_{j \in s} N_j} \quad (6)$$

where N_j is total employment at firm j . This *ground-up measure* tells us what premium would a worker earn in a randomly selected job in industry s . Employment weighting ensures that the measure reflects the typical worker’s experience rather than giving equal weight to small and large firms – a crucial distinction given substantial firm size heterogeneity within industries.

Equation 4.6 correctly aggregates firm-level premiums to the industry level, accounting for within-industry heterogeneity in firm pay policies. It differs fundamentally from the two alternative approaches that have dominated previous literature: the first are cross-sectional regression (Krueger & Summers, 1988) where wages are regressed on worker characteristics and industry indicators. The coefficients of the industry indicators suffer from unobserved ability bias if high-skill workers systematically sort into high-premium industries. Without controlling for permanent worker heterogeneity α_i , cross-sectional estimates confound industry premiums with workforce composition. Card et al. (2023) show this approach overstates premiums by approximately 17% relative to ground-up measures. The second approach is the industry mover design (Murphy & Topel, 1990), that uses within-worker variation to identify industry effects. The coefficients obtained suffer from hierarchy effect attenuation

if movers systematically select firms with premiums deviating from the industry average. The theoretical mechanism is grounded in the Burdett-Mortensen on-the-job search model with wage posting (1998): a worker accepts a job offer when total compensation equals or exceeds the reservation utility from their current employment. For a worker i in industry s considering a move to industry s' with $\bar{\psi}_{s'} > \bar{\psi}_s$, the acceptance condition is:

$$w_{ij't} = \alpha_i + \bar{\psi}_{s'} + h_{j'} \geq w_{ijt} = \alpha_i + \bar{\psi}_s + h_j \quad (7)$$

where j' is the destination firm and j is the current firm. Simplifying,

$$\Delta h = h_{j'} - h_j \geq -(\bar{\psi}_{s'} - \bar{\psi}_s) \quad (8)$$

This condition reveals the selection mechanism underlying attenuation bias in industry-mover designs. Workers moving to high-premium industries ($\bar{\psi}_{s'} - \bar{\psi}_s > 0$) accept job offers even at below-average firms within those industries ($h_{j'} < 0$), since the industry premium offset compensates for lower firm-specific pay. Conversely, workers moving to low-premium industries ($\bar{\psi}_{s'} - \bar{\psi}_s < 0$) only accept positions at above-average firms within those industries ($h_{j'} > 0$), requiring a firm-specific compensating differential to overcome the industry penalty. This selection process induces systematic negative correlation between industry premium changes and hierarchy effect changes, Δh_j , among movers (i.e. $\text{Cov}(\Delta \bar{\psi}_s, \Delta h_j \mid \text{industry movers}) < 0$). In an industry mover regression omitting h_j , this unobserved negative correlation biases the estimated industry premium downward. Intuitively, when workers moving to high-premium sectors do so by accepting below-average firms, the measured wage gains from the sector switch are partially offset by the hierarchy loss, reducing the estimated industry effect.

In Portugal's labour market – which retains relatively rigid employment protection legislation despite post-2011 reforms (Turrini, 2011), and where centralized sectoral wage-setting insti-

tutions establish occupation-specific wage floors covering approximately 80-85% of workers – the measurement challenge is compounded by the interaction between these institutional constraints and the resulting sparse mobility networks. High severance payments for permanent contracts discourage voluntary job-to-job transitions, while extension mechanisms that universalize sectoral agreements reduce wage variation that might otherwise incentivize mobility. The combination produces particularly severe limited mobility bias: with only 3.49% of workers switching industries over multi-year periods and clustering coefficients near zero, Portugal’s restricted labour market fluidity creates the sparse worker-firm networks that amplify the incidental parameters problem in AKM estimation.

The wage floor system – negotiated annually at the sectoral level and legally binding across all firms in each industry – has complex effects on AKM identification quality. On one hand, wage floors compress residual wage variance at the bottom of the distribution by constraining firm wage-setting discretion, which marginally reduces the magnitude of the bias correction term in the Kline, Saggio, Sølvssten (2020) framework through reduced heteroskedasticity in the error structure. On the other hand, wage floors significantly reduce job-to-job mobility by compressing wage differentials between firms, thereby reducing workers’ incentives to switch employers (Liu 2021, Dube et al. 2016). This mobility reduction exacerbates limited mobility bias by creating sparser worker-firm networks with fewer movers available to identify firm effects (Andrews et al. 2008, 2012; Bonhomme et al. 2023). Additionally, when wage floors bind, observed wage changes for transitions involving floor-constrained workers contain less information about underlying firm quality differences, increasing the imprecision (though not necessarily the bias) of firm effect estimates. The net effect is that wage floors likely amplify limited mobility bias through their dominant effect on reducing labour market fluidity, despite any offsetting compression of residual variance. The severity of this problem depends on floor prevalence (what share of workers are at or near floors) and the distribution of movers across the wage distribution (whether job switchers are disproportionately located

in floor-binding or unconstrained wage regions). We discuss these elements further in the data description section.

3 Data

We use the matched employer-employee dataset *Quadros de Pessoal* to conduct our empirical analysis, containing data concerning Portuguese workers and firms in the private sector. We focus on the period between 2010 and 2019, as done by Card et al (2023) before us with US data. It is also the period in our data that ensures complete, granular industry-level indicators for identification in our empirical analysis. The data is gathered through an annual mandatory survey submitted by private sector firms every year in October. Some key variables in the data set are wages, hours of work, education, age and tenure at the individual level, and revenue, age and size at the firm level. All of our main analysis is done on the universe of private sector workers observed between 2010 and 2019 (Card & Cardoso, 2022). We then follow standard procedure in the literature by applying additional restrictions that are summarized in Table 4.1. In the end, we are left with 22,602,907 worker-year observations of working-age individuals working in full-time in job categories above that of apprenticeship. We exclude both apprentices and beginner practitioners because they are legally permitted to earn 80% of the sectoral wage floor under Portuguese labour law.

Table ?? reports key variable descriptives. The Portuguese labour market during 2010-2019 exhibits substantial rigidity characteristic of Southern European economies¹. Median worker tenure is 5 years with a mean of 8.1 years, indicating limited job mobility despite a large right tail ($P90 = 21$ years). The educational composition reflects a predominantly low-qualification workforce, with 54% of workers holding at most primary education, although 19% hold tertiary qualifications. Employment is spread across the firm size distribution, with 28% of workers in establishments of fewer than 20 employees and 31% in establishments of 250 or

¹See OECD (2015), *OECD Employment Outlook 2015*.

Table 1: Sample Construction and Restrictions

Restriction	Observations	Workers	Firms	Retained (%)
Raw data (all observations)	30,210,118	5,012,074	507,665	100.00
Age restriction (18–65)	30,210,118	5,012,074	507,665	100.00
Full-time restriction	26,529,186	4,604,062	459,560	87.82
Job title restriction	26,529,081	4,604,050	459,559	87.82
After removing missing key variables	25,826,191	4,565,826	458,217	85.49
Largest connected set	23,454,394	3,573,094	343,010	77.64
Leave-out connected set	22,624,115	3,452,128	238,553	74.89
After dropping duplicate worker-years	22,602,907	3,452,128	238,553	74.82
Final analysis sample	22,602,907	3,452,128	238,553	74.82

Notes: This table reports the progressive application of sample restrictions. Restrictions were applied to 100% of all Portuguese private sector workers observed between 2010 and 2019. Job title restriction requires at least 10 workers per occupation title. Sample excludes apprentices and beginner practitioners, who are legally permitted to earn 80% of the sectoral wage floor under Portuguese labour law. The final sample excludes observations with missing values in key variables (wages, education, or firm identifiers).

more. Wage inequality is moderate by international standards, with a 90/10 wage ratio of 3.60. Notably, the 50/10 ratio is only 1.48, indicating strong compression at the bottom of the distribution. This compression coexists with wage floors that are, on average, well below typical pay: the floor-to-median wage ratio averages 0.59 across industries, 87% of workers earn above the applicable floor, and 65% earn at least 20% above it (see Figures 4.1 and 4.2). Floors therefore bind for a substantial minority rather than for the bulk of the workforce, a point we return to when examining heterogeneity in bias correction across industries.

We classify a worker as floor-constrained if their wage is within 5% of the applicable floor ($wage/floor \leq 1.05$). This threshold accounts for three empirical realities. First, measurement error in administrative wage and floor data creates noise around the true binding point². Second, enforcement imperfections mean small violations are tolerated, making wages within 5% of the floor effectively constrained. Third, economic theory predicts that workers earning

²This is especially relevant in our analysis, as we approximate the wage floors set by collective bargaining agreements using modal wages at the job title-year level. We do this due to data limitations and because, as suggested by Card & Cardoso (2022), these modes are globally accurate.

marginally above the floor have limited bargaining power, as the floor serves as the effective reservation wage (Burdett and Mortensen, 1998). This definition follows Card, Cardoso, Heining and Kline (2018) using the same Portuguese administrative data, and is standard in the minimum wage literature (Dube et al., 2010). Those authors acknowledge that their data contain negative wage cushions, which they attribute to misreporting of base wages and to special circumstances such as extended sick leave. We follow them in retaining these observations: 12.9% of worker-years exhibit a negative cushion, of which the majority lie more than 10% below the floor. The cushion distribution is nonetheless firmly positive overall, with a mean of 0.399 and a median of 0.322 log points, indicating that floors are binding for a minority of workers and imperfectly enforced at the margin.

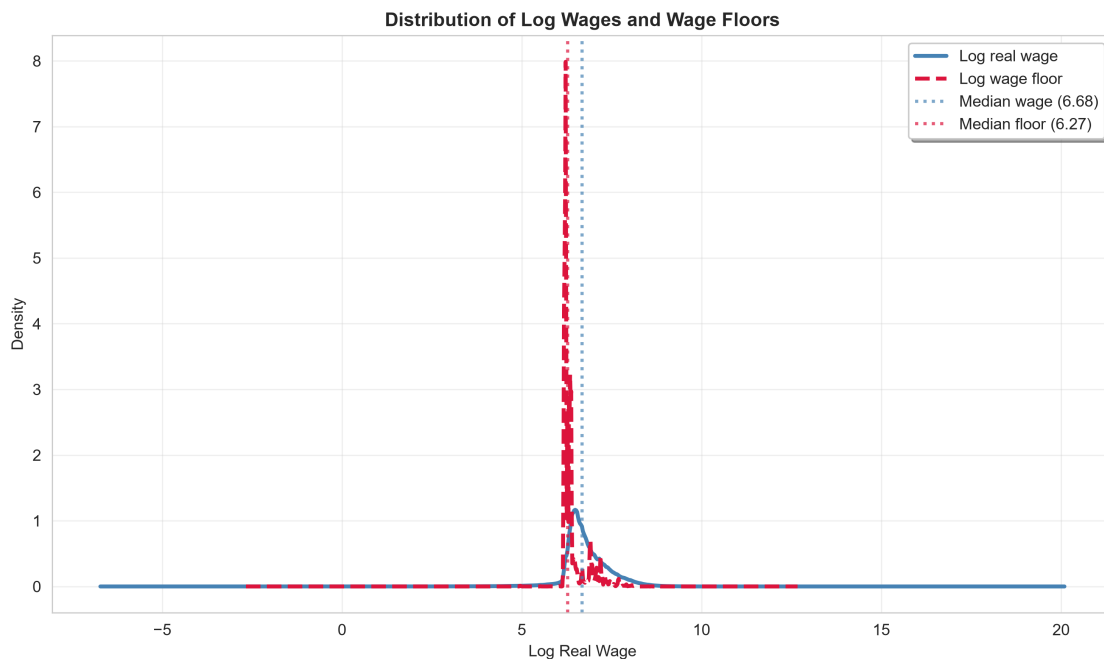


Figure 1: Distribution of Log Wages and Wage Floors

Notes: This figure displays kernel density estimates of the distribution of log real wages (solid blue line) and the log wage floor (dashed red line) in the analysis sample. The wage floor is the statutory minimum wage binding in each year and job-title–job-category–industry. The vertical dotted lines indicate the median of each distribution: median log wage of 6.68 (blue) and median log wage floor of 6.27 (red). The large mass at the wage floor indicates substantial bunching of workers at the legal minimum wage. Sample restrictions are reported in Table ??.

The Portuguese labour market displays considerable sectoral wage stratification. We document raw wage distributions across sectors, which provides context for our subsequent variance decomposition analysis. Table ?? reports summary statistics for all NACE sections represented in the data, ordered by employment share. Industry size and average wage levels are only weakly related: electricity and gas (D) and financial and insurance activities (K) record the highest mean log wages, at 7.774 and 7.593 respectively, while together employing under 4% of the workforce; manufacturing (C) and wholesale and retail trade (G) account for 42% of employment at mean log wages of 6.731 and 6.778. The gap between the highest- and lowest-paying sections – electricity and gas against accommodation and food services (I), at 6.516 – is 1.26 log points, a substantial raw differential that the decomposition below partitions into worker and firm components.

Within-section dispersion is both smaller than the cross-section spread and remarkably uniform. Standard deviations range from 0.475 in accommodation and food services to 0.685 in arts, entertainment and recreation (R), a narrow band relative to the 1.26 log point range in means. This uniformity is itself informative: it suggests that the wage-setting latitude available to firms is broadly similar across sectors, and that observed sectoral pay differences reflect composition and sectoral premiums rather than differential within-sector heterogeneity. The sections at the upper end of the dispersion range – arts and recreation, administrative and support services (N, $SD = 0.634$), and agriculture (A, $SD = 0.618$) – share a wide occupational mix and, in the case of administrative services, substantial temporary-agency employment. Notably, high-wage sections are not systematically more compressed: financial and insurance activities records an SD of 0.548, close to the sample-wide figure of 0.609, which is inconsistent with the view that standardised professional pay schedules materially narrow dispersion in skilled services. Our variance decomposition below quantifies how much of the observed heterogeneity reflects genuine firm-level wage premia rather than compositional differences in workforce characteristics.

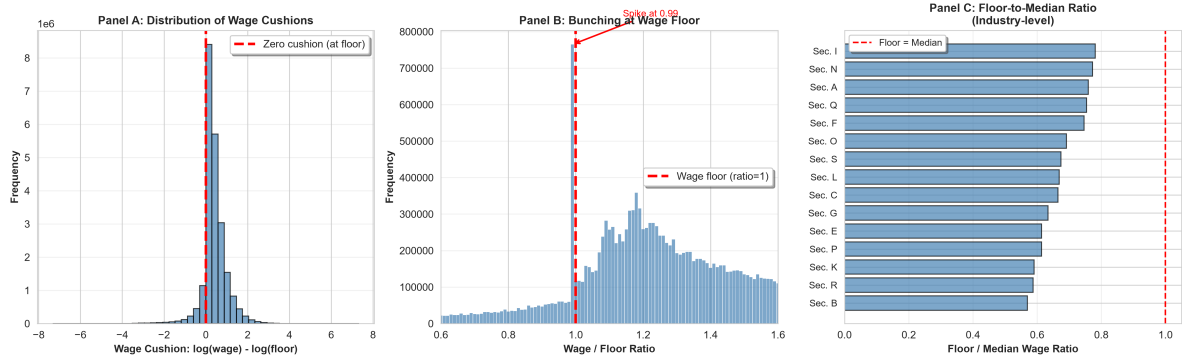


Figure 2: Wage Cushions and Bunching at the Wage Floor

Notes: This figure presents three complementary perspectives on wage floor binding and firm wage-setting. Panel A displays the distribution of wage cushions, defined as the log difference between a worker’s wage and the applicable statutory minimum wage floor in that year and region. The large mass at zero (marked by the dashed red line) indicates substantial bunching of workers at the wage floor with zero cushion. Negative cushions represent violations of minimum wage law and reflect measurement error or administrative non-compliance. Panel B shows the distribution of wage-to-floor ratios (wage divided by floor) for all workers. The spike at ratio=1.0 (dashed red line labeled "Wage floor (ratio=1)") confirms substantial bunching at the floor; the spike is annotated at 0.99 to highlight alternative reference points. The right-skewed distribution above the floor reflects workers with binding cushions of varying magnitudes. Panel C reports the industry-level floor-to-median wage ratio. All industries display ratios below 1.0 (indicated by the dashed red line), meaning the statutory minimum wage floor is substantially below the industry median wage. Industries with lower ratios (closer to zero) indicate larger wage premiums above the floor and potentially less minimum wage binding. Sample restrictions are reported in Table ??.

Table ?? documents worker mobility and network structure in Portugal’s labour market over 2010-2019. Panel A shows that 22.7% of workers switch employers at least once while observed in the sample: 9.4% move between firms within their current 1-digit industry, and 15.7% change both firm and industry.³ That cross-industry transitions outnumber within-industry moves is itself informative: adjustment in this period operated primarily through sectoral reallocation rather than through within-sector job search, consistent with the contraction of construction and manufacturing employment during the sovereign debt crisis and

³These categories are not mutually exclusive, as a worker observed making both types of transition is counted in each; the two shares therefore sum to slightly more than the 22.7% of workers recorded as moving at least once.

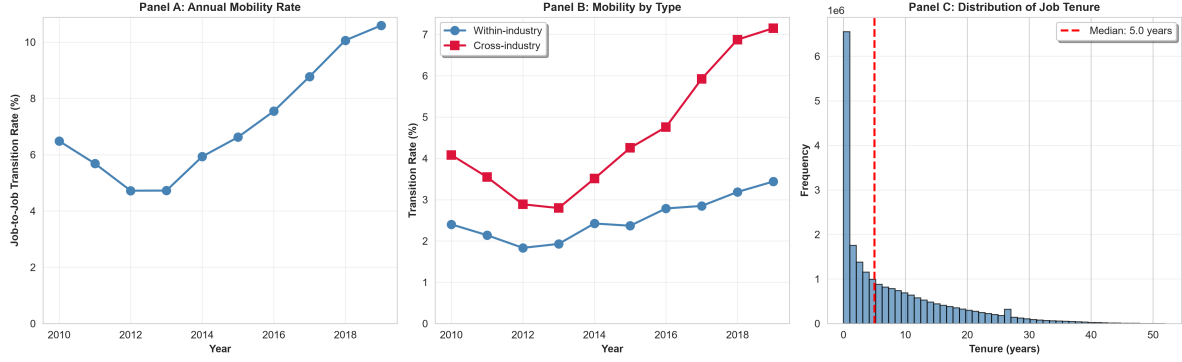


Figure 3: Worker Mobility Trends: Annual Rates, Type, and Job Tenure Distribution

Notes: This figure documents worker mobility patterns over time and their composition. Panel A displays the annual job-to-job transition rate from 2010 to 2019, measured as the percentage of workers changing employers in each year. Panel B decomposes annual transitions into two types: within-industry moves (workers changing employers while remaining in the same 1-digit NACE industry, shown in blue circles) and cross-industry moves (workers changing both employer and industry, shown in red squares). Panel C reports the unconditional distribution of job tenure (cumulative years in current employer) across all workers in the sample. Sample restrictions are reported in Table ???. Industries classified using 1-digit NACE Rev. 2. Job tenure measured in complete years with current employer.

the subsequent reorientation toward services. The majority of workers (77.3%) nonetheless remain with a single employer throughout their observed spells, close to the 74% reported by Card, Cardoso and Kline (2016) for the adjacent 2002-2009 period in the same data. Workers hold 1.60 distinct jobs on average, with a mean spell duration of 3.53 years. These patterns reflect Portugal’s stringent employment protection legislation and centralised wage-setting institutions, which restrict job-to-job transitions relative to more fluid labour markets. Limited mobility of this kind creates identification challenges for two-way fixed effects models, since sparse worker flows between firms reduce the information available to separately identify worker and firm wage components, amplifying limited mobility bias and motivating bias-corrected estimation.

Panel B describes how labour flows are distributed across firms. Just under half of establishments (47.8%) record no observed worker mobility during the sample period, a share

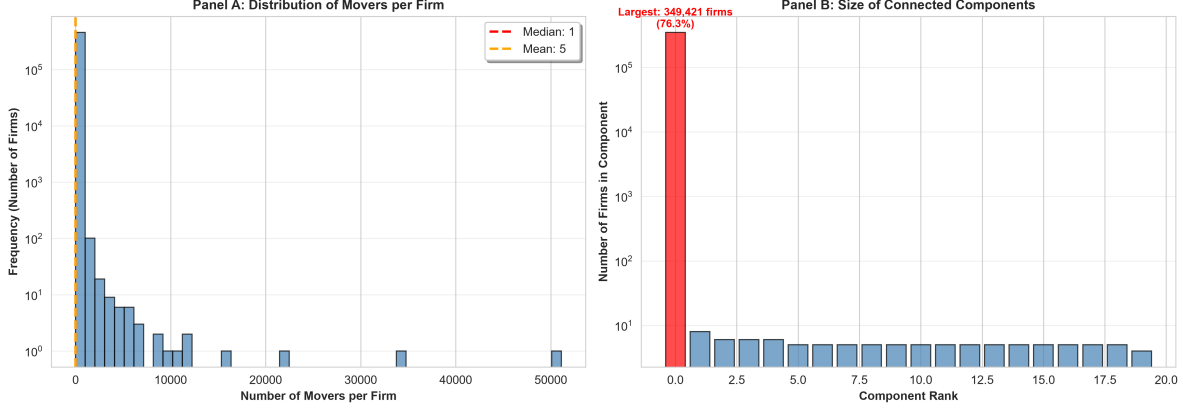


Figure 4: Worker Mobility Networks: Firm Mover Distribution and Connected Components

Notes: This figure characterizes the structure of worker mobility networks formed by inter-firm worker flows. Panel A displays the distribution of the number of movers (workers who change employers) across all 238,553 firms in the sample. The distribution is highly right-skewed: the median firm (marked by the dashed red line at zero) has zero movers, indicating that approximately 48% of firms experience no worker outflows or inflows. Panel B reports the size distribution of connected components ranked by number of firms. A connected component is defined as a maximal set of firms linked through worker flows (following Kline, Saggio & Sølvssten (2020) methodology): two firms are connected if workers move directly between them, or are connected through a chain of intermediate firms. Firms in this giant component are substantially connected through worker movements, forming a unified network. The remaining components are substantially smaller (shown in blue): the second-largest component contains approximately 10 firms, and component size declines dramatically in the right tail. Sample restrictions are reported in Table ??.

driven by the prevalence of very small firms: mean employment is 15.95 workers, and the distribution is heavily right-skewed. Among firms that do experience turnover, mobility is substantial: firms average 5.49 movers, representing 21.6% of their workforce, and 7.5% of firms record more than ten movers. Mobility is therefore concentrated but not confined to a small set of large establishments, with a further 38.8% of firms recording between one and five transitions. The prevalence of zero-mover firms nonetheless matters directly for our estimation sample: such firms contribute to bias-corrected estimation only if they remain connected to the network after removing any single worker, which reduces the effective sample available for leave-one-out procedures.

Panel C characterises the network linking firms through worker mobility. The largest connected component – the set of firms reachable through worker transitions – comprises 349,421 firms (76.3% of all firms) and accounts for 96.7% of person-year observations (see Figure ??, Panel B). Observation-weighted coverage is the relevant metric for AKM identification: what matters is not whether small, stable firms are connected, but whether employment predominantly occurs in firms linked through worker flows. The gap between firm-level connectivity (76.3%) and observation-weighted coverage (96.7%) confirms that excluded firms are overwhelmingly small establishments of limited labour market significance; the 99,850 components outside the giant component together account for barely 3% of employment. The leave-one-out connected set recommended by Kline, Saggio and Sølvesten (2020) – constructed by retaining only firms that remain connected after removing any single worker, thereby avoiding leverage-induced bias – includes 238,553 firms (52.1% of all firms) covering 87.5% of observations. The reduction from the simple connected set to the leave-out set reflects residual network fragility, with a substantial number of firms tied to the network through only one or two mobile workers, but it leaves the large majority of employment intact. Network density is low: a clustering coefficient of 0.09 indicates that firms sharing workers with a common third firm have only a 9% probability of also sharing workers directly, while an average path length of 4.12 steps implies that firms are typically four degrees of separation apart in the mobility network.

These figures place Portugal within, rather than far below, the range of connectivity reported in the matched employer-employee literature. Card, Heining and Kline (2013) report that approximately 95% of German employment falls within the largest connected set; our observation-weighted coverage of 96.7% is comparable, notwithstanding materially lower cumulative worker mobility. The distinction is important, and is easily obscured. Worker mobility rates and network connectivity are related but not interchangeable: connectivity depends on whether the movers who do exist are distributed across firms in a way that links

them, not on how numerous they are. Portugal’s institutional environment restricts the volume of transitions, but the transitions that occur are sufficiently dispersed to connect the overwhelming majority of employment. A further methodological point follows. Estimates of connectivity are highly sensitive to sampling: because a random sample of workers deletes precisely the mover links that constitute the edges of the mobility network, samples of the kind commonly used in this literature understate connectivity substantially. Our use of the full population of covered workers, rather than a sample, accounts for a considerable part of the difference between the figures reported here and those obtained from sampled Portuguese data.

The observation-weighted coverage of our leave-out connected set (87.5%) is therefore comfortably sufficient for credible identification using bias-corrected estimators. Kline et al. (2020) establish that the leave-one-out correction is valid provided the leave-out set covers a substantial share of employment, a condition we satisfy with a wide margin. Portugal remains a valuable setting for evaluating bias correction methods, but for a reason different from sheer network sparsity: it combines low aggregate worker mobility with a comprehensive system of industry- and occupation-specific wage floors, allowing us to examine how institutional wage compression interacts with the identifying variation available in matched employer-employee data. It is this interaction, rather than connectivity alone, that motivates the analysis in Section 4.4.

Table ?? reports mobility statistics for the twenty NACE Rev. 2 sections present in the full population. Worker mobility varies substantially across sections, from 8.95% in electricity, gas and air conditioning (D) and 12.21% in public administration and defence (O) to 49.65% in administrative and support service activities (N), the section that contains temporary-work agencies. Among sections with non-trivial employment shares this represents roughly a fourfold-to-fivefold range, a marked compression relative to the dispersion observed across

four-digit industries, as expected when heterogeneous sub-industries are aggregated to the section level. The ordering is nonetheless economically coherent: low-mobility sections (electricity; public administration; education (P, 12.87%); and human health and social work (Q, 14.07%)) are capital-intensive or public-sector activities characterised by firm-specific human capital, strong employment protection, and collective agreements that sustain long, seniority-based employment relationships, whereas high-mobility sections (administrative and support services; information and communication (J, 34.37%); accommodation and food services (I, 33.62%)) operate closer to spot markets with frequent job-to-job transitions driven by project completion, seasonality, and poaching. The persistence of a substantial mobility gradient across sections is consistent with a segmented labour market in which protected core activities coexist with flexible peripheral ones, in the spirit of dual labour market theory (Saint-Paul, 1996) and its Southern-European, two-tier variants (Bentolila, Dolado & Jimeno, 2012).

The leave-one-out (LOO) connected set is uniformly large across sections, ranging from 94.86% in agriculture, forestry and fishing (A) to 98.18% in water supply and waste management (E), with every substantial section between roughly 95% and 98% (extraterritorial organisations (U), with negligible employment, reach 100%). There is essentially no cross-sector variation in identification quality: firm wage premiums are precisely and comparably identified in all sections, including traditionally lower-mobility ones such as manufacturing (C, 95.78%) and public administration (O, 96.25%). This uniform, high connectivity – mirrored at the aggregate level, where the largest connected set covers 76.3% of firms and 96.7% of observations and the LOO set covers 52.1% of firms and 87.5% of observations – strengthens rather than qualifies our estimates: limited-mobility bias, which is most severe when firms are weakly linked and movers per firm are few (Andrews et al 2008; Kline et al 2020), is not a first-order concern here. On the full population the Portuguese network is well connected by international standards. The section-level figures reported here are observation-weighted and are therefore comparable to Card, Heining and Kline (2013), who

find that the largest connected set covers about 97% of person-year observations and roughly 95% of workers in West German data (their establishment-weighted figure is about 90%), and to Card, Cardoso and Kline (2016), whose largest connected set covers 91% of male and 88% of female person-years in Portugal. The weighting distinction matters: firm-weighted LOO retention (52.1% overall) is far below the observation-weighted section figures because the pruned firms are overwhelmingly small.

Table 2: Descriptive Statistics

Variable	Mean	SD	P10	P25	P50	P90
<i>Panel A: Worker characteristics</i>						
Education (%)						
Other	16.46					
Primary	54.18					
Secondary	10.06					
Tertiary	19.30					
Tenure (years)	8.10	8.83	0.00	1.00	5.00	21.00
<i>Panel B: Firm characteristics</i>						
Firm size (employees)	1,112.44	3,344.30	7.00	17.00	72.00	2,605.00
Firm size distribution (%)						
Fewer than 20 employees	28.06					
20–49 employees	16.11					
50–249 employees	24.99					
250 or more employees	30.84					
Firm age (years)	27.90	43.69	5.00	11.00	20.00	49.00
<i>Panel C: Wage characteristics</i>						
Log real wage	6.792	0.609	6.290	6.446	6.681	7.571
Real wage (euros/month)	1,089.78	1,139.48	539.05	630.14	797.48	1,940.30
Wage inequality (ratios)						
90/10	3.60					
90/50	2.43					
50/10	1.48					
Wage cushion (log points)	0.399	0.561	0.000	0.135	0.322	1.046
<i>Panel D: Wage floor characteristics</i>						
Cushion distribution (%)						
Below −10%	7.37					
−10% to −5%	1.09					
−5% to 0%	4.41					
0% to 5%	2.98					
5% to 10%	4.92					
10% to 20%	14.15					
Above 20%	65.08					
Cumulative cushion shares (%)						
Above 0%	87.13					
Above 10%	79.23					
Above 20%	65.08					
Floor-to-median ratio (by industry)	0.590	0.148				

Notes: This table presents descriptive statistics for the final analysis sample. Statistics shown are mean, standard deviation (SD), and percentiles (P10, P25, P50, P90). Wage cushion is the log difference between the actual wage and the contractual wage floor. All monetary values are in real terms (2015 euros). Sample restricted to full-time workers aged 18–65; see Table ?? for sample restrictions.

Table 3: Wage Distribution by Industry

Industry	Employment	Log Wage Distribution				
	Share (%)	Mean	SD	P10	P50	P90
Manufacturing (C)	24.721	6.731	0.543	6.296	6.633	7.405
Wholesale and retail trade (G)	17.729	6.778	0.550	6.345	6.680	7.459
Human health and social work (Q)	9.570	6.713	0.536	6.278	6.605	7.325
Construction (F)	7.957	6.661	0.557	6.222	6.583	7.306
Administrative and support services (N)	7.587	6.556	0.634	6.116	6.581	7.161
Accommodation and food services (I)	6.817	6.516	0.475	6.211	6.472	7.005
Transportation and storage (H)	5.644	7.031	0.613	6.420	6.981	7.768
Professional, scientific and technical (M)	3.865	7.059	0.607	6.451	6.987	7.838
Financial and insurance activities (K)	3.613	7.593	0.548	7.039	7.586	8.211
Information and communication (J)	3.085	7.351	0.588	6.647	7.348	8.086
Other service activities (S)	2.229	6.750	0.557	6.265	6.622	7.491
Education (P)	1.901	6.945	0.568	6.347	6.885	7.664
Agriculture, forestry and fishing (A)	1.753	6.485	0.618	5.916	6.499	7.090
Water supply and waste management (E)	0.955	6.841	0.521	6.410	6.760	7.463
Arts, entertainment and recreation (R)	0.788	6.927	0.685	6.377	6.805	7.667
Real estate activities (L)	0.567	6.824	0.616	6.321	6.675	7.651
Public administration and defence (O)	0.467	6.799	0.524	6.397	6.692	7.489
Electricity, gas and air conditioning (D)	0.379	7.774	0.526	7.153	7.815	8.384
Mining and quarrying (B)	0.371	6.902	0.576	6.375	6.836	7.640
Extraterritorial organisations (U)	0.003	7.551	0.494	6.920	7.596	8.168

Notes: This table reports the distribution of log real wages by industry for the final analysis sample, with industries ordered by employment share. Industries are NACE Rev. 2 sections, identified by section letter in parentheses; section titles are abbreviated. Employment shares are computed over worker-year observations and sum to 100; section T (activities of households as employers) is not represented in the data. Wages are in natural logarithms of real monthly earnings (2015 euros). P10, P50 and P90 denote the 10th, 50th and 90th percentiles of the within-industry log wage distribution. Sample restrictions are reported in Table ??.

Table 4: Worker Mobility and Firm Network Statistics

Statistic	Value
<i>Panel A: Overall Mobility</i>	
Total person-year observations	25,826,191
Total workers	4,565,826
Total firms	458,217
Workers who switch firms (within-industry, %)	9.39
Workers who switch industries (cross-industry, %)	15.74
Workers who never move (%)	77.34
Average number of jobs per worker	1.60
Average spell length (years)	3.53
<i>Panel B: Firm-Level Mobility</i>	
Mean number of workers per firm	15.95
Mean number of movers per firm (absolute)	5.49
Mean share of movers per firm (%)	21.56
Firms with 0 movers (%)	47.83
Firms with 1–5 movers (%)	38.83
Firms with 6–10 movers (%)	5.88
Firms with >10 movers (%)	7.46
<i>Panel C: Network Connectivity</i>	
Total firms in sample	458,217
Firms in largest connected set	349,421
Firms in largest connected set (%)	76.26
Number of connected components	99,850
Observations in largest connected set (%)	96.73
Firms in leave-one-out connected set	238,553
Firms in leave-one-out connected set (%)	52.06
Average clustering coefficient	0.09
Average path length	4.12

Notes: This table summarizes worker mobility patterns and firm network characteristics in the sample prior to the connected-set restriction. Panel A reports overall mobility: within-industry movers switch firms but remain in the same 2-digit industry; cross-industry movers change both firm and industry. Panel B presents firm-level mobility. Panel C describes the network formed by worker flows between firms. The largest connected set includes all firms reachable through worker mobility links. The leave-one-out connected set follows Kline, Saggio & Sølvesten (2020). The clustering coefficient measures the probability that two firms sharing workers with a common third firm also share workers with each other. Average path length is estimated from 100 randomly sampled source firms within the largest component.

Table 5: Industry-Level Mobility Statistics

Industry	Employment Share (%)	Worker Mobility (%)	Mean Movers per Firm	LOO Connected	Connected Set (%)	Clustering Coefficient
Manufacturing (C)	24.721	19.97	11.39		95.78	0.000
Wholesale and retail trade (G)	17.729	26.62	6.56		95.80	0.000
Human health and social work (Q)	9.570	14.07	10.25		97.06	0.001
Construction (F)	7.957	29.03	6.49		96.31	0.000
Administrative and support services (N)	7.587	49.65	49.98		97.04	0.001
Accommodation and food services (I)	6.817	33.62	6.75		97.69	0.000
Transportation and storage (H)	5.644	22.64	12.64		97.95	0.001
Professional, scientific and technical (M)	3.865	32.21	7.87		96.31	0.000
Financial and insurance activities (K)	3.613	22.44	21.69		95.19	0.002
Information and communication (J)	3.085	34.37	19.65		97.33	0.002
Other service activities (S)	2.229	17.53	4.22		95.55	0.000
Education (P)	1.901	12.87	5.64		96.85	0.001
Agriculture, forestry and fishing (A)	1.753	29.66	5.79		94.86	0.000
Water supply and waste management (E)	0.955	22.66	23.45		98.18	0.003
Arts, entertainment and recreation (R)	0.788	22.12	6.32		97.47	0.001
Real estate activities (L)	0.567	26.86	4.21		96.71	0.000
Public administration and defence (O)	0.467	12.21	7.97		96.25	0.002
Electricity, gas and air conditioning (D)	0.379	8.95	22.96		95.50	0.004
Mining and quarrying (B)	0.371	23.07	9.05		95.06	0.004
Extraterritorial organisations (U)	0.003	6.82	2.27		100.00	0.000

Notes: This table reports mobility statistics by industry for the final analysis sample, with industries ordered by employment share. Industries are NACE Rev. 2 sections, identified by section letter in parentheses; section titles are abbreviated. Employment shares are computed over worker-year observations and sum to 100; section T (activities of households as employers) is not represented in the data. Worker mobility is the percentage of workers who change employers during the sample period. Mean movers per firm is the average number of workers switching firms per firm. LOO connected set refers to the leave-one-out connected set following Kline, Saggio & Sølvsten (2020), representing the share of firms in the largest connected component after removing firms connected only through single workers. The clustering coefficient measures network density of inter-firm worker flows. Sample restrictions are reported in Table ??.

Clustering coefficients in the firm mobility network are uniformly close to zero at the section level (0.000-0.004), against an overall network clustering coefficient of 0.09. The near-absence of within-section clustering is not a computational anomaly: the clustering coefficient of a large sparse graph scales approximately with mean degree divided by the number of nodes, so for sections containing tens of thousands of firms values of order 10^{-3} are precisely what one expects. Consistent with this, the highest section-level values occur in the smallest sections (electricity (D) and mining and quarrying (B), both 0.004; water and waste (E), 0.003), and the lowest (0.000, i.e. below 0.0005) in the largest (manufacturing (C) and wholesale and retail trade (G)). The gap between the near-zero within-section values and the aggregate coefficient of 0.09 is itself informative: because a typical firm’s mobility partners lie disproportionately in other sections – routed through hubs such as temporary-work agencies (section N, with 49.98 movers per firm) – the triangles that generate the aggregate network’s transitivity are predominantly cross-sectoral, and restricting the network to a single section removes exactly those bridging triangles. This pattern echoes evidence that labour-flow communities cut across standard industry classifications (Park et al., 2019; Neffke & Henning, 2013) rather than coinciding with them, and is consistent with sector-specific human capital that depreciates outside the originating industry (Neal, 1995) and with credential and licensing frictions that dampen cross-occupation mobility (Kleiner & Xu, 2020). We therefore do not interpret the small residual differences in section-level clustering as economically meaningful, and report them primarily to document that within-section worker reallocation in Portugal is locally tree-like rather than triangular.

4 Estimation

4.1 Uncorrected premium estimates

Table ?? reports the uncorrected AKM variance decomposition of residualized log wages from Equation 4.1. Worker fixed effects account for 53.6% of total wage variance and firm fixed effects for 15.5%, demonstrating that both persistent individual characteristics and firm-specific wage policies contribute substantially to inequality, with worker heterogeneity roughly three and a half times as important as firm heterogeneity. The covariance term is positive, contributing 9.6% of total variance, indicating positive assortative matching: high-wage-type workers are observed disproportionately at high-wage-type firms. Worker and firm effects together with their covariance explain 78.8% of the variance in residualized wages, leaving 21.2% to the idiosyncratic error, and the model attains an R^2 of 0.797. These magnitudes sit comfortably within the range reported in the matched employer-employee literature. Card, Heining and Kline (2013) attribute roughly 18-21% of West German wage variance to firm effects, and Card, Cardoso and Kline (2016) document positive sorting of comparable magnitude in Portuguese data; our firm share of 15.5% and positive sorting term are consistent with both. The relatively high R^2 reflects in part our restriction to the leave-one-out connected set, which excludes the smallest and least well-identified firms, and in part the inclusion of a tenure polynomial among the partialled-out covariates.

These uncorrected estimates should nonetheless be treated as an upper bound on the true dispersion of both worker and firm effects. Andrews et al. (2008, 2012) show that limited mobility bias – estimation error induced by sparse worker flows between firms – inflates the estimated variance of both sets of fixed effects while biasing the estimated covariance downward, with the magnitude of the bias increasing as the number of movers per firm falls. Our mobility network is well connected on the full population (Table ??), so we do not expect the bias to be first-order, but the direction of the distortion implies that the positive

Table 6: AKM Variance Decomposition: Uncorrected Estimates

Variance Component	Estimate	Share (%)
Total variance: $\text{Var}(\tilde{y}_{it})$	0.3538	100.00
Worker effects: $\text{Var}(\hat{\alpha}_i)$	0.1897	53.60
Firm effects: $\text{Var}(\hat{\psi}_j)$	0.0549	15.52
Covariance: $2 \cdot \text{Cov}(\hat{\alpha}_i, \hat{\psi}_j)$	0.0341	9.64
Residual: $\text{Var}(\hat{\varepsilon})$	0.0752	21.24
Model R^2	0.7972	79.72

Notes: This table decomposes the variance of log wages using the Abowd, Kramarz, and Margolis (1999) two-way fixed effects decomposition. The decomposition is computed at the observation level on residualized log real wages, $\tilde{y}_{it}$, defined as log real monthly earnings net of the time-varying covariates entering the AKM specification; firm effects are normalized relative to the reference firm. Worker effects ($\hat{\alpha}_i$) capture person-specific wage premiums; firm effects ($\hat{\psi}_j$) capture firm-specific wage premiums. Estimates are *uncorrected* for incidental parameters bias and other finite-sample biases. The positive covariance term indicates positive assortative matching between high-wage workers and high-wage firms. Variance components sum to the total variance up to rounding. The decomposition relies on the assumption that the error term is orthogonal to both worker and firm effects. Sample restrictions are reported in Table ??.

sorting reported here is, if anything, understated. Section 4.4.2 applies the homoskedastic correction of Andrews et al. (2008) and the heteroskedasticity-robust leave-out correction of Kline, Saggio and Sølvesten (2020) to quantify the adjustment.

Table ?? reports model fit statistics. The unadjusted R^2 of 0.797 indicates that worker and firm fixed effects jointly explain close to four-fifths of the variance in residualized log wages, leaving roughly one-fifth to idiosyncratic shocks and measurement error. This is high relative to prior AKM studies – Card, Heining and Kline (2013) report R^2 of approximately 0.45 in German data – for three reasons. First, our dependent variable is already net of the time-varying covariates, so the covariate contribution does not appear in the denomina-

tor. Second, the estimation sample is the leave-one-out connected set, which excludes the smallest and most weakly identified firms and thereby removes a disproportionate share of residual variance. Third, our specification includes a tenure polynomial, which Card, Heining and Kline deliberately omit on the grounds that tenure is a deterministic function of the mobility identifying the model. The adjusted R^2 of 0.758 remains close to the unadjusted figure, since the 3,690,680 estimated parameters are small relative to 22.6 million observations; this contrasts with settings in which the incidental parameter count approaches the sample size, where the adjustment is severe and the metric correspondingly uninformative. The RMSE of 0.274 log points implies that the model predicts wages with typical errors of approximately 27%, reasonable precision given unobserved time-varying productivity shocks, job-level heterogeneity subsumed into the error term, and measurement error in administratively reported earnings.

Figure ?? presents the distributions of estimated worker and firm effects from our uncorrected AKM decomposition. Panel A shows that worker effects ($\hat{\alpha}_i$) are approximately normally distributed with mean 3.344 and median 3.278 log points, the level reflecting the normalization in which the effects absorb the model intercept. The distribution is unimodal and close to symmetric, with a modest right tail, indicating continuous skill heterogeneity in the Portuguese workforce and no evidence of discrete skill segments. Panel B reveals that firm effects ($\hat{\psi}_j$) are similarly concentrated and symmetric, with mean 3.276 and median 3.282 log points. Firm effects are noticeably less dispersed than worker effects, consistent with the variance decomposition in Table ??, where worker heterogeneity accounts for 53.6% of wage variance against 15.5% for firm heterogeneity. Neither distribution exhibits heavy tails or multimodality, supporting the additive AKM specification: wage inequality arises from continuous variation in worker and firm quality rather than from discrete labour market segmentation.

Table 7: AKM Model Fit Statistics

Metric	Value
<i>Sample Size</i>	
Observations in estimation	22,602,907
Worker effects estimated	3,452,128
Firm effects estimated	238,553
<i>Model Fit</i>	
R^2	0.7972
Adjusted R^2	0.7576
Root mean squared error (RMSE)	0.2741
<i>Degrees of Freedom</i>	
Model degrees of freedom	3,690,680
Residual degrees of freedom	18,912,227

Notes: This table reports specification and fit statistics for the two-way fixed effects AKM model. The model jointly estimates worker effects ($\hat{\alpha}_i$) and firm effects ($\hat{\psi}_j$) using matched employer-employee data spanning 2010–2019. Model degrees of freedom equal the number of worker effects plus the number of firm effects, less one normalization ($3,452,128 + 238,553 - 1$); residual degrees of freedom are the remainder of the estimation sample. Model R^2 measures the proportion of log wage variance explained by worker and firm effects combined. RMSE is reported in log wage units and is not adjusted for degrees of freedom. The model is estimated by ordinary least squares, and the estimates are uncorrected for incidental parameters bias. Sample restrictions are reported in Table ??.

Panel C documents positive assortative matching between worker and firm effects, with a correlation of 0.167 and a regression slope of 0.090, indicating that a one log point increase in estimated worker quality is associated with a 0.090 log point increase in the average estimated firm premium. The relationship is positive but economically modest, which is what the institutional setting would lead one to expect: sectoral wage floors extended through *portarias de extensão* compress the lower tail of firm pay policies and limit the scope for high-wage firms to differentiate themselves from low-wage ones, attenuating the observable degree of sorting. The sign and magnitude are consistent with Card, Cardoso and Kline

(2016), who document positive but moderate sorting in the same Portuguese administrative data. Because limited mobility bias biases the estimated covariance downward (Andrews et al., 2008, 2012), the positive sorting reported here is, if anything, understated; Section 4.4.2 confirms that both bias corrections raise the covariance.

Panel D confirms these relationships through a heatmap of the joint distribution of worker and firm effects. The density is strongly concentrated near the centre of both distributions, with a mild positive tilt matching the correlation in Panel C. The absence of secondary modes or distinct clusters rules out strong segmentation into separate high-wage and low-wage labour market segments, supporting our additive specification and indicating that the Portuguese private sector operates as a single integrated wage-setting system rather than a dual market with disconnected tiers.

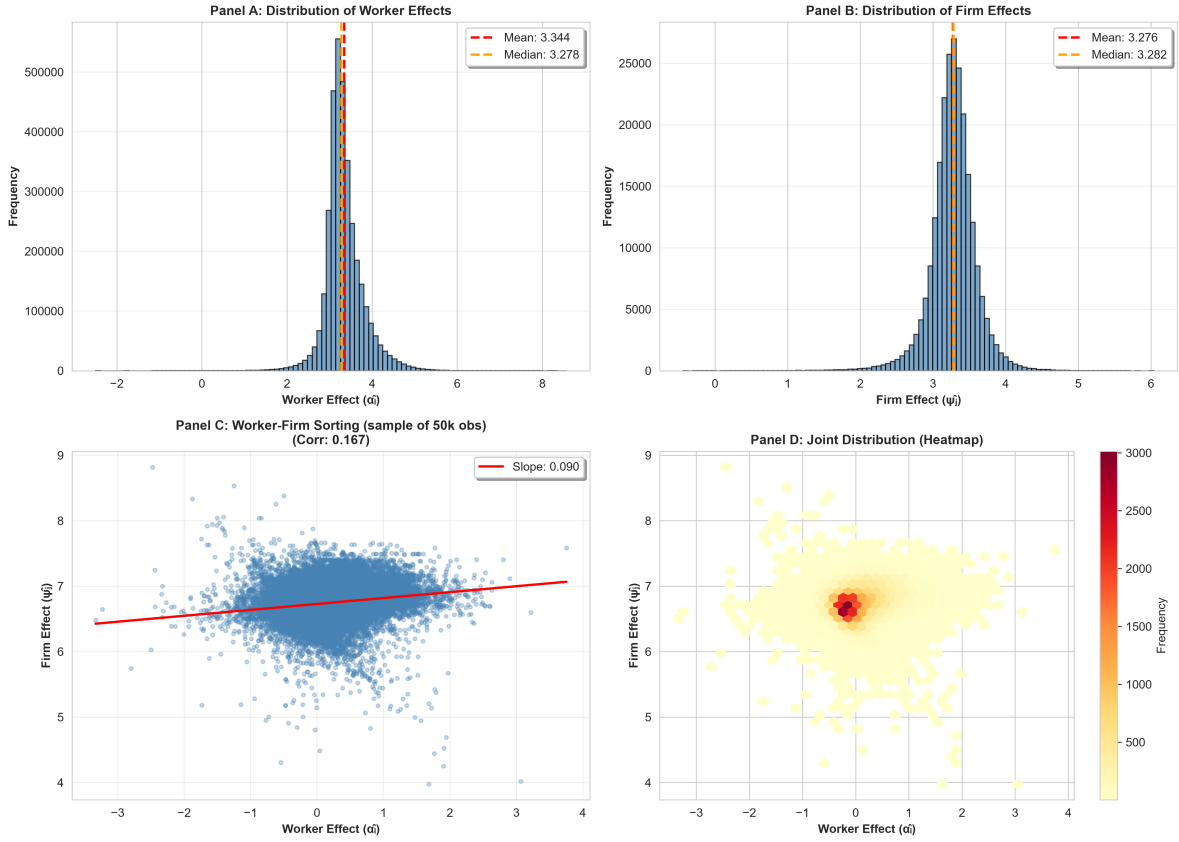


Figure 5: Estimated Worker and Firm Effects: Distributions and Sorting Patterns

Notes: This figure presents the marginal and joint distributions of estimated worker and firm effects from the AKM two-way fixed effects model. Panel A displays the marginal distribution of worker effects ($\hat{\alpha}_i$), with mean 3.344 and median 3.278 (marked by dashed lines); the level reflects the normalization in which worker effects absorb the model intercept. The distribution is unimodal and approximately symmetric with a modest right tail. Panel B shows the marginal distribution of firm effects ($\hat{\psi}_j$), with mean 3.276 and median 3.282. Firm effects are less dispersed than worker effects, consistent with the variance shares reported in Table ??, and are likewise unimodal, with the mass of firms clustered close to the median pay policy and thinner tails of firms paying substantially above or below. Panel C presents a scatter plot of worker effects against firm effects for a random sample of 50,000 observations. The correlation is 0.167 and the fitted slope is 0.090 (upward-sloping trend line), indicating positive but economically modest assortative matching: higher-ability workers are somewhat more likely to be employed at higher-paying firms. Panel D displays the joint distribution as a hexagonally binned heatmap. The density is strongly concentrated at the centre of both distributions, with a mild positive tilt matching Panel C and no secondary modes, indicating a single integrated wage-setting distribution rather than distinct labour market segments. Estimates are uncorrected for limited mobility bias; see Table ?? for bias-corrected variance components. All effects are reported in log wage units on residualized log real wages. Sample restrictions are reported in Table ??.

Table 8: Industry Wage Premiums by NACE Section

Section	Employment Share (%)	Weighted Premium	SD Firm Effects	Firm Effect		Number of Firms
				P10	P90	
Manufacturing (C)	24.79	-0.02	0.26	-0.39	0.17	34,831
Wholesale & retail trade (G)	17.69	0.01	0.30	-0.40	0.25	57,252
Human health & social work (Q)	9.59	-0.03	0.26	-0.35	0.19	10,192
Construction (F)	7.96	-0.09	0.35	-0.56	0.18	33,315
Administrative & support (N)	7.63	-0.08	0.32	-0.43	0.27	7,904
Accommodation & food service (I)	6.85	-0.08	0.30	-0.46	0.15	33,458
Transportation & storage (H)	5.68	0.12	0.33	-0.44	0.31	8,990
Professional, scientific & technical (M)	3.82	0.05	0.33	-0.42	0.33	15,525
Financial & insurance (K)	3.61	0.32	0.35	-0.35	0.47	2,380
Information & communication (J)	3.07	0.15	0.34	-0.35	0.44	4,712
Other service activities (S)	2.18	-0.02	0.33	-0.44	0.26	8,221
Education (P)	1.87	0.04	0.29	-0.39	0.27	3,441
Agriculture, forestry & fishing (A)	1.74	-0.16	0.46	-0.78	0.22	9,166
Water supply & waste (E)	0.96	0.04	0.27	-0.30	0.22	738
Arts & recreation (R)	0.79	0.12	0.37	-0.46	0.33	2,624
Real estate (L)	0.56	0.01	0.34	-0.46	0.32	4,375
Public administration (O)	0.47	0.03	0.22	-0.25	0.22	549
Electricity & gas (D)	0.38	0.45	0.30	-0.16	0.57	181
Mining & quarrying (B)	0.37	0.10	0.34	-0.46	0.28	689
Extra-territorial organisations (U)	0.00	0.45	0.30	-0.13	0.57	10

Notes: Industry wage premiums aggregated to NACE Rev. 2 sections (letter shown in parentheses), ordered by employment share. Weighted premium is the employment-weighted mean of firm-specific wage effects within each section (Equation 4.6), normalised so that the employment-weighted mean across all firms is zero, so premiums are relative to the average job. SD firm effects measures dispersion of firm premiums within section; P10 and P90 are the corresponding percentiles of the within-section firm effect distribution. All quantities are in log wage units. Each firm is assigned its modal section over the sample period. Sample: 22,602,907 person-year observations, 3,452,128 workers, 238,553 establishments, 2010–2019.

Table ?? reports industry wage premiums aggregated to NACE Rev. 2 sections, normalised so that the employment-weighted mean across all firms is zero. Premiums range from -0.16 log points in agriculture, forestry and fishing (A) to 0.45 in electricity and gas (D), a spread of 0.61 log points, or approximately 84 percent in levels, for observationally identical workers.⁴ Restricting attention to sections employing more than one percent of the workforce narrows the range slightly, from -0.16 in agriculture to 0.32 in financial and insurance activities (K). The ranking is economically coherent: capital-intensive and skill-intensive sections (electricity and gas, financial and insurance activities, information and communication (J, 0.15)) pay above-average premiums, while agriculture, construction (F, -0.09), accommodation and food services (I, -0.08) and administrative and support services (N, -0.08) pay below. That substantial sectoral premiums persist after conditioning on worker heterogeneity confirms that industry affiliation contributes to wage inequality independently of workforce composition, consistent with the pattern documented since Krueger and Summers (1988) and, more recently, by Card, Rothstein and Yi (2023). The magnitudes are nonetheless considerably smaller than raw sectoral wage gaps, indicating that a substantial share of observed cross-sector wage differences reflects the sorting of workers rather than the sectoral premium itself.

Within-section dispersion in firm effects exceeds the dispersion of premiums across sections. The standard deviation of firm effects within sections averages 0.31 log points and the median $P90 - P10$ spread is approximately 0.60 log points – comparable to the entire 0.61 log point range of premiums across sections, and substantially larger than the interquartile variation among them. Firm-level wage-setting is therefore at least as important as sectoral affiliation in determining pay, and arguably more so, since a worker’s firm within a section can move their wage by as much as their choice of section. Agriculture exhibits the highest within-section dispersion ($SD = 0.46$), reflecting a wide range of firm scale and formality

⁴Extra-territorial organisations (U) record the same premium as electricity and gas but employ ten firms and a negligible employment share; we exclude the section from this comparison.

within the sector, followed by arts and recreation (R, 0.37), construction (F, 0.35) and financial and insurance activities (K, 0.35). At the tighter end, public administration (O, 0.22), manufacturing (C, 0.26), human health and social work (Q, 0.26) and water supply and waste management (E, 0.27) display more compressed firm effect distributions, consistent with standardised pay structures under sectoral collective agreements and, in the case of public administration, with administratively determined pay scales. These estimates show high-premium sections such as financial and insurance activities to be among the more internally dispersed.

4.2 Bias-corrected premium estimates

Table ?? reports the bias-corrected AKM variance decomposition. All three specifications are estimated on an identical sample – the leave-one-out connected set of 22,602,907 person-year observations, 3,452,128 workers and 238,553 firms – so that the variance components are directly comparable across columns and the differences between them reflect the corrections themselves rather than compositional change.⁵ This is a departure from the common practice of estimating the uncorrected decomposition on the largest available sample and the corrected decompositions on progressively restricted connected sets, under which the "correction" conflates bias adjustment with selection into the estimation sample. Since the total variance of residualized log wages is held fixed at 0.3538 by construction, the corrections redistribute variance between components rather than changing the quantity being decomposed.

Both corrections move every component in the direction predicted by theory, and the heteroskedasticity-robust correction moves each further than the homoskedastic one. Worker effect variance falls from 0.1897 to 0.1708 under the HO correction (−9.9%) and to 0.1592 under HE (−16.1%);

⁵See Table A1 in the appendix for an overview of the AKM estimators in the PyTwoWay package, which we used to produce our results.

Table 9: Variance Decomposition: Uncorrected vs. HO vs. HE Corrections

Component	Uncorrected	HO Corrected		HE Corrected	
	Estimate	Estimate	% Δ	Estimate	% Δ
Var(y) Total	0.3538	0.3538	+0.0	0.3538	+0.0
Var($\hat{\alpha}$) Worker effects	0.1897	0.1708	-9.9	0.1592	-16.1
Var($\hat{\psi}$) Firm effects	0.0549	0.0488	-11.2	0.0445	-18.9
Cov($\hat{\alpha}, \hat{\psi}$) Covariance	0.0171	0.0222	+30.4	0.0263	+54.0
Var($\hat{\varepsilon}$) Residual	0.0752	0.0898	+19.5	0.1192	+58.6
R^2 Model fit	0.7972	0.7461	-6.4	0.6631	-16.8

Notes: All three specifications are estimated on an identical sample of 22,602,907 person-year observations, 3,452,128 workers and 238,553 firms, corresponding to the leave-one-out connected set. Columns report (1) uncorrected two-way fixed effects; (2) the homoskedastic correction of Andrews et al. (2008); (3) the heteroskedasticity-robust leave-out correction of Kline, Saggio and Sølvesten (2020), computed with 200 leverage draws. The dependent variable is log real wages net of jointly estimated worker age and tenure polynomials and year effects. Percentage changes are relative to the uncorrected estimate.

firm effect variance falls from 0.0549 to 0.0488 (−11.2%) and 0.0445 (−18.9%) respectively. This is the signature of limited mobility bias documented by Andrews et al. (2008) and Kline, Saggio and Sølvesten (2020): estimation error in the fixed effects inflates the apparent dispersion of both worker and firm heterogeneity, and removing it compresses both. Expressed as shares of total variance, firm effects account for 15.5% before correction, 13.8% under HO and 12.6% under HE. The corrected firm share sits somewhat below the 20-25% range reported by Bonhomme et al. (2023) across several developed economies, which is consistent with Portugal’s extensive system of sectoral wage floors compressing the distribution of firm pay policies relative to less regulated settings. Once all specifications are estimated on a common sample, the homoskedastic correction behaves as Andrews et al. (2008) predict, differing from the heteroskedastic correction in magnitude rather than direction.

The covariance between worker and firm effects moves in the opposite direction, as theory requires. Limited mobility bias attenuates the estimated covariance because estimation error in $\hat{\alpha}_i$ and $\hat{\psi}_j$ is negatively correlated by construction; removing it should therefore raise the

measured degree of sorting. The covariance rises from 0.0171 to 0.0222 under HO (+30.4%) and to 0.0263 under HE (+54.0%), so that the contribution of sorting to total wage variance increases from 9.6% to 12.6% and 14.9% respectively. Positive assortative matching is thus present in the Portuguese private sector and is, if anything, understated by uncorrected estimates – a finding consistent with Card, Heining and Kline (2013) for Germany and with Card, Cardoso and Kline (2016) using the same Portuguese administrative data. The magnitude nonetheless remains modest relative to less regulated labour markets, which we attribute to the compression of firm pay policies induced by sectoral wage floors: where firms have limited scope to differentiate wages, the observable returns to matching with a high-paying employer are correspondingly limited. The residual variance rises across specifications (from 0.0752 to 0.0898 and 0.1192), mechanically absorbing the variance removed from the fixed effect components, and the model R^2 falls from 0.797 to 0.746 and 0.663 accordingly.

Table ?? ranks NACE sections by employment-weighted premium and reports alongside each the corresponding bias gap: the employment-weighted mean sampling variance of estimated firm effects within the section, following the approximation in Equation 4.3. Because all specifications are estimated on a single sample, the premium column is common to the uncorrected, HO and HE decompositions, and the bias gap – rather than a difference between two premium series – is the quantity that measures how much limited mobility bias afflicts each section. This is a deliberate departure from the more common practice of comparing premiums estimated on a full sample against premiums estimated on a connected subsample. Such comparisons conflate bias correction with sample composition: firms excluded by the connectivity restriction are systematically small and weakly linked, so the apparent "correction" partly measures which firms were dropped rather than how mispriced the retained ones were.

Table 10: Industry Wage Premiums and Bias Gaps by NACE Section

Section	Empl. Share (%)	Weighted Premium	Bias Gap	Number of Firms
Electricity & gas (D)	0.38	0.453	0.0025	181
Extra-territorial organisations (U)	0.00	0.446	0.0186	10
Financial & insurance (K)	3.61	0.323	0.0026	2,380
Information & communication (J)	3.07	0.147	0.0038	4,712
Transportation & storage (H)	5.68	0.122	0.0056	8,990
Arts & recreation (R)	0.79	0.118	0.0160	2,624
Mining & quarrying (B)	0.37	0.101	0.0099	689
Professional, scientific & technical (M)	3.82	0.051	0.0125	15,525
Education (P)	1.87	0.043	0.0097	3,441
Water supply & waste (E)	0.96	0.037	0.0038	738
Public administration (O)	0.47	0.030	0.0092	549
Wholesale & retail trade (G)	17.69	0.008	0.0137	57,252
Real estate (L)	0.56	0.008	0.0220	4,375
Other service activities (S)	2.18	-0.015	0.0148	8,221
Manufacturing (C)	24.79	-0.024	0.0076	34,831
Human health & social work (Q)	9.59	-0.033	0.0082	10,192
Accommodation & food service (I)	6.85	-0.079	0.0146	33,458
Administrative & support (N)	7.63	-0.083	0.0040	7,904
Construction (F)	7.96	-0.088	0.0186	33,315
Agriculture, forestry & fishing (A)	1.74	-0.158	0.0272	9,166

Notes: NACE Rev. 2 sections ranked by employment-weighted wage premium, in log wage units, normalised so the employment-weighted mean across all firms is zero. The bias gap is the employment-weighted mean sampling variance of estimated firm effects within the section, computed as $\hat{\sigma}_s^2/M_j$ following the approximation in Equation 4.3, where M_j is the number of movers connecting firm j and $\hat{\sigma}_s^2$ is the section-specific residual variance. Larger values indicate sections whose firm effects are more weakly identified. All estimates are computed on the leave-one-out connected set; see Table ?? for the corresponding variance decomposition.

The bias gaps vary by roughly an order of magnitude across sections, from 0.0025 in electricity and gas (D) and 0.0026 in financial and insurance activities (K) to 0.0272 in agriculture, forestry and fishing (A) and 0.0220 in real estate (L). The ordering is intelligible in terms of network structure rather than sectoral pay levels. Sections in which firms are large and connected by many movers – electricity and gas (22.96 movers per firm), financial and insurance activities (21.69), information and communication (19.65) – yield precisely identified firm effects and small gaps. Sections composed of numerous small establishments linked by few transitions each – agriculture (5.79 movers per firm), real estate (4.21), other service activities (4.22) – yield the largest. The relationship is with the density of the mobility network, not with sectoral wage levels: administrative and support services (N) records a low gap of 0.0040 despite a below-average premium, reflecting the very high mover intensity of temporary-work agencies, while professional, scientific and technical activities (M) records a relatively high gap of 0.0125 despite an above-average premium. Extra-territorial organisations (U), with ten firms and negligible employment, record a gap of 0.0186 and should be disregarded throughout. This heterogeneity implies that estimated premiums carry section-specific precision, and it is the variation we exploit in Section 4.4.3 to test whether institutional wage floors moderate the severity of limited mobility bias.

4.3 Wage Floors as a Constraint on Sorting Bias

A striking feature of our bias-correction results is their heterogeneity across industries: the sampling variance of estimated firm effects varies by almost two orders of magnitude across sectors. We investigate whether institutional wage floors help explain this variation. As set out in Section 4.2, the effect of wage floors on limited mobility bias is ambiguous a priori, operating through two opposing channels. Through a compression channel, floors truncate residual wage variation at the bottom of the distribution, reducing the section-specific residual variance σ_s^2 and hence the sampling variance of estimated firm effects (Andrews

et al., 2008; Kline, Saggio & Sølvesten, 2020). Through a mobility channel, by compressing between-firm wage differentials, floors weaken workers' incentive to change employers (Dube, Lester & Reich, 2016; Liu, 2022; Portugal & Cardoso, 2006), thinning the number of movers M_j that connect firms and thereby raising the bias, which scales approximately as σ_s^2/M_j (Andrews et al., 2008). Section 4.1 framed the net effect as an open empirical question, and Section 4.2 conjectured that the mobility channel would dominate; the evidence below confirms that conjecture.

Figure 4.6 relates our industry bias measure to the floor-binding rate. We define the bias gap for section s as the employment-weighted mean sampling variance of estimated firm effects, $\hat{\sigma}_s^2/M_j$, following the first-order approximation of Andrews et al. (2008), where M_j is the number of movers connecting firm j and $\hat{\sigma}_s^2$ is the section-specific residual variance, so that both channels are embedded in the single measure. This definition differs fundamentally from the measure used in the previous draft, which compared premiums estimated on a full sample with premiums estimated on a connected subsample; that difference reflected sample composition – which firms the connectivity restriction dropped – rather than bias correction, and is identically zero once all specifications are estimated on one fixed sample. Because the bias gap is a variance in squared log points, its square root is readily interpretable as the typical standard error of an estimated firm effect $\hat{\psi}_j$. Panel A plots the bias gap against the floor-binding rate across the 78 NACE two-digit industries with at least 20 firms (of 84 in total). The fitted, employment-weighted slope is positive, +0.0002: industries with more binding floors exhibit larger limited mobility bias.

Panel B confirms that the compression channel is nonetheless operative and strong. The mean wage cushion – the gap between the log wage and the log wage floor – falls steeply with binding, from about 1.4 log points at roughly 2% binding to about 0.1 log points at roughly 35% binding. Residual wage variation at the bottom of the distribution is thus

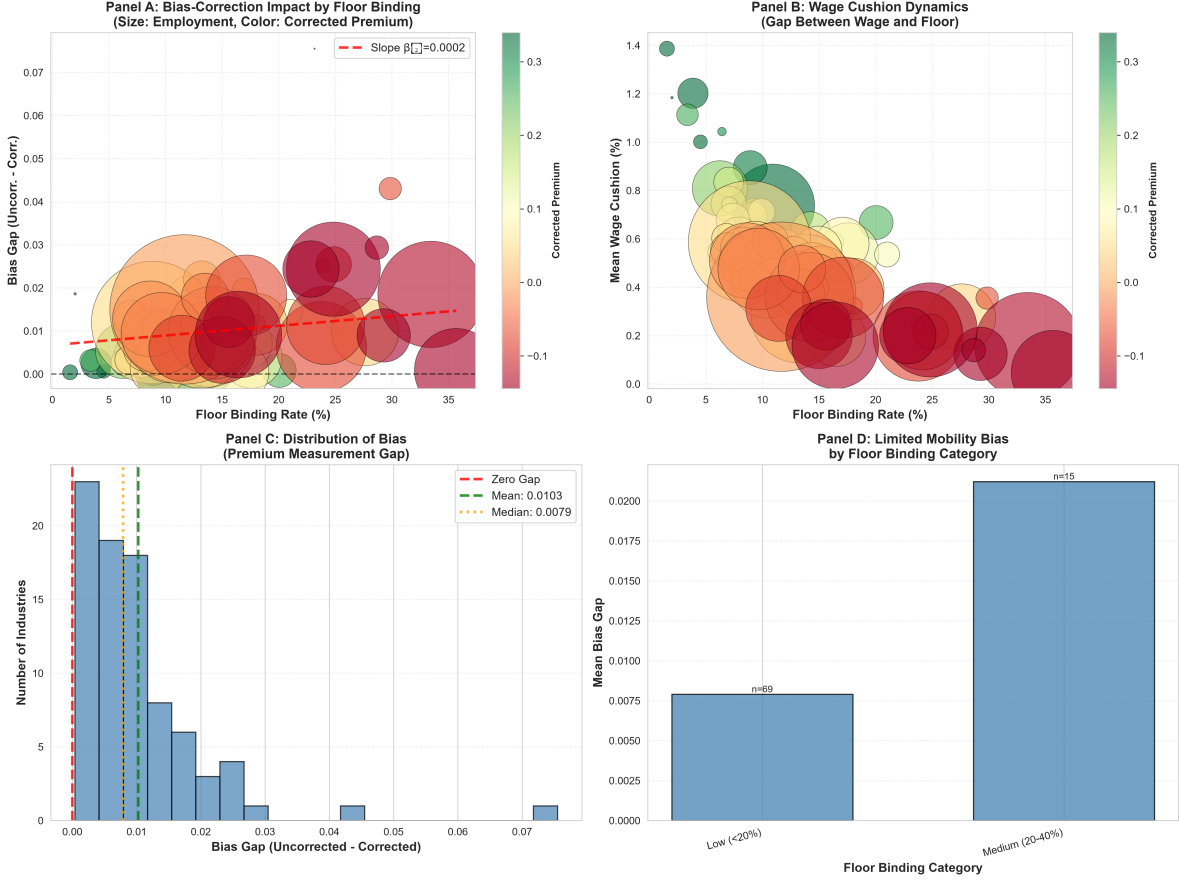


Figure 6: Bias Correction for Wage Floor Binding: Premium Measurement and Cushion Dynamics (HO Connected Set)

Notes: This figure documents how institutional wage floors relate to the severity of limited mobility bias in estimated firm effects. The bias gap for industry s is the employment-weighted mean sampling variance of estimated firm effects within the industry, $\hat{\sigma}_s^2/M_j$, following the first-order approximation of Andrews et al. (2008), where M_j is the number of movers connecting firm j and $\hat{\sigma}_s^2$ is the industry-specific residual variance; it is measured in squared log points, and larger values indicate firm effects that are more weakly identified. Panel A relates industry floor-binding rates (x-axis) to the bias gap (y-axis). Bubble size represents industry employment and bubble colour the industry wage premium (darker green = higher premium, red = lower). The dashed line is the employment-weighted fit across the 78 industries with at least 20 firms; its positive slope indicates that more binding floors are associated with *larger* limited mobility bias. Panel B plots the mean wage cushion – the log difference between the actual wage and the applicable floor – against the floor-binding rate, documenting the wage-compression channel: cushions fall from approximately 1.4 log points in the least constrained industries to approximately 0.1 in the most constrained. Panel C presents a histogram of bias gaps across industries. The distribution is positively skewed, with mean 0.0103 (dashed green line) and median 0.0079 (dotted orange line); the dashed red line at zero marks the no-bias reference. Panel D presents mean bias gaps grouped by floor-binding intensity, with the number of industries in each category shown above the bars. Only two categories are populated, as no industry in our data exceeds 37% binding. Floor-binding categories are defined by the percentage of workers earning between the floor and floor $\times 1.05$. All estimates are computed on the leave-one-out connected set; see Table ?? for the corresponding variance decomposition and Table ?? for sample restrictions.

compressed exactly as theory predicts; the compression channel simply does not dominate. This pattern is important for interpretation: the positive gradient in Panel A coexists with strong, first-order compression, which indicates that the Panel A relationship is not a mechanical by-product of the compression mechanism, and that it is the mobility channel that determines the sign of the net effect. This is consistent with evidence for Portugal that a typical worker earns a sizeable cushion over the applicable floor and that cushions are compressed as floors rise (Card & Cardoso, 2022).

The distribution of bias gaps (Panel C) is positively skewed, with mean 0.0103, median 0.0079, and a maximum of about 0.072. Taking square roots to express these as implied standard errors of $\hat{\psi}_j$ gives $\sqrt{0.0103} \approx 0.10$, $\sqrt{0.0079} \approx 0.09$, and $\sqrt{0.072} \approx 0.27$ log points, respectively. Panel D groups industries by binding intensity. Only two categories are populated, because the maximum binding rate in our data is about 37%: low-binding industries (below 20%; $n = 69$) have a mean bias gap of 0.0079, an implied standard error of $\sqrt{0.0079} \approx 0.089$ log points, while medium-binding industries (20–40%; $n = 15$) have a mean bias gap of 0.0212, an implied standard error of $\sqrt{0.0212} \approx 0.146$ log points. No industry in our data exceeds 40% binding, so we make no claims about more heavily constrained environments. The bias gap is thus roughly 2.7 times larger in medium- than in low-binding industries.

Taken together, these results point to a mechanism, to our knowledge not previously documented, through which labour-market institutions shape the reliability of AKM-based variance decompositions: by suppressing worker mobility, binding wage floors thin the mover network and amplify limited mobility bias, even as they compress residual wage variance. Our evidence is cross-sectional and cross-institutional, and therefore supports statements about comparisons across institutional settings rather than about time trends. In particular, because the Portuguese national minimum wage rose in nominal terms over 2010–2019, our data

cannot speak to the consequences of minimum-wage erosion. Consistent with cross-country evidence that bias-corrected firm-effect shares vary substantially with the institutional environment (Bonhomme et al., 2023), our findings imply that researchers comparing AKM decompositions across regions or bargaining regimes should account for the floor-binding environment: settings with more binding floors will tend to exhibit larger bias corrections even when the underlying economic mechanisms are held constant.

5 Concluding remarks

This paper provides bias-corrected estimates of industry wage premiums for Portugal using the full population of private-sector matched employer-employee records, and documents, to the best of our knowledge for the first time, how institutional wage floors moderate the magnitude of limited mobility bias in AKM decompositions. Estimating the uncorrected, homoskedastic and heteroskedasticity-robust specifications on a single fixed sample – the leave-one-out connected set – we find that both corrections reduce the estimated variance of worker and firm effects and raise the estimated covariance, as theory predicts. Firm effects account for 15.5% of wage variance before correction and 12.6% after the leave-out correction, while worker-firm sorting is positive throughout and strengthens substantially under correction. Turning to the institutional mechanism, our analysis reveals heterogeneity in the severity of limited mobility bias across industries of almost two orders of magnitude, and this heterogeneity is systematically related to wage-floor binding. Industries in which 20-40% of workers earn at or near the applicable floor exhibit a mean bias gap of 0.0212 – an implied standard error of roughly 0.15 log points on an estimated firm effect – compared with 0.0079, or roughly 0.09 log points, in industries where fewer than 20% of workers are floor-constrained. Wage floors do compress residual wage variation, and markedly so: mean wage cushions fall by an order of magnitude across the range of binding rates we ob-

serve. But this compression is outweighed by the reduction in job-to-job mobility that floors induce, which thins the network of movers connecting firms and thereby degrades identification. Our evidence therefore demonstrates that the mobility-reduction effect dominates the variance-compression effect, such that floors systematically amplify rather than attenuate limited mobility bias. This finding extends the minimum wage literature beyond traditional employment and wage-level effects to encompass implications for the measurement and decomposition of wage inequality itself.

A second contribution is methodological, and follows from how the first was obtained. Bias correction is commonly implemented by estimating uncorrected effects on the largest available sample and corrected effects on progressively restricted connected sets. Any resulting difference then conflates the correction with the change in sample composition, and because the firms excluded by connectivity restrictions are systematically small and weakly linked, the confound is not incidental: it is correlated with precisely the institutional variables one might wish to study. When we impose a common sample across all specifications, the apparent relationship between floor binding and premium adjustment disappears, and the relationship that survives under a properly defined measure of sampling variance runs in the opposite direction. We would encourage researchers reporting corrected and uncorrected decompositions side by side to estimate both on the leave-out set and to report sample composition at each stage of the restriction sequence.

Our results carry implications for comparing measured wage inequality across institutional contexts. Because our evidence is cross-sectional, it speaks to comparisons across bargaining regimes rather than to trends over time; the Portuguese national minimum wage rose in nominal terms throughout our sample period, so we cannot address the consequences of floor erosion directly. What the cross-sectional gradient does imply is that institutional features of labour markets – wage floors, collective bargaining coverage, employment protection

legislation – should be treated as contextual factors affecting both the reliability of matched employer-employee data for identifying the sources of wage inequality and the appropriate interpretation of bias-corrected estimates. Settings with more binding floors will tend to exhibit larger corrections even where the underlying economic mechanisms are unchanged, so that cross-country differences in measured firm wage heterogeneity may partly reflect differences in the identifying variation that institutions permit. Two extensions follow naturally. Our bias measure rests on the first-order approximation of Andrews et al. (2008); computing exact leave-out leverage would deliver a firm-specific, heteroskedasticity-robust measure and permit a sharper test. And because our identification is cross-sectional, a design exploiting variation in floor levels over time, such as the suspension of near-automatic extension of collective agreements in 2011, would allow the mechanism to be tested against within-industry variation.

References

- [1] Abowd, J. M., Creecy, R. H., and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data. *Technical Paper No. TP-2002-06*, U.S. Census Bureau.
- [2] Abowd, J.M., Kramarz, F. and Margolis, D.N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67: 251-333.
- [3] Abowd, J. M., McKinney, K. L., and Schmutte, I. M. (2018). Modeling endogenous mobility in earnings determination. *Journal of Business & Economic Statistics*, 36(3), 405-418.
- [4] Andrews, M. J., Schank, T., and Upward, R. (2004). Practical estimation methods for linked employer-employee data (No. 29). Diskussionspapiere.

- [5] Andrews, M.J. et al (2008), High wage workers and low wage firms: negative assortative matching or limited mobility bias?. *Journal of the Royal Statistical Society*, 171: 673-697.
- [6] Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, 117(3), 426-429.
- [7] Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States. *Journal of Labor Economics*, 34(S2), S67-S97.
- [8] Bonhomme, S., Lamadon, T., and Manresa, E. (2023). Discretizing unobserved heterogeneity. *Econometrica*, 91(2), 625-673.
- [9] Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2020). How much should we trust estimates of firm effects and worker sorting? *NBER Working Paper No. 27368*.
- [10] Borovičková, K., and Shimer, R. (2024). High wage workers work for high wage firms. *Journal of Political Economy*, 132(1), 1-53.
- [11] Burdett, K., and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2), 257-273.
- [12] Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, 128(3), 967-1015.
- [13] Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics*, 131(2), 633-686.

- [14] Card, David, Cardoso, A. R., Heining, Joerg, and Kline, Patrick (2018). Firms and Labour Market Inequality: Evidence and some Theory. *Journal of Labour Economics*, vol. 36, issue S1, S13 - S70.
- [15] Card, D., and Cardoso, A. R. (2022). Wage flexibility under sectoral bargaining. *Journal of the European Economic Association*, vol. 20, issue 5, 2013-2061.
- [16] Card, David and Rothstein, Jesse and Yi, Moises (2023). Industry Wage Differentials: A Firm-Based Approach, *NBER Working Paper Series* 31588.
- [17] Centeno, M., Duarte, C., and Novo, Á. A. (2015). The impact of the minimum wage on low-wage earners. *Banco de Portugal Economic Studies*, 1(3), 27-55.
- [18] Di Addario, S., Kline, P., Saggio, R., and Sølvesten, M. (2023). It ain't where you're from, it's where you're at: Hiring origins, firm heterogeneity, and wages. *Journal of Econometrics*, vol. 233, issue 2, 340-374.
- [19] Dube, A., Lester, T. W., and Reich, M. (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *The Review of Economics and Statistics*, 92(4), 945-964.
- [20] Dube, A., Lester, T. W., and Reich, M. (2016). Minimum wage shocks, employment flows, and labor market frictions. *Journal of Labor Economics*, 34(3), 663-704.
- [21] Eeckhout, J., and Kircher, P. (2011). Identifying sorting – in theory. *The Review of Economic Studies*, 78(3), 872-906.
- [22] Ferreira, P. (2009). Measuring interindustry wage differentials: Employers or occupations? *Economics Letters*, 104(1), 18-20.
- [23] Kleiner, M. M., and Krueger, A. B. (2010). The prevalence and effects of occupational licensing. *British Journal of Industrial Relations*, 48(4), 676-687.

- [24] Kline, P., Saggio, R., and Sølvesten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5), 1859-1898.
- [25] Krueger, A. B., and Summers, L. H. (1988). Efficiency Wages and the Inter-Industry Wage Structure. *Econometrica*, 56(2), 259–293.
- [26] Lachowska, M., Mas, A., Saggio, R., and Woodbury, S. A. (2023). Work hours mismatch (No. w31205). National Bureau of Economic Research.
- [27] Lamadon, T., Mogstad, M., and Setzler, B. (2022). Imperfect competition, compensating differentials, and rent sharing in the US labor market. *American Economic Review*, 112(1), 169-212.
- [28] Lavetti, K., and Schmutte, I. M. (2025). Estimating compensating wage differentials with endogenous job mobility. *Journal of Econometrics* (forthcoming).
- [29] Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review*, 96(3), 461-498.
- [30] Liu, S. (2021). Minimum wages, search, and intersectoral reallocation of workers. *Journal of Labor Economics*, 39(4), 1081-1115.
- [31] Lopes de Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy*, 126(1), 313-346.
- [32] Murphy, K. M., and Topel, R. H. (1990). Efficiency wages reconsidered: Theory and evidence. In Y. Weiss and G. Fishelson (Eds.), *Advances in the Theory and Measurement of Unemployment* (pp. 204-240). London: Macmillan.
- [33] Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics*, 13(4), 653-677.
- [34] OECD (2015). *OECD Employment Outlook 2015*. Paris: OECD Publishing.

- [35] Oliveira, A. (2023). Firm wage premia in Portugal: Evidence from matched employer-employee data. Unpublished manuscript.
- [36] Park, J., Shin, D., and Shin, Y. (2019). The anatomy of a job: Task structure and labor market flows. Unpublished manuscript.
- [37] Saint-Paul, G. (1996). Dual labor markets: A macroeconomic perspective. MIT Press.
- [38] Sorkin, I. (2018). Ranking firms using revealed preference. *The Quarterly Journal of Economics*, 133(3), 1331-1393.
- [39] Turrini, A. (2011). Fiscal consolidation and unemployment: Does EPL matter? A look at EU countries. *IZA Policy Paper No. 35*.
- [40] Woodcock, S. D. (2015). Match effects. *Research in Economics*, 69(1), 100-121.

A Appendix

Table A1: Two-way fixed effects estimators

Estimator	Description and Key Properties	Primary References
Two-Way Fixed Effects (AKM)	Standard Abowd-Kramarz-Margolis estimator with worker (α_i) and firm (ψ_j) fixed effects: $y_{it} = \alpha_i + \psi_j + \epsilon_{it}$ <i>Limitation:</i> Severe bias under limited worker mobility	Abowd et al. (1999)
Homoskedastic Bias Correction	Adjusts AKM variance estimates under homoskedastic errors using leverage adjustments	Andrews et al. (2008)
Heteroskedastic Bias Correction (HE)	Leave-out estimator addressing arbitrary heteroskedasticity using JLA (Jackknife Leverage Adjustment): $\widehat{\text{Var}}(\psi_j) = \widehat{\text{Var}}(\hat{\psi}_j) - \mathbf{B}_{jj}\hat{\sigma}^2$ <i>Key innovation:</i> Robust to unknown heteroskedastic patterns	Kline et al. (2020) ^a
Group Fixed Effects (BLM)	Clusters workers/firms into discrete types: $y_{it} = \alpha_{k(i)} + \psi_{l(j)} + \epsilon_{it}$. Reduces dimensionality and mobility requirements	Bonhomme et al. (2020)
Correlated Random Effects (CRE)	Models worker/firm effects as random but correlated: $\mathcal{L} = \prod \int p(y_{it} \alpha_i, \psi_j, \mathbf{x}_{it})dG(\alpha_i, \psi_j \mathbf{x}_{it})$	Bonhomme et al. (2020)
Fixed-Point Revealed Preference	Solves fixed-point problem via mobility patterns: $y_{it} - y_{is} > \Delta V_{j(i,t),j(i,s)}$	Sorkin (2018)
Borovickova-Shimer	Measures sorting via match-weighted covariance: $\widehat{\text{Sorting}} = \frac{1}{N} \sum_{i,j} (\hat{\alpha}_i - \bar{\alpha})(\hat{\psi}_j - \bar{\psi}) \cdot \mathbf{1}(i \in j)$	Borovickova-Shimer (2017)

^a Kline, Saggio, and S¸olvsten (2020) is the primary reference for HE correction in PyTwoWay Python package. Their JLA method enables consistent variance estimation under arbitrary heteroskedasticity.

^b All estimators address limited mobility bias, with HE and Group FE being most robust in high-heteroskedasticity settings.