SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

WHAT DO (THOUSANDS OF) UNIONS DO ? UNION-SPECIFIC PAY PREMIA AND INEQUALITY

Ellora Derenoncourt, François Gerard, Lorenzo Lagos, and Claire Montialoux

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2025-12

What Do (Thousands of) Unions Do? Union-Specific Pay Premia and Inequality*

Ellora Derenoncourt
Princeton University

François Gerard
University College London and IFS

Lorenzo Lagos
Brown University

Claire Montialoux
Sciences-Po and CNRS

August 1, 2025

Abstract

We study the role of union heterogeneity in shaping wages and inequality among unionized workers. Using linked employer-employee data from Brazil and job moves across multi-firm unions, we estimate over 4,800 union-specific pay premia. Unions explain 3–4% of earnings variation. While unions raise wages on average, the standard deviation in union effects is large (6-7%). Validating our approach, wages fall in markets with higher vs. lower union premia following a nationwide right-to-work law. Linking premia to detailed data on union attributes, we find that unions with strike activity, collective bargaining agreements, internal competition, and skilled leaders secure higher wages. High-premium unions compress wage gaps by education while the average union exacerbates them. Post right-to-work, however, worker support for high-premium unions falls when between-group bargaining differentials are large. Our findings show that unions are not a monolith—their structure and actions shape their wage effects and, consequently, worker support.

*Ellora Derenoncourt (ellora.derenoncourt@princeton.edu), François Gerard (f.gerard@ucl.ac.uk), Lorenzo Lagos (lorenzo_lagos@brown.edu), and Claire Montialoux (claire.montialoux@sciencespo.fr). We thank audiences at Brown, Princeton, UW-Madison, Stanford, UMD College Park, UBC, UCL, and RF-Berlin for helpful comments and suggestions. We are particularly thankful to Ihsaan Bassier, Henry Farber, Attila Lindner, Rafael Machado Parente, Suresh Naidu, Garima Sharma, and Damián Vergara for detailed feedback on the project. Dylan Black, Cedric Christensen, Jai Kapoor, and Sai Pitre provided excellent research assistance.

The effect of unions on wages has been a classic topic of interest in economics for over a century (Smith, 1776; Commons, 1921; Lester, 1941; Rees, 1962; Freeman and Medoff, 1984; Lewis, 1986). Most contributions by economists to this topic treat unions as a monolith, estimating a single average union premium in each setting in question (Drakopoulos and Katselidis, 2014). More precisely, although we recognize that unions can have heterogeneous effects on workers (Ashenfelter, 1972; DiNardo et al., 1996; Card, 1996, 2001) and that union attributes affect what these organizations do (Farber, 1978; Budde et al., 2024; Corradini et al., 2025), there is no comprehensive empirical evidence on how their impact on wages varies across individual unions.

In this paper, we estimate union-specific pay premia for thousands of individual unions, using administrative data from Brazil. We build on methods for studying firm pay premia and use the presence of multi-firm unions and worker moves to separate out firm vs. union effects on pay. Unions account for a substantial portion of the variance in earnings, being at least as important as local labor market and industry effects combined. We corroborate our estimates by showing that a right-to-work reform reduced wages in labor markets with higher vs. lower premium unions. Our estimates also correlate with known determinants of union wage effects, such as strike activity. Shifting focus to the effect of unions on between-group inequality, we show that unions, on average, modestly increase inequality by education, primarily due to differential premia across groups within unions. These average effects mask substantial heterogeneity—in particular, higher premium unions tend to reduce wage gaps. Finally, we show that unions that negotiate higher wages face a tradeoff between compression and worker support in the post right-to-work era.

Both data limitations and institutional features pose challenges to identifying union-specific pay premia. On the data side, systematic information on which union represents workers across different establishments is uncommon. This is critical for using job transitions as a source of variation in wages that can pin down the role of unions. On the institutional side, the unionization of workplaces is usually an endogenous outcome, implying that the selection of unions to establishments can confound estimates of a union's ability to influence wages (Beauregard et al., 2024). Moreover, disentangling firm-specific wage effects from union-specific pay premia requires that firms map to multiple unions and/or that unions map to multiple firms.

We overcome both these challenges by studying the effect of unions in the rich empirical setting of Brazil, where thousands of unions conduct establishment-level bargain-

ing, and unionization is widespread. Linking multiple administrative data sources, we identify the primary union representing every establishment in the formal private sector. We then exploit worker moves across firms and within unions in employer-employee matched data, tracking workers' wages from 2009 to 2017. In Brazil, workers are unionized by default based on the primary economic activity of their firm; as such, there is limited scope for union-to-firm selection. Furthermore, multi-firm unions are prevalent, meaning that many job transitions imply a change in firm but not a change in union, which is critical for separating firm from union effects on wages. Lastly, Brazil is an important setting to study wage inequality, as it is one of the world's most unequal emerging economies.

We estimate the pay premia of individual unions by adapting the classic AKM framework, which decomposes earnings into worker and establishment effects using worker mobility across establishments. We replace the establishment fixed effect in the original AKM with the sum of a union and firm effect, which we denote the union-firm AKM (UF-AKM). At its heart, our approach assumes additivity of firm and union effects and that the correlation between worker and establishment wage effects are adequately accounted for by this augmented model—our baseline model also includes industry and local labor market fixed effects. The extremely strong correlation (over 0.98) between the estimated establishment fixed effects from a classic AKM and the sum of our UF-AKM estimated fixed effects validates this approach. Estimation relies on two types of worker moves across establishments: those that entail changes in both firm and union and those involving a change in firm but not in union. Because unionization is widespread in our setting, our connected set is large, and limited mobility bias is not a concern.

We find that unions account for 3-4% of the variance in log earnings—as much as industry and local labor markets combined. The variance in unions makes up approximately one third of this (1.0-1.4%) while the covariance in union and worker effects makes up the rest. Despite unions explaining a small share of overall inequality, the standard deviation in union effects is substantial and economically meaningful: 6.5-7.3 log points, underscoring the importance of understanding heterogeneity across unions.

To interpret our effects as union-specific pay premia, we normalize our UF-AKM estimates to unions that lack the operating capacity to influence wages, measured as those with low revenue (using an approach akin to the one proposed in Card et al., 2016, for the normalization of establishment effects). Brazil's average premium across unions (1.8%) is similar to average premia in other contexts with high collective bargaining coverage, e.g.,

France (2%), Germany (2-4%), and Portugal (1%) (Breda, 2015; Hirsch and Mueller, 2020; Jäger et al., 2024). Premia in these contexts tend to be lower than in a Wagner-style system like the U.S., where workers face substantial legal hurdles and employer opposition. Arguably, the unions that do form in such a setting are akin to the right tail of unions in our context. Indeed the mean pay premium in the top quartile of unions in Brazil is 10.7%, which is similar to the average union premium in the U.S.

We next conduct two empirical exercises that both validate our interpretation of UF-AKM estimates as union wage effects and contribute to our understanding of determinants of union efficacy. First, we study the effects of a major reform in 2017 that weakened labor unions by introducing right-to-work nationwide, thus making mandatory worker contributions voluntary. The reform reduced aggregate union revenue by 90%. We aggregate our estimated union pay premia to the labor market level (defined as state-sector combinations) and merge them to labor force survey data that span the years of the reform. We then show that wages in high vs. low union-premia markets fell by 3.8% after the reform, a large effect considering the mean country-wide premium of 1.8%. These results support the interpretation of our union premia estimates as capturing the effects of unions on wages.

Second, we compile an exhaustive dataset of union attributes across multiple sources—including union, CBA, and strike registries—and use a data-driven approach to document key correlates of union premia. Our findings connect to several core areas in the industrial relations literature: the role of union size, the effect of strikes, and the importance of union leadership. We find that size is strongly negatively correlated with pay premia, with smaller, more local unions securing higher pay. In terms of union actions, going on strike and negotiating firm-specific CBAs have the most positive effects on union pay premia. We also show that shared governance, representative leadership, and having skilled union leaders all correlate positively with our estimated premia.

In the final part of our paper, we focus on the role of unions in between-group inequality, highlighting the importance of heterogeneity across unions in this domain. First, we estimate the effects of individual unions on earnings gaps by education, race, and gender. On average, unions in Brazil slightly exacerbate these gaps. However, just as there is substantial heterogeneity in union pay premia, there is considerable heterogeneity in union effects on inequality. Importantly, we find that unions with high pay premia tend to compress gaps by education, consistent with cross-country evidence (Jäger et al., 2024). Lastly, we show that worker support after right-to-work hinges on unions engaging in balanced

bargaining, or not greatly favoring one group over another. These results highlight the trade-off unions face in terms of worker support and reducing inequality among covered workers.

Our estimates of union-specific pay premia relate to a core literature on union wage effects (Card, 1996; Lemieux, 1998; DiNardo and Lee, 2004; Lee and Mas, 2012; Frandsen, 2021; Farber et al., 2021; Baker et al., 2024; Beauregard et al., 2024). A common feature across this work is the estimation of an average union premium, where heterogeneity is limited to differential effects of the average union over time or across groups of workers. We contribute to this literature by providing the first estimates of union-specific wage effects, which explain a non-trivial portion of the variance in wages.[1] We also contribute to the study of unions' effects on between-group inequality (Ashenfelter, 1972; DiNardo et al., 1996; Card, 2001; Biasi and Sarsons, 2022; Corradini et al., 2025; Budde et al., 2024). In particular, our UF-AKM framework disentangles two channels through which unions affect this inequality: differential sorting across unions and differential premia within unions, finding that the latter matters more for between-group inequality in Brazil.

Our paper also contributes to the literature on determinants of union pay effects, including an extensive literature on strikes (Card, 1990; Massenkoff and Wilmers, 2024; Lyon et al., 2024); collective bargaining agreements (Bassier, 2022; Lagos, 2024); size and workforce homogeneity (Alesina and La Ferrara, 2000; Farber, 2001; Biasi et al., 2025); and leadership (Corradini et al., 2025; Boudreau et al., 2025; Kaplan and Naidu, 2025). In addition to providing evidence on these well-studied channels, we also shed light on less explored attributes of unions, such as their internal governance structure or the presence of specific clauses in collective bargaining agreements.

Finally, we contribute to a large literature using the AKM framework to understand the role of firms in labor markets (Card, 2011; Card et al., 2013, 2016, 2018; Arellano-Bover and San, 2020; Gerard et al., 2021; Engbom and Moser, 2022). While the theoretical underpinning of the firm pay premia come from different sources of employers' wage setting power, another important institution that may be constraining this power is unions (Azkarate-Askasua and Zerecero, 2023; Dodini et al., 2024; Lagos, 2024). This paper augments the AKM framework to incorporate union constraints on wage setting by employers.

The paper is organized as follows. Section 1 provides background information on

[1] We also contribute to the literature on right-to-work, providing evidence the law reduces union wage effects (Lumsden and Petersen, 1975; Biasi and Sarsons, 2022; Fortin et al., 2023).

unionization in Brazil. Section 2 presents the data and describes the analysis sample. Section 3 explains and executes our empirical strategy, presenting a variance decomposition to discuss the role of unions in wage inequality among unionized workers. Section 4 documents dynamism across markets with different union premia around the 2017 right-to-work reform and relates our estimated pay premia to detailed union attributes. Section 5 analyzes the role of unions in between-group inequality, documenting heterogeneity in differential bargaining and its influence on worker support for unions. Section 6 concludes.

1 Unionization in Brazil

Before providing background on industrial relations in Brazil, we distinguish between three core concepts often used interchangeably when referring to union pay premia.

1.1 Core concepts

Union pay premia can stem from either unionization, coverage, or membership. When numerous workers want to bargain collectively, the entity that represents those workers is a union. We say that workers are *unionized* when employers (or employer associations) recognize an entity as having the right to represent its workers in collective bargaining. The union can perform many actions on behalf of its workers, the most prominent being the negotiation of a collective bargaining agreement (CBA). When such an agreement exists, the workers for whom the CBA applies are said to have *coverage*. Lastly, workers can affiliate to unions directly—regardless of coverage or employer recognition of the union—granting them *membership*.

This paper focuses on the union pay premium stemming from *unionization*. Specifically, our empirical approach estimates the wage effects resulting from workers being represented by a given union relative to some reference union (see Section 3.2). Their interpretation as union-specific pay premia arises after normalizing these effects relative to those of unions without the operational capacity to influence wages (see Section 4.1 for details).

Work on the premia associated with unionization has primarily emphasized wage effects of the average union, where the counterfactual is the absence of unionization.[2]

[2]See Gregg Lewis (1986) for a review of the foundational U.S. literature or Jäger et al. (2024) for a recent

There is little evidence on the relative effects of different unions, even in settings where multiple unions operate. In the U.S., for example, there are around 15,000 local unions (McCarten, 2021), yet the literature focuses on an average union premium.

Furthermore, in "Wagner Act" systems, such as the U.S. and Canada, where workers must actively organize, petition, and vote to form a union, unions must overcome substantial challenges to exist. The effect of these "surviving unions" may be very different from effects in settings where unionization is more widespread or even the default, as in several other industrialized economies.[3] In the latter case, unions may survive without much worker support and thus could ignore the demands from their constituents or even act to their detriment. Faced with different incentives and institutional constraints, unions are likely to behave differently, underscoring the need to study heterogeneity in union effects.

1.2 Institutional setting

Unionization In some respects, industrial relations in Brazil echo the U.S. context, with a large number of unions (over 6,000) and most CBAs covering individual firms rather than sectors. In other respects, collective bargaining in Brazil more closely resembles that of other non-Wagner-Act countries, characterized by significantly greater union involvement in private sector wage setting and more extensive state regulation of collective bargaining rights and CBA coverage (Lagos, 2024).

In particular, unionization in Brazil exists by default with pre-determined assignment of unions to workplaces. This default exists because the right to represent workers collectively is defined by category-geography cells. For historical reasons, the category-geography cells that pertain to each union do not map to industries and local labor markets (henceforth LLMs). The category component is vaguely defined, usually referring to an economic activity, but occasionally referring to a type of job. Neither of these maps to industry or occupation codes. Moreover, the geographic component varies widely since it allows for any grouping of municipalities. As such, the median share of workers covered by a given union in an industry-LLM pair is 57%, and the average union represents workers in establishments across 13 industries and 7 LLMs (see Table C1 for examples).

comparative review.

[3] Indeed, in the U.S. and other low collective bargaining coverage countries, the average union premium is large while high coverage places, such as Portugal and Finland, have premia close to zero (see Jäger et al. (2024)).

Because unions are pre-assigned to workplaces, there is limited scope for union-to-firm selection, a key feature we leverage in our empirical design (see Section 3.3). Instead, all new establishments undergo a process called *enquadramento sindical*, in which their primary economic activity and geographic location determine their union. If unions believe an assignment is incorrect, they can contest it through the labor courts. New unions may form by claiming representation over a cell that currently lacks a local union, subject to approval by the Labor Ministry. As a result of this assignment system, the vast majority of establishments are affiliated with a single designated union (Figure C1).

The outcome of the above process is a fragmented landscape of thousands of unions, each of which face limited external competition when it comes to representing workers in its category-geography cell. However, even in systems that allow for multiplicity—i.e., allowing more than one union to vie for the right to represent workers in a given cell—in practice, unions often face limited competition. For instance, in France and Germany, the state mandates that unions must meet certain representativeness criteria, and it is often difficult for more than one union to meet these criteria. In other settings, such as the U.S., turnover in unions or the raiding of one union's shop by another is rare and contentious.[4] Instead, internal competition through union elections and leadership plays a more significant role in affecting union policy than external competition, something we argue is the case in Brazil as well and which we explore in Section 4.3.

Coverage Unions provide extensive CBA coverage in Brazil: the coverage rate is about 50% in the formal private sector. CBAs have universal coverage, meaning that an agreement applies to all workers from the union's cell at employers represented in the negotiation, regardless of union membership. The high coverage rate is driven by sectoral CBAs, but the majority of CBAs are at the firm level. Sectoral CBAs are negotiated with employer associations representing firms in a corresponding category-geography cell. Given their wider scope, they tend to set general floors and standards for a sector. Firm-level CBAs are negotiated with individual employers, and their narrower scope allows for the inclusion of augmenting provisions (Horn, 2009; Lagos, 2024; Corradini et al., 2025).

Although every worker is formally represented by a union, not all workers are covered by a CBA. Some unions may choose not to engage in collective bargaining or may be unsuccessful in securing an agreement. Others may selectively negotiate firm-level CBAs with certain employers while neglecting others. Coverage is further affected by the lim-

[4] Affiliates of the AFL and CIO signed a no-raiding pact in 1954; by 2024, fewer than 3% of union elections featured more than one union (Cole, 1969; Krislov, 1955, 1991; National Labor Review Board, 2024).

ited duration of CBAs—most expire within a year, with a legal maximum of two years. In short, union discretion over collective bargaining activities and the need for frequent renegotiation mean that coverage in Brazil is far from 100%.

Membership The union membership rate (i.e., union density) in Brazil is approximately 15%.[5] Membership is voluntary and not needed for coverage. Instead, it provides access to private consumption benefits offered by unions, such as discounted private health plans and recreational facilities. To become a member, workers must pay membership fees, which are distinct from mandatory annual unionization contributions (see Section 2.1). In this context, union density need not reflect worker support for the union's role in collective bargaining.

2 Data

We combine multiple data sources to map unions to establishments and workers in Brazil and to build a rich database of union characteristics. This section presents the data sources used and describes the sample constructed for our analysis (details in Appendix C).

2.1 Data sources

Relação Anual de Informações Sociais (RAIS) Linked employer-employee data from RAIS allows us to track workers and their hourly earnings across jobs. The data cover the universe of the formal sector in Brazil. On the employers' side, information is reported at the establishment level using a unique 14-digit identifier known as CNPJ. The first 8 digit of the CNPJ are unique firm identifiers, allowing us to assign establishments to firms. On the workers' side, rich individual characteristics are reported including gender, race, and educational attainment. We have access to the 2009-2017 RAIS annual files.

Guia de Recolhimento de Contribuição Sindical Urbana (GRCSU) The primary source of revenue for most unions in Brazil is mandatory unionization contributions paid by the workers each union represents. These annual payments, equivalent to approximately a day's wage, are deducted from workers' paychecks by their employer and transferred

[5] For comparison, density in the U.S. is 9.9% today (Bureau of Labor Statistics, 2025).

to the union.[6] The GRCSU data include the amounts paid by each establishment to the unions representing their workers, excluding the agricultural sector. As such, these data allow us to determine the union with representation rights at a given establishment (see Section 2.2). Crucially, a labor reform enacted in late 2017 made unionization contributions voluntary (opt-in). We therefore interpret contributions that persist beyond 2017 as a proxy for workers' support of their union. We use the 2009-2021 GRCSU annual files.

Supplementary unions data We draw on multiple sources to build a novel dataset of union-specific attributes. These sources include the union registry *Cadastro Nacional de Entidades Sindicais* (CNES), which provides information on union structure, elections, and leadership; the CBA registry *Sistema Mediador* (SM), which includes information on the clauses included in the sectoral and firm-level CBAs negotiated by each union; and the strike registry *Sistema de Acompanhamento de Greves* (SAG), which provides data on strikes at the establishment level. Importantly, we can link each union leader to RAIS and extract information on their gender, race, and education.

Pesquisa Nacional por Amostra de Domicílios Contínua (PNAD-Contínua) This nationally representative household survey collects data covering a wide range of topics, including employment, formality, and earnings. We use the 2015-2019 quarterly PNAD-Contínua files to perform out-of-sample validation checks of our union-specific pay premia estimates, which are restricted to the 2009-2017 RAIS data.

2.2 Sample construction

Our baseline sample consists of longitudinal data tracking workers' wages across private sector jobs, where each establishment of employment is mapped to a single firm and a single union. We use the hourly earnings of workers employed in establishments in the private sector throughout the month of December, following the RAIS data construction in Gerard et al. (2021). Since firms identifiers are available in RAIS, the only remaining task is to assign a unique union to each establishment.

We assign each establishment a primary union, which is defined as the union that receives the most contributions from a given establishment in the GRCSU data. This ab-

[6]Other sources of union revenue include membership dues and assistance fees. The former are incurred by workers who voluntarily decide to become members. The latter are levied occasionally on workers gaining CBA coverage, i.e., the union asks for a fee to cover the costs of negotiating the CBA.

stracts from the possibility that establishments can have subsets of workers represented by different unions, e.g., when one of the few occupation-based unions overlaps with the one corresponding to the establishment's economic activity. However, Figure C1 shows that our abstraction is close to reality. Specifically, 83% of the establishments in GRCSU contribute to a single union. Even in the cases where this fails, there seems to be a dominant union since the primary one receives 75% of the contributions, on average. Lastly, there is negligible variation in primary unions within establishments over time (Figure C2).

3 Union-Firm AKM

We introduce an augmented version of the classic AKM decomposition of earnings into worker and establishment effects that allows us to estimate the pay premia of thousands of unions in Brazil. We term this approach the union-firm AKM, henceforth UF-AKM. As we explain in detail below, our approach relies on the ability to map individual unions to the establishments whose workers they represent.

3.1 The role of unions in explaining establishment variation in pay

The standard AKM specification that estimates worker and establishment fixed effects using worker moves is as follows:

$$y_{it} = \alpha_i + \psi_{j(i,t)} + \beta X_{it} + r_{it}, \tag{1}$$

where y_{it} is log earnings for worker i in year t; α_i is the worker fixed effect i; and $\psi_{j(i,t)}$ is the fixed effect for the establishment j where i works in year t. We include in X_{it} an unrestricted set of year dummies and quadratic and cubic terms in age, all fully interacted with educational attainment (Card et al., 2018).

This approach does not account for the role that labor market institutions, such as unions, play in shaping (or constraining) employers' pay practices, and hence, their contribution to estimated establishment fixed effects, $\hat{\psi}_j$. In fact, the source of heterogeneity in $\hat{\psi}_j$ is a subject of current debate, with proposed explanations including variation in monopsony power to heterogeneity in industry premia (Card et al., 2024).

Our UF-AKM approach is motivated by descriptive facts pertaining to the role of unions in explaining variation in the AKM establishment pay premia among multi-establishment

firms. Figure B1a shows that regressions of $\hat{\psi}_j$ on union fixed effects have an adjusted R-squared of 55%. This is greater than that of LLM and industry fixed effects, but less than that of firm fixed effects. Delving deeper, Figure B1b depicts the Shapely values of each fixed effect group, i.e., firm, industry, LLM, and union. We find that the marginal contribution of adding union fixed effects in explaining $\hat{\psi}_j$ is 14%. Despite the fact that there is noise in $\hat{\psi}_j$, the finding that firm, industry, LLM, and union fixed effects account for nearly all of the observed variation in that sample provides strong motivation for our approach, which we explore in greater detail in the following section.

3.2 Estimation of the UF-AKM

A number of designs can be used to generate estimates of union-specific pay premia. One approach is to run the AKM model with union rather than establishment fixed effects. While this union-mover approach would account for time-invariant worker attributes, such estimates might be prone to hierarchy bias, as has already been documented for the case of industry-mover designs (Card et al., 2023). Concretely, movers from low- to high-premium unions may be systematically coming from higher paying workplaces covered by the origin union into the lower paying workplaces covered by the destination union.

A second approach would be to aggregate $\hat{\psi}_j$ up to the union level. This "ground-up" approach—proposed in Card et al. (2023)—has the benefits of addressing hierarchy bias and being flexible in implementation. For example, Beauregard et al. (2024) implement this approach to estimate the union premium in Canada. However, a limitation of the "ground up" estimates is that it is difficult to net out the contribution of the pay premia of workplaces where each union operates. This may be feasible when one is interested in the average union premium, but not for estimating union-specific pay premia.[7]

We follow a third design that we term an "institutionalist" approach. The idea is to split the establishment effect into its essential components directly within the AKM specification (and without having to restrict the sample to multi-establishment firms, as in the exercises described in Section 3.1). In particular, we are interested in distinguishing between a *firm-wide pay policy* and a *union-specific bargaining effect*. In the simplest UF-AKM, this would amount to replacing ψ_j with the firm and union fixed effects of each corresponding establishment. The key assumption of this approach is that the characteristics replacing ψ_j adequately account for the correlation between worker and establishment

[7] Beauregard et al. (2024) regress $\hat{\psi}_j$ on log value added per worker for unionized and non-unionized jobs to disentangle union selection into higher premium workplaces from union rent-extraction.

wage effects (Kline, 2024). This assumption implies falsification tests, which we discuss in Section 3.3.

The specification in a UF-AKM is one where the establishment fixed effect is replaced by the sum of a firm and a union fixed effect:

$$y_{it} = \alpha_i + \phi_{f(i,t)} + \eta_{u(i,t)} + \beta X_{it} + \varepsilon_{it}, \tag{2}$$

where $\phi_{f(i,t)}$ is the fixed effect for the firm f where i works in year t; $\eta_{u(i,t)}$ is the fixed effect for the union u representing i in year t; and all other variables are defined as in Equation (1). Our baseline specification also includes LLM and industry fixed effects to better satisfy the aforementioned assumption, but we omit this for simplicity of exposition.[8]

Worker mobility Estimation in these models relies on worker mobility, which in the case of the UF-AKM, requires specifying whether a move across establishments implies a change in firm and/or union. The fact that unions represent workers across multiple firms and that some multi-establishment firms have different unions assigned to their establishments enriches the types of moves that exist in our data.

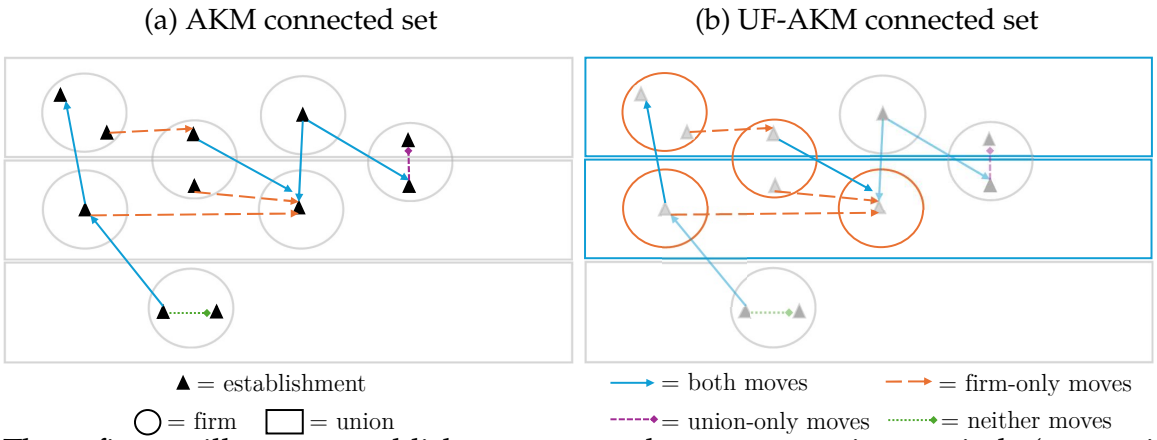
Denoting $\mathcal{M}$ as the set of moves across establishments, we group these events into four mutually exclusive and collectively exhaustive types. First, "both" moves ($\mathcal{B}$) are those where both the firm and the union change. Second, "firm-only" moves ($\mathcal{F}$) are those where the firm changes, but the union does not. Third, "union-only" moves ($\mathcal{U}$) are those where the union changes, but the firm does not. Fourth, "neither" moves ($\mathcal{N}$) are those where neither the firm nor the union change.

Estimation approach We start with examples illustrating the role different move types play in estimating AKM and UF-AKM parameters before proceeding in more detail.

Consider a standard, two-period AKM ($T = 2$) without time-varying covariates, described by the following equation: $y_{it} = \alpha_i + \psi_{j(i,t)} + r_{it}$, with $t \in \{1, 2\}$. In this simple example, the connected set of establishments—based on worker moves, as illustrated in Figure 1a—are used to estimate the relative effect of $\hat{\psi}_k$ versus $\hat{\psi}_l$, where k and l are establishments. That is, the wage change worker i experiences in moving from ℓ to k reflects

[8]For LLMs we use microregions, i.e., neighboring municipalities grouped into 543 units by IBGE that capture local labor markets. For industries we use three-digit CNAE codes, i.e., 285 detailed groupings based on IBGE's classification code for economic activities. IBGE is the Brazilian Institute of Geography and Statistics.

Figure 1: Types of Movers and Connected Sets



These figures illustrate establishment moves between two time periods (categorized by types) used to determine the connected set in which the AKM and UF-AKM parameters can be estimated. Figure 1a highlights the connected set of establishments (depicted by triangles) based on all move events $\mathcal{M}$ used for estimating AKMs. Figure 1a highlights the connected set of firms and unions (depicted by circles and rectangles, respectively) based on "both" moves $\mathcal{B}$ and "firm-only" moves $\mathcal{F}$ used for estimating UF-AKMs. Transitions within a firm—i.e., "union-only" moves $\mathcal{U}$ and "neither" moves $\mathcal{N}$—are not required for estimating UF-AKMs.

the relative effect of establishment k, i.e., $\psi_k - \psi_\ell$, as the worker effect is held constant. This approach attributes the variation in wages to establishments (depicted as triangles). However, unions (depicted as rectangles) could in fact be exerting their own influence on wages above and beyond the pay policies of firms (depicted as circles).

Figure 1b illustrates the analogous two-period example in the context of the UF-AKM, which is described by the following equation: $y_{it} = \alpha_i + \phi_{f(i,t)} + \eta_{u(i,t)} + \varepsilon_{it}$, with $t \in \{1, 2\}$. Our setting is one in which an establishment j consists of a non-unique combination of firm identifiers $f(j)$ and union identifiers $u(j)$. Within a set of firms connected by firm-only moves ($\mathcal{F}$), one can estimate the relative firm effects. This corresponds to the highlighted circles connected by the long-dashed arrows. Specifically, the wage change worker i experiences in moving from $f(\ell)$ to $f(k)$ that share the same union reflects the relative effect of firm $f(k)$, i.e., $\phi_{f(k)} - \phi_{f(\ell)}$, as both the worker and union effect are held constant. Moreover, among the unions corresponding to this connected set, the presence of both-moves ($\mathcal{B}$) allows us to estimate the relative union effects.[9] This corresponds to

[9]The existence of multi-establishment firms with different primary unions across establishments (e.g., the circles intersecting two rectangles in Figure 1b) is critical so that there is more than one union in the connected set of firms from firm-only moves.

the highlighted rectangles connected by the solid arrows. Although only the worker effect is held constant for a mover from $u(\ell)$ to $u(k)$, we can plug in our estimates for the relative firm effects, allowing us to estimate the relative union effects, i.e., $\eta_{u(k)} - \eta_{u(\ell)}$.

In short, we use the largest connected set of firms from $\mathcal{F}$ moves and their corresponding unions connected by $\mathcal{B}$ moves, to estimate the firm and union fixed effects of the UF-AKM. We can then back out the worker fixed effects with: $\hat{\alpha}_i = \frac{1}{T}\sum_t (y_{it} - \hat{\phi}_{f(i,t)} - \hat{\eta}_{u(i,t)})$. Importantly, $\mathcal{U}$ and $\mathcal{N}$ moves are not required for estimation of the UF-AKM parameters. In fact, one could drop these moves from the connected set because mobility within firms often implies transfers or promotions, where the symmetry condition of the UF-AKM specification does not hold (see Section 3.3 for details).[10] Lastly, including additional fixed effects (i.e., for local labor markets and industries) requires restricting the connected set to the units of this fixed effect that share at least one union-firm pair with another of these units in the connected set.

3.3 Identification assumptions and falsification tests

The "institutionalist" assumption we make is that establishment pay premiums are determined by *firm-wide pay policies* on the one hand, and *union-specific bargaining effects* on the other. Furthermore, the UF-AKM specification imposes that these two components are additively separable. In our Online Appendix, we derive a wage posting model constrained by collective bargaining that provides a micro-foundation for Equation (1).

We need five assumptions ($A1$-$A5$) to be satisfied so that mobility between establishments separately identifies the treatment effect of working for firm $f(j)$ and being represented by union $u(j)$. All assumptions except for $A1$ are analogues of the usual AKM assumptions applied to firms and unions.[11] We discuss each assumption in what follows, before exploring the implied falsification tests.

Preliminaries Define $D_{i,t}$ as the establishment where worker i works in year t, i.e., a random variable representation of the $j(i,t)$ index in the AKM, where realizations are denoted as d_j. Following our UF-AKM notation, $f(d_j)$ and $u(d_j)$ refer to the firm and union corresponding to establishment d_j, respectively. Let $y_{it}(D_{i,1}, D_{i,2})$ represent the potential

[10] In practice, we only drop the $\mathcal{U}$ moves since those could be used to pin down $\hat{\eta}_{u(i,t)}$ in the same way $\mathcal{F}$ moves pin down $\hat{\phi}_{f(i,t)}$. Leaving the $\mathcal{N}$ moves is innocuous, so we leave them in the connected set.

[11] In our setting, the establishment determines their firm-union pair. Hence, assumptions $A2$-$A5$ (which apply to firms and unions) are weaker than their AKM analogues (which apply to establishments). However, this would no longer be the case in settings where workers can opt out of unionization.

log earnings for worker i at $t \in \{1, 2\}$ for any employment realization in these two periods. Similarly, $y^f_{it}(\cdot, \cdot)$ and $y^u_{it}(\cdot, \cdot)$ denote potential outcome functions for firm and union pay, respectively, that may not necessarily be the same as the one defined for establishments.

We are interested in identifying the average treatment effects (ATEs) on pay of being employed by the firm (represented by the union) of a given establishment, relative to the firm (union) of some reference establishment. That is, the ATEs of interest are:

$$\mathbb{E}[y^f_{it}(f(d_k), f(d_k)) - y^f_{it}(f(d_\ell), f(d_\ell))] \quad \text{and} \quad \mathbb{E}[y^u_{it}(u(d_k), u(d_k)) - y^u_{it}(u(d_\ell), u(d_\ell))],$$

where d_k is the establishment of interest and d_ℓ is the reference establishment.

A1. Additivity without establishment effects

$$y_{it}(D_{i,1}, D_{i,2}) = y^f_{it}(f(D_{i,1}), f(D_{i,2})) + y^u_{it}(u(D_{i,1}), u(D_{i,2}))$$

The potential outcomes from working at an establishment equals to the sum of the potential outcomes from the corresponding firms and unions. This assumption implies that no other characteristic of the establishment systematically influence pay, which is why our baseline UF-AKM includes LLM and industry fixed effects. Furthermore, since there is no interaction effect, this assumption implies homogeneity in union effects across employers.[12] It is worth noting that a version of this assumption is implicit in traditional AKMs that rely on firm-level data—that is, they assume that the potential outcome function for firms subsumes the potential outcome function for establishments.

A2. Exclusion

$$\begin{aligned} y^f_{i1}(f(D_{i,1}), f(D_{i,2})) &= y^f_{i1}(f(D_{i,1})) \quad \text{and} \quad y^u_{i1}(u(D_{i,1}), u(D_{i,2})) = y^u_{i1}(u(D_{i,1})) \\ y^f_{i2}(f(D_{i,1}), f(D_{i,2})) &= y^f_{i2}(f(D_{i,2})) \quad \text{and} \quad y^u_{i2}(u(D_{i,1}), u(D_{i,2})) = y^u_{i2}(u(D_{i,2})) \end{aligned}$$

The firm- and union-specific effects do not depend on one's past or future workplaces. This assumption is violated in sequential models of the labor market. Within the AKM framework, Di Addario et al. (2023) provide evidence that establishment-of-origin effects

[12] The lack of interaction effects does not imply that unions do not respond to employers. Instead, this corresponds to the proportionality assumption in our economic model (see our Online Appendix).

do not play a role in explaining wages.

A3. Parallel trends

$$\mathbb{E}[y^f_{i2}(f(d_k)) - y^f_{i1}(f(d_k))|D_{i,1} = d_k, D_{i,2} = d_\ell] = 0 \quad \forall \ell \neq k$$
$$\mathbb{E}[y^u_{i2}(u(d_k)) - y^u_{i1}(u(d_k))|D_{i,1} = d_k, D_{i,2} = d_\ell] = 0 \quad \forall \ell \neq k$$

All movers who, at $t = 2$, left the firm (or union) assigned to d_k would, on average, have followed a common earnings trend had they not moved between the two years, regardless of their firm (or union) at destination (recall that all our specifications include a flexible set of time effects). This assumption is violated in learning models of the labor market.

A4. Stationarity

$$\begin{aligned}\mathbb{E}[y^f_{i2}(f(d_\ell)) - y^f_{i2}(f(d_k))|D_{i,1} = d_k, D_{i,2} = d_\ell] \\ = \mathbb{E}[y^f_{i1}(f(d_\ell)) - y^f_{i1}(f(d_k))|D_{i,1} = d_k, D_{i,2} = d_\ell] \quad \forall \ell \neq k \\ \mathbb{E}[y^u_{i2}(u(d_\ell)) - y^u_{i2}(u(d_k))|D_{i,1} = d_k, D_{i,2} = d_\ell] \\ = \mathbb{E}[y^u_{i1}(u(d_\ell)) - y^u_{i1}(u(d_k))|D_{i,1} = d_k, D_{i,2} = d_\ell] \quad \forall \ell \neq k\end{aligned}$$

The average treatment effect among movers working for the firm (or represented by the union) assigned to d_ℓ instead of d_k is the same regardless of the year. This assumption is violated if firm and union effects are time-varying.

A5. No selection on gains

$$D_{i,1}, D_{i,2} \perp y^f_{it}(f(d_\ell)) - y^f_{it}(f(d_k)) \quad \forall \ell \neq k$$
$$D_{i,1}, D_{i,2} \perp y^u_{it}(u(d_\ell)) - y^u_{it}(u(d_k)) \quad \forall \ell \neq k$$

Workers do not select their establishment based on comparative advantage in firm- or union-specific pay. As in the usual AKM, this assumption is violated in models with match effects, which have been shown to explain only a small part of the variation in log earnings (Bonhomme et al., 2019). Nonetheless, since these restrictions are unlikely to be satisfied exactly, economists tend to interpret the AKM as a useful statistical model.

Falsification tests Under assumptions $A1$-$A5$, fitting OLS to the UF-AKM specification means that $\hat{\phi}_f$ and $\hat{\eta}_u$ are unbiased estimates of the ATEs defined in our preliminaries—proof provided in our Online Appendix.[13] In other words, the assumptions imply an exogenous mobility condition (EMC) for the UF-AKM, i.e., $\mathbb{E}[\varepsilon_{it}|\alpha_i, \phi_{f(i,t)}, \eta_{u(i,t)}, X_{it}] = 0$. We now show falsification tests for these identification assumptions.

Figure B2 contains falsification tests for additivity ($A1$) and stationarity ($A4$) using a split sample approach to adjust for attenuation bias due to noise in AKM and UF-AKM estimates. First, Figure B2a reveals that the sum of the UF-AKM components approximates that of the AKM establishment fixed effect, suggesting that characteristics replacing the establishment effects account for the relevant factors affecting pay. Specifically, we randomly divide workers into two non-overlapping samples. We then regress the sum of the estimates for the firm, union, LLM, and industry fixed effects from the UF-AKM in one sample on the corresponding establishment's $\hat{\psi}_j$ from an AKM in the same sample, instrumenting the latter with its equivalent from the other sample. The slope of this regression is 0.986 (SE=0.008), which is statistically indistinguishable from one.[14] Second, Figure B2b shows that the union effects appear to be stationary. That is, we regress estimates of $\hat{\eta}_u$ from a late-period UF-AKM (2013-2017) in one sample on the $\hat{\eta}_u$ from an early-period UF-AKM (2009-2013) in the same sample, where the latter are instrumented by the $\hat{\eta}_u$ from a full-period UF-AKM (2009-2017) in the other sample. We obtain a slope estimate of 1.082 (SE=0.045), which is statistically indistinguishable from one.

UF-AKM analogues of the traditional mover event studies and symmetry test used to falsify the AKM identification assumptions also support our approach. Figures B4 and B5 show these falsification tests by type of mover. Those corresponding to the "both" and "firm-only" movers reveal parallel trends before the move ($A3$) and after the move (EMC), as well as symmetry in the gains/losses for moves up/down the firm and union ladders ($A1$). The figures for the "union-only" and "neither" movers are interesting as well. The "union-only" moves lack symmetry from moves down the union ladder, given downward rigidity from transitions within the same firm, which justifies their exclusion from our UF-AKM. The "neither" moves are essentially flat, suggesting that there is no ladder across establishments in the same firm-union pair ($A1$).

[13] These assumptions imply that a "two-step" approach—i.e., regressing $\hat{\psi}_{i,j(i,t)}$ on firm and union fixed effects—would also give unbiased estimates of the ATEs of interest (Kline, 2024). While the former may be simpler from an econometric standpoint, the UF-AKM (i.e., the "one-step" approach) provides clarity on the economic meaning of the assumptions, which justifies important restrictions on the connected set.

[14] In further support of the additivity of firm and union effects, Figure B3 reveals no meaningful patterns in the average residuals of the UF-AKM by cells of $\hat{\phi}_f$ and $\hat{\eta}_u$ deciles.

3.4 UF-AKM variance decomposition

To explore the role unions play in overall inequality among unionized workers, we run a variance decomposition of the UF-AKM. The main components of the UF-AKM variance decomposition (ignoring X_{it}) are:

$$Var(y_{it}) = \underbrace{Var(\alpha_i)}_{\text{transferable skills}} + \underbrace{Var(\phi_{f(i,t)}) + Var(\eta_{u(i,t)})}_{\text{pay setting and union bargaining}} + \underbrace{2Cov(\alpha_i, \phi_{f(i,t)}) + 2Cov(\alpha_i, \eta_{u(i,t)})}_{\text{assortative matching}} + \underbrace{2Cov(\phi_{f(i,t)}, \eta_{u(i,t)})}_{\text{union-to-firm selection}} + Var(\varepsilon_{it}).$$

Simply put, the variance components represent the direct influence that heterogeneity in transferable skills across workers, pay setting policies across firms, and bargaining effects across unions have on the variance of log earnings. The covariance components involving the person effects capture assortative matching of workers to firms and unions, respectively. The term $2Cov(\phi_{f(i,t)}, \eta_{u(i,t)})$ represents the selection of firms to unions. Lastly, $Var(\varepsilon_{it})$ is the variance in log earnings that is not explained by the UF-AKM.[15]

Limited mobility bias A well-documented issue with these variance decompositions is limited mobility bias, which occurs when sampling errors in the unbiased estimates introduce biases when taking second moments. Specifically, if there are few movers relative to fixed effects, variance terms are overestimated and covariance terms are underestimated (details in our Online Appendix). Luckily, in our setting, we have 9.6 million "both" movers estimating only 4,805 union fixed effects.[16] Moreover, in settings with no union-to-firm selection, such as Brazil, we expect that $Cov(\hat{\phi}_f, \hat{\eta}_u) \approx 0$. Under limited mobility bias, this covariance is negative. However, as discussed below, we find effectively zero covariance, suggesting that limited mobility bias is not a concern in our case.

Variance decomposition results Table 1 shows the results of the UF-AKM variance decomposition in the largest connected set of four separate samples: the entire country; the more developed Southeast region (where the minimum wage is less binding—see Gerard et al. (2021)); and an early period (2009-2013) and a late period (2013-2017) to demonstrate

[15] As shown in Table A1, the adjusted R-squared of the UF-AKM is high and nearly the same as that of an AKM (91%). This is only 2 percentage points slower than a match effects model.

[16] The high degree of worker mobility in RAIS implies that corrections for limited mobility bias barely affect the shares of the AKM variance decomposition (Kline et al., 2020; Gerard et al., 2021).

dynamic stability of the results. There is substantial inequality in these samples, with the standard deviation of log earnings ranging from 61.5 to 64.5 log points.

Table 1: UF-AKM Variance Decomposition of Log Earnings

	Brazil	Southeast	Early period	Late period
	(1)	(2)	(3)	(4)
Mean log earnings	2.218	2.304	2.202	2.292
Standard deviation of log earnings	0.621	0.641	0.645	0.615
Standard deviation (across person-year observations)				
Person effects	0.499	0.517	0.540	0.516
Firm effects	0.203	0.204	0.197	0.196
Union effects	0.069	0.065	0.070	0.073
Percentage of variance in log earnings due to				
Variance in person effects	64.4	65.1	70.1	70.4
Variance in firm effects	10.7	10.1	9.3	10.1
Variance in union effects	1.2	1.0	1.2	1.4
Variance in LLM/industry effects	0.7	0.8	0.7	0.6
Covariance in person and firm effects	13.5	14.2	12.0	9.6
Covariance in person and union effects	2.5	2.3	2.4	2.7
Covariance in person and LLM/industry effects	2.3	3.1	3.1	2.6
Covariance in firm and union effects	-0.5	-0.6	-0.5	-0.8
Adjusted R^2	0.908	0.914	0.925	0.927
Number of firms	991,423	551,715	581,934	560,761
(firm-only moves)	(5,960,686)	(3,189,732)	(2,598,592)	(2,403,694)
Number of unions	4,805	2,192	4,461	4,458
(both moves)	(9,575,586)	(5,146,155)	(3,573,323)	(3,284,332)

This table summarizes the results from estimating UF-AKM models for log earnings using person-year observations in the largest connected set of four samples: all Brazil, the Southeast region, an early period (2009-2013), and a late period (2013-2017). The specification estimated is provided in Equation (2) with the inclusion of local labor market and industry fixed effects.

The variance and covariance of person and firm effects account for 89-91% of inequality. Heterogerenity in transefarble skills is the largest factor, explaining 64.4-70.4% of the variance in log earnings (standard deviation of 49.9 to 54.0 log points). Heterogeneity in firm-wide pay policies explains 9.3-10.7% of inequality, with the standard deviation ranging from 19.6 to 20.4 log points. Positive assortative matching of workers to high paying firms is substantial (from 9.6% to 14.2%), sometimes accounting for more than the variance in firm effects. These results align with the general take-away of AKM variance decompositions in Brazil (Gerard et al., 2021; Engbom and Moser, 2022).

Unions account for an additional 3-4% of the variance in log earnings (summing the variance in unions with the covariance in person and union effects). As a share of overall

inequality, the variance in unions is not a decisive factor (1.0-1.2%), but the heterogeneity in union effects is still substantial with a standard deviation from 6.5-7.3 log points. Interestingly, positive assortative matching of workers to effective unions is a larger factor for overall inequality (2.3-2.7%).[17] A priori, this correlation may seem counterintuitive as prior literature shows that lower skill workers may benefit more from unions and therefore sort towards unionized workplaces. However, in our setting, unionization is by default with unions pre-assigned to workplaces. We explore the effect of unions on wage compression further in Section 5.

Finally, the role of local labor markets and industries combined is no larger than that of unions. As in the union case, positive assortative matching is a larger contributor to the variance in log earnings than the heterogeneity across LLMs and industries. Nonetheless, the inclusion of these fixed effects is relevant. Table A1 compares the variance decompositions for an AKM, a basic UF-AKM (i.e., without LLM and industry fixed effects), and the baseline UF-AKM in our connected set. The role of person effects remains essentially the same, but how the variance of establishment effects and positive assortative matching are split into firm and union effects varies depending on the inclusion of LLM and industry fixed effects. In terms of variance in establishment effects, the contribution of firms and unions shrinks from 76% and 12% in the basic UF-AKM to 73% and 8% in the baseline UF-AKM, respectively. In terms of covariance of worker and establishment effects, the contribution of firms and unions shrinks from 77% and 23% in the basic UF-AKM to 74% and 14% in the baseline UF-AKM, respectively.

We have documented the relationship between unions and overall inequality among unionized workers. But are these estimates actually capturing something about unions? We explore this question and delve deeper into union heterogeneity in the next section.

4 Heterogeneity in Union-Specific Pay Premia

While the UF-AKM variance decomposition shows that union heterogeneity is not a primary contributor to overall inequality among unionized workers in Brazil, the 7 log point standard deviation in union effects reveals meaningful heterogeneity in pay premia across unions. This section focuses on said heterogeneity, starting with a normalization proce-

[17] Figure B6 shows that this pattern is strongest in low-skill jobs among low-skill workers; a weaker pattern emerges in high-skill jobs among high-skill workers. This is particular to union effects, i.e., positive assortative matching of workers to firms is similar for low- and high-skill jobs, and holds across the skill distribution.

dure to interpret the UF-AKM's union effects as union pay premia. This normalization allows us to situate our estimated premia in the literature and to compare union premia in Brazil to those of other countries. Our normalization exercise also allows us to speak to unions' effects on between-group inequality in Brazil, which we do in Section 5.[18]

In the remainder of this section, we validate our estimated premia by documenting that a reform that weakened labor unions affected union pay premia. We further show that our premia are correlated with union attributes known to affect bargaining power, union policy, and the ability of unions to secure wage gains. We also uncover correlations for less explored union characteristics. The qualitative takeaway from these exercises does not depend on our specific normalization.

4.1 Interpretation of union pay premia

We define a union pay premium as the wage effects of unionization from a specific union relative to unionization from an ineffective union. We start with our estimated union effects from the UF-AKM, which are relative to some omitted union. We normalize these to a set of unions deemed to lack the operational capacity to influence wages ("ineffective unions"). As such, we rescale our union-specific pay premia $\hat{\eta}$ using a normalizing constant determined by the relation between union pay premia and log union revenue.

Our normalization procedure follows the method used to normalize AKM establishment effects in Card et al. (2016). The original approach aims to identify establishments with negligible rents using log value added per worker. Our approach seeks to identify unions with limited resources to impact wages using log union revenue. The rationale for this normalization is that there is some fixed cost τ for unions to influence wages, and unions without sufficient revenue R_u to meet those costs are ineffective. In other words, we assume $\mathbb{E}[\eta_u | R_u < \tau] = 0$.

Figure 2a depicts the relationship between union pay premia and log union revenue that pins down the normalizing constant. The scatterplot reveals an increasing relationship between bins of log union revenue (averaged across 2009-2017 GRCSU files) and the mean of the estimated UF-AKM union effects in the bin. The fitted line corresponds to a

[18] To explore the effect of unions on between-group inequality in Brazil in Section 5, we re-estimate the UF-AKM separately for each group of interest. Applying the same normalization procedure to each set of estimates allows us to estimate differences in the average premia across groups and thus to decompose the contribution of differential sorting from gaps in union premia.

Figure 2: Union Pay Premia: Interpretation and Distribution

(a) Normalization procedure

(b) Histogram of union pay premia

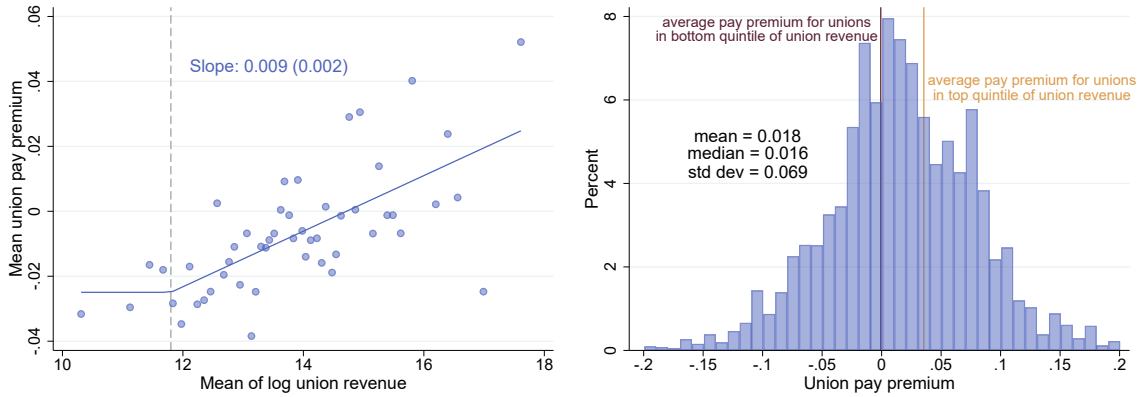


Figure 2a shows a binned scatter plot of the mean union effects (estimated from the UF-AKM) across bins of log union revenue (averaged across 2009-2017 GRCSU files). The fitted line corresponds to the OLS estimates from Equation (3) for the threshold τ (dotted vertical line) that maximizes the adjusted R-squared. The horizontal line represents the normalizing constant $\hat{\beta}_0$; the slope conditional on revenues above the threshold $\hat{\beta}_1$ is reported above the scatterplot. Figure 2b is a histogram of the union-specific pay premia. The normalization procedure affects the mean and median, but not the standard deviation. The vertical lines at 0.000 and 0.035 mark the mean union pay premia among unions in the bottom and top quintile of union revenue. The observation level in both figures are unique unions, weighting by person-year observations.

regression of the following form:

$$\hat{\eta}_u = \beta_0 + \beta_1 \max(0, R_u - \tau) + u_i, \tag{3}$$

where β_0 captures the normalizing constant for a given threshold τ, and β_1 represents the slope above τ. The threshold selected is the one maximizing the adjusted R-squared, using 0.10 increments along the x-axis. Results are essentially unchanged when using 0.01 increments or choosing the runner-up in maximizing R-squared.

The histogram of the normalized union pay premia is shown in Figure 2b. Under this normalization, the average union boosts pay by 1.8 log points. By construction, the mean union pay premium among low revenue unions (bottom quintile of revenue) is near zero. Consistent with the positive slope in Figure 2a, the mean union pay premium among high revenue unions (top quintile of revenue) is 3.5 log points. These magnitudes are closer to estimates of union premia from bargaining coverage in settings with widespread unionization—e.g., 2% in France (Breda, 2015), 2-4% in Germany (Hirsch and Mueller, 2020), and 1% in Portugal (Jäger et al., 2024)—than to estimates of the premia in Wagner-style settings like the U.S. or Canada. Notably, we find that the mean premium in the top quartile of union premia in Brazil is 10.7%, which is similar to the U.S.'s 10-20% (Farber et al., 2021).

The substantial heterogeneity across unions in terms of pay premia (standard deviation of 7 log points) is unaffected by the normalization. However, with the newly defined interpretation of the union effects, it is clear that some unions have a negative effect on wages. On the one hand, this could reflect compensating differentials. That is, unions that focus on non-wage amenities may be forced to dampen wages for workers—we find some evidence of this in Section 4.3. On the other hand, the institutional setting in Brazil is conducive to the existence of co-opted unions. In other words, since unions face little competition and have a stable source of revenue from mandatory contributions, some may ignore workers' interest and be more loyal to the interests of their leaders. These issues were among the stated motivations for a 2017 labor reform, which we now explore as a validation exercise for our estimates of union-specific pay premia.

4.2 Impact of weakening unions on labor markets

To validate that our UF-AKM union effects capture the impact of unions on wages, we leverage a 2017 labor reform that weakened unions across Brazil. Specifically, we show

how labor markets with varying union quality experienced differential wage effects from the reform that align with our interpretation of the union pay premia.

In November 2017, Law 13.467 came into effect, marking the most substantial overhaul to Brazil's labor laws since the 1940s. This labor reform introduced significant changes aimed at increasing labor market flexibility and reducing the influence of unions. For our purposes, the most relevant aspect of the reform was the abolition of mandatory unionization dues (Section 1.2). However, the reform also legalized outsourcing for core business functions, introduced provisions for mutual termination agreements, and established guidelines for remote work, among other changes.

The reform weakened unions by decimating their most reliable revenue source. By making unionization contributions voluntary, the reform created a free-rider problem similar to that introduced by right-to-work laws in the U.S. That is, with guaranteed union representation and universal coverage there is little incentive to contribute to one's union, even when one benefits greatly from its actions.[19] Figure 3a shows that the total revenue form unionization contributions decreased by 90% after the reform. This significant decline in revenues forced many unions to reduce staff and operations.

Difference-in-differences To assess the effect of the reform, we employ a difference-in-differences approach using the following specification:

$$y_{it} = \sum_{j \neq 0} \beta_j (D_{m(i)} \times \delta_{t=j}) + \mu_{m(i)} + \delta_t + \gamma X_{m(i)t} + \varepsilon_{it}, \tag{4}$$

where y_{it} is a time-varying outcome for worker i, e.g., log earnings; $D_{m(i)}$ is a dummy for labor markets where the average union pay premium is high; δ_t and $\mu_{m(i)}$ are quarter and labor market fixed effects, respectively; and $X_{m(i)t}$ are time-varying controls, e.g., state-by-quarter and sector-by-quarter fixed effects. The omitted quarter $j = 0$ is 2017q3, i.e., the quarter before the implementation of the labor reform.

We run this regression using quarterly household survey data (PNAD-Contínua).[20] These data maintain representativeness at the state-by-sector level, which we denote as a labor market. Hence, we take the average union pay premium at the labor market level in RAIS and merge this information into PNAD-Contínua.[21] To address noise in the premia,

[19] By leaving the union monopoly intact, the reform failed to introduce any incentive for workers to direct potential contributions to a union that best represents their interests.

[20] We do not have access to RAIS files after 2017.

[21] Figure B7a reveals that the more prosperous regions of the country have better unions on average.

especially at this level of aggregation, we inversely weight the regression in Equation (4) by the standard deviation in pay premia within a labor market. Standard errors are clustered at the state level, but results are robust to clustering at the labor market level.

Results The analysis focuses on the differential effects of the 2017 labor reform in markets with high vs. low union pay premia. In particular, the treatment dummy $D_{m(i)}$ equals one for labor markets in the top tercile of union pay premia and zero for those in the bottom tercile. From the 324 labor markets in our PNAD sample, all 5 geographic regions and 12 sectors are represented in the top and bottom terciles.

Figure 3b shows that weakening unions causes a 3.8% wage decrease in markets with high premium unions. The magnitude of the decline is reasonable, given that the average union pay premium is 1.8 log points. We are unable to reject parallel trends prior to the reform, supporting the identification assumption that wages would have evolved similarly in both markets had the reform not taken place. Interestingly, the effects are delayed by one-year, aligning with the expiration of most CBAs negotiated prior to the reform.

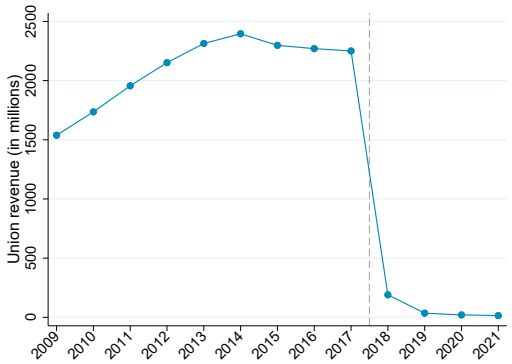
An advantage of PNAD-Contínua is that the survey includes information on employment by formality status, allowing us to explore the effects of weakening unions on this margin. Figure B8a shows that there are no meaningful employment effects despite lower wages in markets with high premium unions. Figure B8b reveals that weakening unions does not affect formality. This suggests that incentives for firms to hire formally given lower wages and more flexibility may be offset by the weakening of unions' role in enforcing formality in the workplace (Kohli, 2025). Future work should aim to further improve our understanding of unions' impact on formality, given that this is a critical margin in developing countries.

Our results indicate that our union premia estimates do capture union effects on pay since wages in labor markets with higher union pay premia declined strongly relative to those with lower union pay premia following the reform. Additionally, this exercise validates the use of union revenue as a normalization factor, given the reform's stark impact on wages after the expiration of extant CBAs.

Figure B7b shows that unions in the hospitality sector are ineffective on average while the sectors with the highest average union premia are extractive industries followed by manufacturing.

Figure 3: Transition from Mandatory to Voluntary Union Contributions

(a) Trends in union revenue



(b) Impact on log earnings

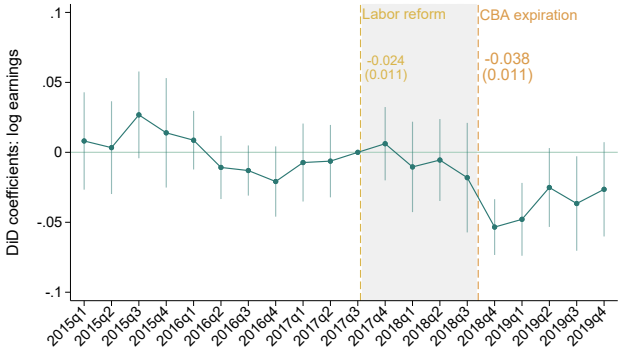


Figure 3a shows the trend in total revenues from unionization contributions based on annual GRCSU files. Prior to November 2017, these contributions were mandatory, i.e., deducted from workers' paychecks. Following the 2017 labor reform, unionization contributions became voluntary, which resulted in very few workers opting into making contributions. Figure 3b plots the dynamic DiD coefficients from Equation (4) using the PNAD-Contínua data and log earnings as the dependent variable. The treatment dummy $D_{m(i)}$ equals one for labor markets in the top tercile of union pay premia and equals zero for those in the bottom tercile. Pooled DiD estimates with standard errors for the period after the reform and the expiration of extant CBAs (denoted by the vertical dotted lines) are reported in the figure. Regressions use the standard weights provided by PNAD-Contínua, and we also inversely weight the regression by the standard deviation in union pay premia within a labor market. Standard errors are clustered at the state level.

4.3 Union attributes associated with pay premia

As further evidence that the UF-AKM captures unions' role in pay, we now explore the association between our estimated union pay premia and a rich set of union attributes. Though we use a data-driven approach to identify attributes that correlate with the premia, the results reported below align with existing evidence on key characteristics of effective unions. Additionally, with our extensive dataset on union characteristics, we shed light on understudied union characteristics and their importance for union wage effects.

Union attributes We build a novel dataset with rich information on union attributes by complementing our main RAIS-GRCSU sample with supplementary data sources, including union, CBA, and strike registries (see Section 2.1 for details). For exposition purposes, we group these attributes into three broad categories: 1) union type; 2) union actions; and 3) governance and leadership. Refer to our Online Appendix for precise definitions on these union attributes.

The majority of unions in the UF-AKM connected set are local, industry-based unions that offer coverage at the municipality level (Table A2). In other words, few unions are federations (confederations) that represent a category of workers at the state (national) level, or unions whose category pertains to an occupation rather than the establishment's main economic activity. Most of them are affiliated with some union central, with *Força Sindical* and *CUT* ranking the highest in terms of number of unions affiliated. There is also substantial variation in revenue and workers covered. The mean union density is 23% but the distribution is right-skewed, e.g., the median membership rate is only 6%.

Only 22% of unions in our sample engage in any strike during the sample period, while 93% negotiate at least one CBA (Table A3). The average number of firm-level CBAs negotiated by unions surpasses the average number of sectoral CBAs by about 65%. On average, these CBAs have a length of 2,786 words and 28.2 clauses. Based on the amenity value estimates of firm-level CBAs from Lagos (2024), the non-wage amenities negotiated have an average value equivalent to a 5 log point increase in wages. Out of the 24 clause subgroups provided in the CBA registry, the four most common ones negotiated are those regulating the workday (e.g., shift scheduling), bargaining (e.g., fines for CBA non-compliance), assistances (e.g., transport vouchers), and wage adjustments (e.g., wage floors).

Collective governance, internal competition, and parity in union boards are uncommon (Table A4). Only 5% of unions operate under a collective leadership structure, as

opposed to a presidential model, and just 13% mandate the convening of a general assembly before leadership elections. The participation rate in union elections is 64% on average, but the vote share of the winning slate of candidates is extremely high (IQR of 89-99%). In fact, nearly all union elections have a single slate of candidates, meaning that there is little internal competition for union leadership. There is substantial variation in the share of union board members who are female, nonwhite, low education, and blue collar. Remarkably, a minority of unions have boards with parity in representation of these worker groups. Lastly, a standardized average of the union board members' AKM worker effects reveals a wide range of skill across union leadership (IQR of -60 to 54 log points).

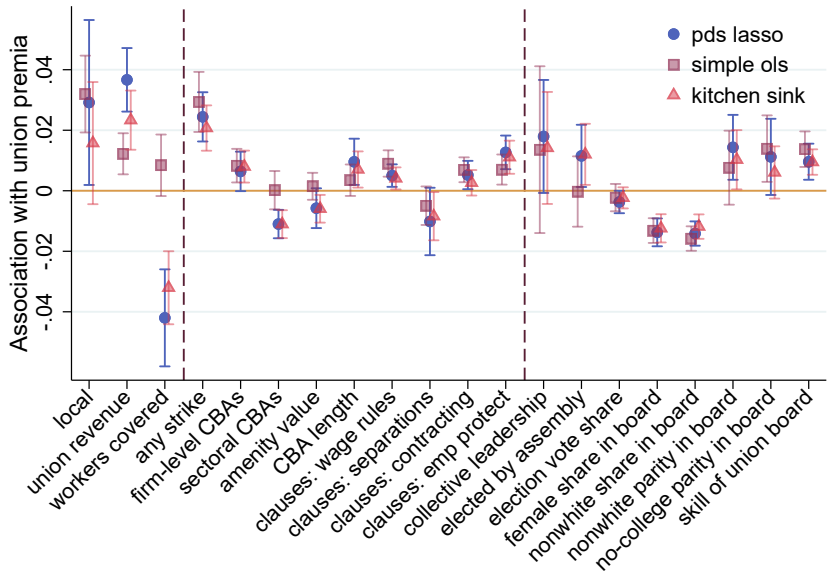
Correlation results To provide a summary of the relevant associations between union-specific pay premia and the 56 different attributes, we first reduce the dimensionality of attributes. Specifically, we run an adaptive LASSO of the premia on the attributes, ensuring all continuous attributes are normalized. Out of the 56 attributes, 27 are selected.

We run two correlational exercises with the selected attributes. First, we run a kitchen-sink regression of the premia on each attribute, adding all other selected attributes as controls. Next, to better address omitted variable bias, we estimate a post-double selection LASSO. The PDS LASSO consists of 1) running a LASSO of the attribute of interest on all other attributes; 2) running a LASSO of the premia on other attributes; and 3) running an OLS of the premia on the attribute of interest, adding as controls all the selected attributes in steps (1) and (2). For reference, we also report simple bivariate OLS estimates.

Figure 4 shows the estimates of the coefficient pertaining to each attribute (shown on the x-axis) from the three correlational exercises described above. Only attributes with a statistically significant estimate are shown in the figure (17 out of 27). The dotted vertical lines partition the attribute space into three categories: a) union type; b) union actions; and c) governance and leadership. Our discussion of the results focuses on the PDS LASSO estimates, but the kitchen-sink OLS estimates are directionally similar. The results are robust to including state and sector fixed effects, which is expected since the UF-AKM already controls for LLMs and industries (Figure B9).

The associations for "union type" attributes show that local unions with a small number of workers and large revenues are associated with higher pay premia. Representation from a local union, as opposed to higher-level entities (i.e., federations and confederations), could be linked to better pay since the union is likely to have a better sense of local

Figure 4: Correlates of Union Pay Premia



This figure shows the estimates of coefficients correlating union-specific pay premia with union attributes (denoted in the x-axis). For each attribute, estimates for three separate analyses are reported: 1) *kitchen sink:* a kitchen-sink regression of the premia on the attribute of interest, adding all other selected attributes as controls; 2) *pds lasso:* a post-double selection LASSO to partially address omitted variables bias with observed attributes; and 3) *simple ols:* a simple bivariate OLS regression of the premia on the attribute of interest. Figure 4 only reports the attributes that have a statistically significant estimate. The dotted vertical lines partition the attribute space into three categories: a) union type; b) union actions; and c) governance and leadership. All regressions are run with unique union observations, weighting by person-year observations.

conditions. Similarly, if there are fewer workers that the union represents, coordinating on collective actions is potentially easier. These results are consistent with prior literature on constituency size and homogeneity (Alesina and La Ferrara, 2000; Farber, 2001). More resources to mobilize workers and exert pressure are possibly behind the positive association between union revenues and pay premia, consistent with the evidence shown in Section 4.2 and with recent work showing that right-to-work laws reduce wages and union density (Fortin et al., 2023). Lastly, given that union membership in this context does not proxy for worker support, it is not surprising that union density is not correlated with higher pay premia.

The associations for "union action" attributes reveal that unions engaging in strikes and collective bargaining boost workers' wages. This finding is consistent with a large literature on the efficacy of strikes in securing wage gains (see Card (1990) for a review; see Massenkoff and Wilmers (2024) and Lyon et al. (2024) for more recent evidence). In terms of CBAs, it is the unions that negotiate many firm-level agreements that are linked to higher premia. In fact, unions negotiating many sectoral CBAs are associated with lower premia, which possibly reflects the fact that unions negotiating these types of agreements are not local. We find suggestive evidence of compensating differentials, since there is a marginally negative association between union pay premia and amenity value. Turning to the CBA clauses, we find that wage rules, contracting, and employment protections are associated with higher pay premia, while clauses on separations are linked to lower pay premia. Among these, the correlation with wage rules is the least surprising since these clauses are often concerned with normative salaries and equalizing pay.

The associations for "governance and leadership" attributes show that collective governance, internal competition, and skilled leadership are linked to higher pay premia. First, collective leadership structures and convening a general assembly before elections are positively correlated with union pay premia. Second, the marginally significant negative correlation with the election vote share could be interpreted as the presence of internal union competition providing incentives for the leadership to deliver on wages. In terms of board composition, a high share of leaders who are either women or nonwhite is associated with lower pay premia. By contrast, nonwhite and low education *parity* in leadership are linked to higher pay premia, suggesting that a balanced representation of low-wage workers in union leadership pushes unions to focus on raising pay (Budde et al., 2024). Finally, union boards with high skilled leaders are associated with securing

higher wages.[22]

In short, many of the correlations uncovered in this analysis align with prior research, which further validates our interpretation of the UF-AKM estimates. Importantly, the associations also point to promising directions for future research on unions, including collective governance structures and internal competition.

5 Unions and Between-Group Inequality

The UF-AKM assumes that all workers represented by a given union experience the same wage premium, on average. However, it is well-documented that union effects are heterogeneous for different groups of workers.[23] In this section, we explore how unions affect wage gaps between groups of workers. We then characterize unions that bargain differentially for high versus low educated workers in order to understand heterogeneity in wage compression effects across unions. We end with a discussion of trade-offs that unions may face between increasing wages and compressing the wage distribution while maintaining worker support through voluntary contributions.

5.1 Wage gap decompositions

The AKM has been used to analyze the role of establishments in explaining wage gaps between groups (Card et al., 2016; Gerard et al., 2021; Guo et al., 2025). Estimating separate AKM models for each group, one can decompose the extent to which the gaps in the overlapping connected sets are explained by worker effects versus establishment effects. Furthermore, one can implement a Oaxaca-Blinder decomposition to determine how much of the role played by employers is driven by cross-establishment sorting, as opposed to within-establishment relative pay setting.

The UF-AKM can be used in an analogous way to determine the role of unions in explaining wage gaps between workers in groups $g \in \{H, L\}$. Using the dual connected set from separate UF-AKM models estimated on each worker group, the wage gap decom-

[22] Boudreau et al. (2025) explores the traits of union leaders and finds that they are positively selected along a wide range of characteristics, including grit, locus of control, and prosociality, potentially linking these traits to the ability to secure better collective outcomes.

[23] See, e.g., Ashenfelter (1972), Farber et al. (2021), and Biasi and Sarsons (2022).

position (ignoring the covariates X_{it}) amounts to:

$$\mathbb{E}\left(y_{it}^{H}\right)-\mathbb{E}\left(y_{it}^{L}\right)=\underbrace{\left(\bar{\alpha}^{H}-\bar{\alpha}^{L}\right)}_{\text{transferable skills}}+\underbrace{\sum_{j}\phi_{f(j)}^{H}\left(\pi_{j}^{H}-\pi_{j}^{L}\right)}_{\text{cross-firm sorting}}+\underbrace{\sum_{j}\left(\phi_{f(j)}^{H}-\phi_{f(j)}^{L}\right)\pi_{j}^{L}}_{\text{within-firm relative pay setting}}$$
$$+\underbrace{\sum_{j}\eta_{u(j)}^{H}\left(\pi_{j}^{H}-\pi_{j}^{L}\right)}_{\text{cross-union sorting}}+\underbrace{\sum_{j}\left(\eta_{u(j)}^{H}-\eta_{u(j)}^{L}\right)\pi_{j}^{L}}_{\text{within-union differential bargaining}}$$

where π_j^g is the proportion of workers from group g in establishment j. As is common in this literature, the sorting components assign the premia of the favored group H to the firms and unions where each group works, while the within-firm and within-union components focus on the gap in pay premia ($H - L$) that exists where the marginalized group L works. As such, only the latter components are sensitive to the normalizing procedure described in Section 4.1.[24]

We implement this wage gap decomposition for three pairs of worker groups: 1) high vs. low education (defined as above high school vs. high school or less); 2) male vs. female; and 3) white vs. nonwhite. Table A5 contains the variance decomposition of the UF-AKM for each of the six worker groups. To account for differential geographic and sectoral distribution between the worker groups, we reweight the marginalized group using DFL weights by LLM-by-sector (DiNardo et al., 1996; Gerard et al., 2021).[25]

The results from the wage gap decomposition by education, gender, and race are in Table 2. The average wage gaps shown in Column (1) are large, especially by education (71.2 log points). We find that firms exacerbate these wage gaps—ranging from 10% to 37% of the average gap, as per Column (2).[26] Columns (3) and (4) reveal that within-firm relative pay setting effects dominate cross-firm sorting effects in the education and gender wage gaps, but not the racial wage gaps.

Unions also exacerbate the wage gaps by education, gender, and race but to a lesser degree than firms—ranging from 2.6% to 7.0% of the average gap, as per Column (5). Interestingly, the gaps in union pay premia are driven by within-union differential bargaining rather than cross-union sorting. Specifically, the effects in Column (6) are at least

[24] Following the literature, we normalize firm effects relative to the average pay premium in the restaurant industry, which is a sector with negligible rents.

[25] Figure B10 shows the normalization procedure of the union effects for the three pairs of worker groups, as well as how the distribution of the union pay premia are affected by the DFL weights.

[26] Our results by gender and race align with the existing literature (Card et al., 2016; Gerard et al., 2021).

Table 2: Role of Firms and Unions in Wage Gaps Among Unionized Workers

			Decomposition			Decomposition	
	Overall wage gap	Gap in firm effects	Relative pay setting	Cross-firm sorting	Gap in union effects	Differential bargaining	Cross-union sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. High vs. low education							
Average statistic	0.712	0.148	0.102	0.046	0.019	0.013	0.006
		20.7%	*14.3%*	*6.5%*	*2.6%*	*1.8%*	*0.9%*
Panel B. Male vs. female							
Average statistic	0.192	0.071	0.049	0.022	0.013	0.009	0.004
		36.9%	*25.4%*	*11.6%*	*7.0%*	*4.7%*	*2.3%*
Panel C. White vs. nonwhite							
Average statistic	0.161	0.016	0.004	0.012	0.007	0.005	0.002
		10.0%	*2.8%*	*7.3%*	*4.5%*	*3.3%*	*1.2%*

This table shows the education, gender, and racial wage gaps in Column (1), along with the contribution of firm and union effects to these gaps in Columns (2) and (5), respectively. The Oaxaca-Blinder decomposition of the gaps in firm and union pay premia (see Section 5.1) are provided, showing within-firm / within-union pay effects in Columns (3) and (6) as well as across-firm / across-union sorting effects in Columns (4) and (7). The percent of the overall wage gap attributed to Columns (2)-(7) are reported under each statistic. The sample in each panel is restricted to the dual connected set from the UF-AKM connected sets of the two worker groups, i.e., high vs. low education (defined as above high school vs. high school or less) in Panel A, male vs. female in Panel B, and white vs. nonwhite in Panel C. Workers in the marginalized group are reweighted using DFL weights by LLM-by-sector cells.

twice as large as those in Column (7). These results are robust to alternative normalization procedures for the union pay premia (Table A6).

While the cross-union sorting effects we estimate align with the positive worker-to-union assortative matching documented in Section 3.4, the differential bargaining effects contrast with prior studies showing that unions compress wages by skill group and gender (Farber et al., 2021). We note however, that this modest exacerbation of gaps reflects the impact of unions on average in a setting with default unionization. In fact, there is vast heterogeneity across unions in differential bargaining effects.[27] In the next section, we focus on differential bargaining by education and what union attributes explain it as compression by skill group is a key focus of the unions literature (Jäger et al., 2024).

5.2 Within-union differential bargaining by worker education

On average, differential bargaining for high versus low educated workers by unions increases the wage gap between these groups by 1.3 log points (Table 2). However, as shown in Figure 5a, underlying this average effect is substantial heterogeneity in how each individual union affects these gaps. Specifically, the standard deviation of the union-specific differential bargaining effects by education is 7.2 log points, which is slightly larger than the overall dispersion in union-specific pay premia.

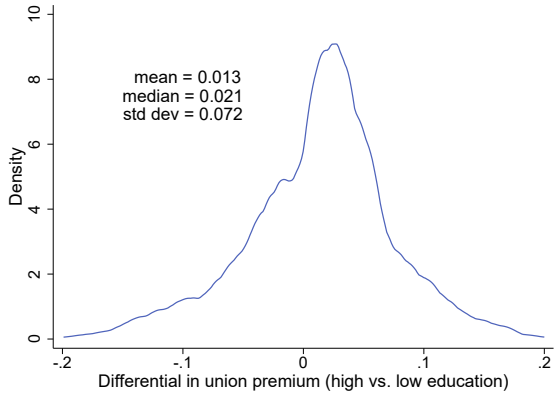
We next show that these union-specific differential bargaining estimates correlate with union attributes in meaningful ways. Following the analysis in Section 4.3, Figure B11 shows that unions that compress wage gaps by education negotiate wage adjustment clauses—including wage floors, wage increases, and wage equalization—and have leadership that is more blue collar, less educated, and nonwhite. Furthermore, unions affiliated to the *Central dos Sindicatos Brasileiros* (CSB) exacerbate these gaps. Interestingly, CSB is a union central that had strong ties with the political party of Michel Temer (PMBD) who pushed for the 2017 Labor Reform as president. It is a union central known for taking a more cooperative stance with employers, hence it is reasonable to posit that its affiliates are less likely to fight for raising wages, particularly at the bottom of the wage distribution.

Importantly, Figure 5b shows that the unions securing high pay premia tend to compress wage gaps by education. In other words, unions that actively fight for raising wages

[27] The standard deviation of union effects on education, gender, and race gaps is 7.2, 6.1, and 5.8 log points, respectively.

Figure 5: Heterogeneity in Union-Specific Differential Bargaining

(a) Distribution of differential bargaining



(b) Wage gains vs. wage compression

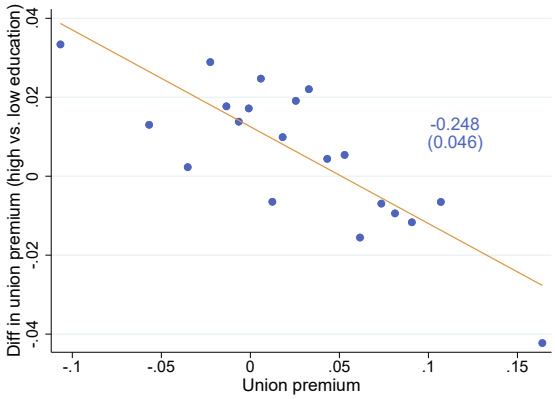


Figure 5a depicts the kernel density of the differential bargaining effects by education, i.e., $(\eta^H_{u(j)} - \eta^L_{u(j)})\pi^L_j$, applying DFL weights by LLM-by-sector on low education workers. The mean, median, and standard deviation of the distribution are reported next to the kernel density. The normalization procedure affects the mean and median, but not the standard deviation. Figure 5b is a binned scatterplot of the differential bargaining effects by education on the union pay premium, including the linear prediction from an OLS. The estimate for the slope coefficient and its standard error are reported next to the scatterplot. Samples are restricted to the dual connected set from the UF-AKMs of the high and low educated workers. The observation level in both figures are unique unions, weighting by DFL-reweighted low education person-year observations.

are the ones that compress the wage distribution.[28] In Wagner-style settings like the U.S., it is likely that most of the unions that obtain certification are those on right-hand side of Figure 5b. That is, since workers have to organize and overcome numerous obstacles in order to form a union, the existing unions are a positively selected sample of the ones that would arise if unionization were far more widespread and with limited external competition.

5.3 Discussion: worker support for unions

Our findings highlight substantial heterogeneity across unions in terms of wage effects and how such gains are shared with different worker groups. Our estimates come from a period in Brazil where unions generally did not depend on worker support for funding, i.e., unionization dues were mandatory for all workers represented by each union. However, following the 2017 labor reform, buy-in from workers became a relevant feature since workers had to opt-in to continue these contributions. As such, the reform provides an opportunity to characterize the type of unions that enjoy worker support by analyzing heterogeneous effects of different unions' ability to maintain these contributions.

We are interested in determining whether unions that secure higher pay premia or bargain differentially across education groups have more worker support. Specifically, we create two categorical variables: one for union premia (i.e., low-mid-high) and one for differential bargaining (i.e., compress-balanced-exacerbate). Following Section 4.2, we then estimate a difference-in-differences specification, in which we interact an indicator for years after 2017 with two sets of dummy variables—categories of unions by their pay premia ($\hat{\eta}_u$) and their differential bargaining effects by education ($\hat{\eta}_u^H - \hat{\eta}_u^L$)—and their interactions. The specification also controls for year and union fixed effects, as well as state-by-year and sector-by-year fixed effects.

Motivated by the shape of the distributions of union premia (Figure 2b) and differentials by education (Figure 5a), our categories are defined in levels. Specifically, low-premium and high-premium unions are those with $\hat{\eta}_u$ below -2 log points and above 2 log points, respectively; mid-premium union are those with $\hat{\eta}_u$ falling in between. Similarly, balanced-bargaining unions are those with $\hat{\eta}_u^H - \hat{\eta}_u^L$ between -2 and 2 log points.

[28] The complementarity between wage gains and wage compression by education does not apply to other forms of wage compression. Figure B12a shows that unions that raise wages also exacerbate gender gaps while Figure B12b reveals no statistically significant effect on racial gaps. Hence, the extent to which unions face trade-offs between wage gains and compression depends on the between-group inequality in question.

Table 3: Worker Support of Unions After the Labor Reform

Panel A	Among compress	Among balanced	Among exacerbate
Union premium	(1a)	(2a)	(3a)
High vs. low	0.027	0.083	0.030
	(0.020)	(0.031)	(0.019)
Mid vs. low	0.016	0.083	0.049
	(0.024)	(0.036)	(0.021)
Panel B	Among high	Among mid	Among low
Differential bargaining	(1b)	(2b)	(3b)
Compress vs. balanced	-0.094	-0.104	-0.037
	(0.021)	(0.032)	(0.030)
Exacerbate vs. balanced	-0.075	-0.057	-0.022
	(0.021)	(0.030)	(0.029)

This table summarizes the results from the following DiD regression: $y_{ut} = \beta_1(Post_t \times Low_u) + \beta_2(Post_t \times High_u) + \beta_3(Post_t \times Compr_u) + \beta_4(Post_t \times Exacer_u) + \beta_5(Post_t \times Low_u \times Compr_u) + \beta_6(Post_t \times Low_u \times Exacer_u) + \beta_7(Post_t \times High_u \times Compr_u) + \beta_8(Post_t \times High_u \times Exacer_u) + \mu_u + \delta_t + \gamma X_{ut} + \varepsilon_{ut}$, where y_{ut} is an indicator for union u receiving any contributions in year t; $Post_t$ is an indicator for years after 2017; δ_t and μ_u are year and union fixed effects, respectively; and X_{ut} are state-by-year and sector-by-year fixed effects. Standard errors are clustered at the union level. Panel A shows the worker support effects for high and mid premium unions (relative to low premium unions) among each of the three categories of differential bargaining. Panel B shows the worker support effects for unions that compress and exacerbate wage gaps by education (relative to unions engaging in balanced bargaining) among each of the three categories of union premia. In both the union premia (low-mid-high) and differential bargaining (compress-balanced-exacerbate) categorical variables, we define their categories as follows: below -2 log points, between -2 and 2 log points, and above 2 log points in union premia and differential bargaining by education, respectively. Estimates for β_1-β_8 are reported in Table A7.

Unions with $\hat{\eta}_u^H - \hat{\eta}_u^L$ below -2 log points are said to compress wage gaps while those with $\hat{\eta}_u^H - \hat{\eta}_u^L$ above 2 log points exacerbate those gaps.[29]

Table 3 summarize the DiD estimates by reporting the implied treatment effects for high- and mid-premium unions (relative to low-premium unions) among each of the three differential bargaining groups, and those for unions that compress and exacerbate wage gaps (relative to those engaging in balanced bargaining) among each of the three pay premia groups. The results in show that high- and mid-premium unions were more likely to continue to receive voluntary unionization contributions, particularly if they engaged in balanced bargaining—a relative increase of 8.3 percentage points. Relatedly, unions exacerbating or compressing wage gaps were less likely to receive contributions after the 2017 reform—a relative decrease of 5.7-10.4 and 7.5-9.4 percentage points among mid-premium and high-premium unions, respectively.[30]

In summary, this analysis suggests that workers support unions that increase wages. This is perhaps not surprising. The interesting lesson, however, is that worker support is stronger when unions are not too favorable towards one group of workers. Hence, unions appear to face a trade-off between compressing wages and maintaining worker support. To put it more bluntly, worker solidarity has its limits.

6 Conclusion

We link data on the specific union representing workers at each establishment in Brazil to employer-employee matched data to provide the first estimates of union-specific pay premia. Using an augmented AKM we term the union-firm AKM, or UF-AKM, we show that unions account for a notable share of the variance in earnings (3-4 log points) among unionized workers, and the standard deviation in union wage effects is large (6-7 log points). A nation-wide right-to-work law significantly lowered wages in markets with higher vs. lower average estimated premia, underscoring that our approach captures union effects on pay. Unions that are small and local, go on strike, and have skilled leadership all tend to have higher premia. High premium unions also reduce between-group inequality, but after right-to-work, such unions face a tradeoff between compression and worker support.

[29] Results are comparable if we define groups by terciles of $\hat{\eta}_u$ and $\hat{\eta}_u^H - \hat{\eta}_u^L$ instead (see Table A8).

[30] Figure B13 provide complementary evidence for these effects by estimating difference-in-differences coefficients for each year from 2013 to 2021, failing to reject the parallel trends assumption.

We show that in settings where data and institutional details allow, an AKM model including separate firm and union effects captures establishment effects on pay and can generate estimates of union-specific pay premia. Brazil combines rich administrative data on earnings and unions with default unionization that limits establishment sorting into unions. The average premium we estimate in this setting is similar to that of other countries with widespread union representation and state-mandated coverage—around 2%. Yet we can also characterize high premium unions that are arguably similar to the more proactive unions in Wagner-style settings such as the U.S. and Canada.

Our framework can also be used to understand union effects on outcomes beyond pay, such as hours or workplace amenities. Further, although we adapt the AKM to study how specific labor unions influence pay, a similar approach could be used to study the effect of employer associations or other labor market institutions that span establishments. Finally, we hope that our rich data on unions can spur further research into the characteristics of unions that not only lead to greater pay premia but that also condition worker support.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, 67(2):251–333.

Alesina, A. and La Ferrara, E. (2000). Participation in Heterogeneous Communities. *The Quarterly Journal of Economics*, 115(3):847–904.

Arellano-Bover, J. and San, S. (2020). The Role of Firms in the Assimilation of Immigrants. *Available at SSRN 3594778*.

Ashenfelter, O. (1972). Racial Discrimination and Trade Unionism. *Journal of Political Economy*, 80(3, Part 1):435–464.

Azkarate-Askasua, M. and Zerecero, M. (2023). Union and Firm Labor Market Power. *Available at SSRN 4323492*.

Baker, M., Halberstam, Y., Kroft, K., Mas, A., and Messacar, D. (2024). The Impact of Unions on Wages in the Public Sector: Evidence from Higher Education. Technical report, National Bureau of Economic Research.

Bassier, I. (2022). Collective bargaining and spillovers in local labor markets. Centre for Economic Performance Discussion Paper Series 1895, London School of Economics and Political Science. Centre for Economic Performance.

Beauregard, P.-L., Lemieux, T., Messacar, D., and Saggio, R. (2024). Why Do Union Jobs Pay More? Evidence from Matched Employer-Employee Data. Technical report, Unpublished manuscript.

Biasi, B., Cullen, Z. B., Gilman, J. H., and Roussille, N. (2025). How Does Wage Inequality Affect the Labor Movement? Technical report, National Bureau of Economic Research.

Biasi, B. and Sarsons, H. (2022). Flexible wages, bargaining, and the gender gap. *The Quarterly Journal of Economics*, 137(1):215–266.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739.

Boudreau, L., Macchiavello, R., Minni, V., and Tanaka, M. (2025). Leaders in Social Movements: Evidence from Unions in Myanmar. *American Economic Review*, 115(6):1975–2000.

Breda, T. (2015). Firms' Rents, Workers' Bargaining Power and the Union Wage Premium. *The Economic Journal*, 125(589):1616–1652.

Budde, J., Dohmen, T., Jäger, S., and Trenkle, S. (2024). Worker Representatives. *IZA Working Paper No. 17152*.

Bureau of Labor Statistics (2025). Union members — 2024. News Release USDL-25-0105, U.S. Department of Labor.

Card, D. (1990). Strikes and Bargaining: A Survey of the Recent Empirical Literature. *The American Economic Review*, 80(2):410–415.

Card, D. (1996). The Effect of Unions on the Structure of Wages: A Longitudinal Analysis. *Econometrica*, pages 957–979.

Card, D. (2001). The Effect of Unions on Wage Inequality in the U.S. Labor Market. *ILR Review*, 54(2):296–315.

Card, D. (2011). The Return of the Firm to Labor Economics. In *Presidential Address presented at the annual conference of the Society of Labor Economists, Vancouver*.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics*, 36(S1):S13–S70.

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.

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.

Card, D., Rothstein, J., and Yi, M. (2023). Location, Location, Location. *NBER Working Paper no. 31587*.

Card, D., Rothstein, J., and Yi, M. (2024). Industry Wage Differentials: A Firm-Based Approach. *Journal of Labor Economics*, 42(S1):S11–S59.

Cole, D. L. (1969). The AFL-CIO's Internal Disputes Plan. *Monthly Lab. Rev.*, 92:12.

Commons, J. R. (1921). *Trade Unionism and Labor Problems*. Selections and documents in economics, ed. by W. Z. Ripley. Ginn and Company, Boston, New York [etc.].

Corradini, V., Lagos, L., and Sharma, G. (2025). Collective Bargaining for Women: How Unions Can Create Female-Friendly Jobs. *The Quarterly Journal of Economics*, 140(3):2053–2105.

Di Addario, S., Kline, P., Saggio, R., and Sølvsten, M. (2023). It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages. *Journal of Econometrics*, 233(2):340–374.

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5):1001–44.

DiNardo, J. and Lee, D. S. (2004). Economic Impacts of New Unionization on Private Sector Employers: 1984–2001. *The Quarterly Journal of Economics*, 119(4):1383–1441.

Dodini, S., Salvanes, K., and Willén, A. (2024). The Dynamics of Power in Labor Markets: Monopolistic Unions versus Monopsonistic Employers. *CESifo Working Paper No. 9495*.

Drakopoulos, S. and Katselidis, I. (2014). The Development of Trade Union Theory and Mainstream Economic Methodology. *Journal of Economic Issues*, 48(4):1133–1149.

Engbom, N. and Moser, C. (2022). Earnings inequality and the minimum wage: Evidence from brazil. *American Economic Review*, 112(12):3803–47.

Farber, H. S. (1978). Individual Preferences and Union Wage Determination: The Case of the United Mine Workers. *Journal of Political Economy*, 86(5):923–942.

Farber, H. S. (2001). Union success in representation elections: Why does unit size matter? *ILR Review*, 54(2):329–348.

Farber, H. S., Herbst, D., Kuziemko, I., and Naidu, S. (2021). Unions and inequality over the twentieth century: New evidence from survey data. *The Quarterly Journal of Economics*, 136(3):1325–1385.

Fortin, N. M., Lemieux, T., and Lloyd, N. (2023). Right-to-Work Laws, Unionization, and Wage Setting. In *Research in Labor Economics: 50th Celebratory Volume*, pages 285–325. Emerald Publishing Limited.

Frandsen, B. R. (2021). The Surprising Impacts of Unionization: Evidence from Matched Employer-Employee Data. *Journal of Labor Economics*, 39(4):861–894.

Freeman, R. B. and Medoff, J. L. (1984). *What Do Unions Do?* Basic Books, New York.

Gerard, F., Lagos, L., Severnini, E., and Card, D. (2021). Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil. *American Economic Review*, 111(10):3418–57.

Gregg Lewis, H. (1986). Union Relative Wage Effects. In *Handbook of Labor Economics*, volume 2, chapter 20, pages 1139–1181. Elsevier.

Guo, N., Zhang, L., Zhang, R., and Zou, B. (2025). Migration Restrictions and the Migrant-Native Wage Gap: The Roles of Wage Setting and Sorting. *Available at SSRN*.

Hirsch, B. and Mueller, S. (2020). Firm Wage Premia, Industrial Relations, and Rent Sharing in Germany. *ILR Review*, 73(5):1119–1146.

Horn, C. H. (2009). *Collective Bargaining in Brazil: A Study on Joint Regulation of the Employment Relationship in Manufacturing*. VDM Verlag.

Jäger, S., Naidu, S., and Schoefer, B. (2024). Collective Bargaining, Unions, and the Wage Structure: An International Perspective. In *Handbook of Labor Economics*, volume 5. Elsevier. Working Paper.

Kaplan, E. and Naidu, S. (2025). Between Government and Market: The Political Economics of Labor Unions. *Annual Review of Economics*, 17.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Kline, P. M. (2024). Firm wage effects. Working Paper 33084, National Bureau of Economic Research.

Kohli, N. (2025). The Role of Unions in Labour Markets with Informality: Evidence from Brazil. *Working Paper*.

Krislov, J. (1955). The Extent and Trends of Raiding among American Unions. *The Quarterly Journal of Economics*, 69(1):145–152.

Krislov, J. (1991). The AFL-CIO's Effort to Minimize Union Membership Conflicts: 1962-1987. *Lab. Stud. J.*, 16:3.

Lagos, L. (2024). Union Bargaining Power and the Amenity-Wage Tradeoff. *IZA Working Paper No. 17034*.

Lee, D. S. and Mas, A. (2012). Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961–1999. *The Quarterly Journal of Economics*, 127(1):333–378.

Lemieux, T. (1998). Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection. *Journal of Labor Economics*, 16(2):261–291.

Lester, R. A. (1941). *The Economics of Labor*. Macmillan, New York.

Lewis, H. G. (1986). *Union Relative Wage Effects: A Survey*. University of Chicago Press, Chicago.

Lumsden, K. and Petersen, C. (1975). The Effect of Right-to-Work Laws on Unionization in the United States. *Journal of Political Economy*, 83(6):1237–1248.

Lyon, M. A., Kraft, M. A., and Steinberg, M. P. (2024). The Causes and Consequences of US Teacher Strikes. Technical report, National Bureau of Economic Research.

Massenkoff, M. and Wilmers, N. (2024). Economic Outcomes of Strikers in an Era of Weak Unions. *Journal of Labor Economics*, 42(1):25–51.

McCarten, C. (2021). Union Local Size. Blog post.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*, 105:142.

National Labor Review Board (2024). Election results. Fiscal year report.

Nickell, S. J. and Andrews, M. (1983). Unions, Real Wages and Employment in Britain 1951-79. *Oxford Economic Papers*, 35:183–206.

Rees, A. (1962). *The Economics of Trade Unions*. University of Chicago Press, Chicago.

Smith, A. (1776). *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Number v. 1 in An Inquiry Into the Nature and Causes of the Wealth of Nations. W. Strahan; and T. Cadell, in the Strand.

Online Appendix

A Appendix Tables

Table A1: Comparison of AKM and UF-AKM

	AKM Connected Set	UF-AKM Connected set		
	AKM	AKM	Basic UF-AKM	UF-AKM
	(1)	(2)	(3)	(4)
Mean log earnings	2.204	2.219	2.218	2.218
Standard deviation of log earnings	0.625	0.621	0.621	0.621
Standard deviation (across person-year observations)				
Person effects	0.488	0.494	0.500	0.499
Establishment effects	0.253	0.237	-	-
Percentage of variance in log earnings due to				
Variance in person effects	60.9	63.1	64.8	64.4
Variance in establishment effects	16.3	14.6	-	-
Covariance in person and establishment effects	18.2	18.4	-	-
Variance in establishment effects captured by				
Variance in firm effects	-	-	75.9	73.2
Variance in union effects	-	-	12.4	8.4
Covariance in establishment and person effects captured by				
Covariance in person and firm effects	-	-	77.4	73.8
Covariance in person and union effects	-	-	22.3	13.8
Adjusted R^2	0.911	0.910	0.908	0.908
Adjusted R^2 from match effects model	0.935	0.932	-	-
Number of establishment	2,710,696	1,333,355	1,364,905	1,364,905
Number of person-year observations	152,872,586	117,107,927	117,314,384	117,314,384

This table summarizes the results from estimating AKM and UF-AKM models using person-year observations from Brazil (2009-2017) in two connected set of establishments: 1) all moves; and 2) "firm-only" moves with the corresponding unions from "both" moves, excluding "union-only" movers. AKM refers to the specification in Equation (1). Basic UF-AKM refers to the specification in Equation (2). UF-AKM refers to the specification in Equation (2) with the inclusion of local labor market and industry fixed effects.

Table A2: Descriptive Statistics of Union Attributes: Union Type

	Mean	Std dev	Q1	Q2	Q3
Union premium	0.018	0.069	-0.020	0.016	0.060
Union type					
Local union	0.96	0.19	1.00	1.00	1.00
Industrial union	0.84	0.37	1.00	1.00	1.00
Municipal coverage	0.73	0.44	0.00	1.00	1.00
Força Sindical central	0.20	0.40	0.00	0.00	0.00
CUT central	0.19	0.39	0.00	0.00	0.00
UGT central	0.17	0.37	0.00	0.00	0.00
NCST central	0.15	0.35	0.00	0.00	0.00
CSB central	0.06	0.24	0.00	0.00	0.00
CTB central	0.06	0.23	0.00	0.00	0.00
Log union revenue	14.00	1.51	13.13	14.02	14.94
Log workers covered	6.27	2.23	4.87	6.75	7.83
Union density	0.23	0.34	0.02	0.06	0.26

This table displays descriptive statistics for attributes describing the type of union. The sample comprises unions in the UF-AKM connected set, i.e., 4,805 unions (Table 1). The first row reports the descriptive statistics for the union-specific pay premia, using person-year observation weights. Subsequent rows report the descriptive statistics for the attributes, using unique union observations.

Table A3: Descriptive Statistics of Union Attributes: Union Actions

	Mean	Std dev	Q1	Q2	Q3
Union premium	0.018	0.069	-0.020	0.016	0.060
Union actions					
Any strike	0.22	0.41	0.00	0.00	0.00
Any CBA	0.93	0.25	1.00	1.00	1.00
Log firm-level CBAs	3.2	2.1	1.4	3.5	4.8
Log sectoral CBAs	2.7	1.5	2.1	2.9	3.6
CBA length	2,768	1,789	1,639	2,583	3,605
CBA clauses	28.2	15.9	17.6	27.5	38.7
Amenity value	0.05	0.07	0.00	0.01	0.07
Clauses: wage adjustments	1.93	1.49	1.00	1.80	2.62
Clauses: wage payment	0.97	0.95	0.24	0.73	1.42
Clauses: other wages	0.10	0.23	0.00	0.00	0.09
Clauses: wage rules	0.84	1.03	0.17	0.56	1.14
Clauses: bonuses	0.53	0.65	0.04	0.33	0.82
Clauses: pays	1.47	1.18	0.55	1.30	2.10
Clauses: assistances	2.43	1.93	0.90	2.19	3.58
Clauses: incentives	0.89	0.88	0.19	0.72	1.31
Clauses: separations	1.29	1.20	0.39	1.00	1.91
Clauses: contract types	0.09	0.21	0.00	0.00	0.07
Clauses: hiring	0.75	0.90	0.07	0.48	1.05
Clauses: contracting	0.66	0.90	0.05	0.36	0.89
Clauses: staffing	0.90	1.21	0.14	0.55	1.16
Clauses: working conditions	0.75	1.04	0.08	0.43	1.00
Clauses: employment protections	1.16	1.15	0.28	0.86	1.71
Clauses: workday	3.67	2.32	2.00	3.50	5.00
Clauses: workplace injuries	0.58	0.60	0.06	0.44	0.91
Clauses: workplace preventions	1.69	1.63	0.53	1.29	2.37
Clauses: vacations	0.45	0.48	0.02	0.32	0.75
Clauses: leaves	0.26	0.47	0.00	0.05	0.33
Clauses: time off	0.30	0.45	0.00	0.12	0.43
Clauses: union-firm relations	1.53	1.44	0.50	1.26	2.18
Clauses: union organization	1.69	1.35	0.62	1.50	2.50
Clauses: bargaining	3.29	3.00	1.70	2.88	4.18

This table displays descriptive statistics for attributes describing actions taken by unions. The sample comprises unions in the UF-AKM connected set, i.e., 4,805 unions (Table 1). The first row reports the descriptive statistics for the union-specific pay premia, using person-year observation weights. Subsequent rows report the descriptive statistics for the attributes, using unique union observations.

Table A4: Descriptive Statistics of Union Attributes: Governance and Leadership

	Mean	Std dev	Q1	Q2	Q3
Union premium	0.018	0.069	-0.020	0.016	0.060
Governance and leadership					
Collective leadership	0.05	0.23	0.00	0.00	0.00
Elected by assembly	0.13	0.34	0.00	0.00	0.00
Election participation	0.64	0.23	0.51	0.67	0.81
Election vote share	0.90	0.16	0.89	0.97	0.99
Female share in board	0.23	0.22	0.05	0.17	0.34
Nonwhite share in board	0.28	0.25	0.07	0.21	0.46
No-college share in board	0.71	0.29	0.57	0.82	0.93
Blue-collar share in board	0.47	0.32	0.14	0.53	0.75
Female parity in board	0.18	0.39	0.00	0.00	0.00
Nonwhite parity in board	0.38	0.48	0.00	0.00	1.00
No-college parity in board	0.41	0.49	0.00	0.00	1.00
Blue-collar parity in board	0.36	0.48	0.00	0.00	1.00
Skill of union board	0.00	1.00	-0.60	-0.06	0.54

This table displays descriptive statistics for attributes describing the governance and leadership structure of unions. The sample comprises unions in the UF-AKM connected set, i.e., 4,805 unions (Table 1). The first row reports the descriptive statistics for the union-specific pay premia, using person-year observation weights. Subsequent rows report the descriptive statistics for the attributes, using unique union observations.

Table A5: UF-AKM Variance Decomposition of Log Earnings by Worker Groups

	High educ	Low educ	Men	Women	White	Nonwhite
	(1)	(2)	(3)	(4)	(5)	(6)
Mean log earnings	3.017	2.078	2.305	2.113	2.338	2.070
Standard deviation of log earnings	0.680	0.496	0.620	0.600	0.652	0.538
Standard deviation (across person-year observations)						
Person effects	0.554	0.369	0.494	0.497	0.531	0.428
Firm effects	0.257	0.187	0.209	0.185	0.212	0.184
Union effects	0.078	0.071	0.070	0.070	0.069	0.073
Percentage of variance in log earnings due to						
Variance in person effects	66.4	55.3	63.6	68.7	66.4	63.3
Variance in firm effects	14.3	14.3	11.3	9.5	10.6	11.7
Variance in union effects	1.3	2.0	1.3	1.4	1.1	1.8
Variance in LLM/industry effects	0.6	1.3	0.8	0.9	0.6	1.2
Covariance in person and firm effects	7.4	8.2	13.1	8.5	13.6	8.8
Covariance in person and union effects	2.0	2.9	2.6	2.5	2.1	2.1
Covariance in person and LLM/industry effects	1.0	3.0	1.8	3.0	2.3	2.1
Adjusted R^2	0.909	0.864	0.901	0.913	0.914	0.881
Number of firms	107,290	928,011	565,688	500,267	690,279	388,708
(firm-only moves)	(593,964)	(5,201,384)	(3,264,804)	(2,090,233)	(3,170,863)	(2,196,260)
Number of unions	3,659	4,497	4,439	3,974	4,560	3,822
(both moves)	(972,388)	(7,801,552)	(5,148,518)	(2,576,116)	(4,757,110)	(3,034,799)

This table summarizes the results from estimating UF-AKM models for log earnings using person-year observations from Brazil (2009-2017) in the largest connected set of six worker groups: (1) workers with more than high school (*high educ*); (2) workers with high school or less (*low educ*); (3) male workers; (4) female workers; (5) white workers; and (6) nonwhite workers. The specification estimated is provided in Equation (2) with the inclusion of local labor market and industry fixed effects.

Table A6: Robustness of Differential Bargaining to Normalization Procedure

		Lowest MSE		Second Lowest MSE		Lowest MSE (granular)	
	Overall wage gap	Differential bargaining	Cross-union sorting	Differential bargaining	Cross-union sorting	Differential bargaining	Cross-union sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. High vs. low education							
Average statistic	0.712	0.013	0.006	0.012	0.006	0.013	0.006
		1.8%	*0.9%*	*1.7%*	*0.9%*	*1.8%*	*0.9%*
Panel B. Male vs. female							
Average statistic	0.192	0.009	0.004	0.009	0.004	0.009	0.004
		4.7%	*2.3%*	*4.8%*	*2.3%*	*4.7%*	*2.3%*
Panel C. White vs. nonwhite							
Average statistic	0.161	0.005	0.002	0.005	0.002	0.005	0.002
		3.3%	*1.2%*	*3.3%*	*1.2%*	*3.3%*	*1.2%*

This table shows the education, gender, and racial wage gaps in Column (1), along with the Oaxaca-Blinder decomposition of the gaps union pay premia (see Section 5.1) using three different normalization procedures. Columns (2)-(3) uses the threshold value that jointly minimizes the mean squared error of Equation (3) on both worker groups using 0.10 intervals of mean log union revenue. Columns (4)-(5) uses the runner-up threshold value in jointly minimizing the mean squared error of Equation (3) on both worker groups using 0.10 intervals of mean log union revenue. Columns (6)-(7) uses the threshold value that jointly minimizes the mean squared error of Equation (3) on both worker groups using 0.01 intervals of mean log union revenue. The differential bargaining effects is the only statistic affected by changing the normalizing constant. The percent of the overall wage gap attributed to Columns (2)-(7) are reported under each statistic. The sample in each panel is restricted to the dual connected set from the UF-AKM connected sets of the two worker groups, i.e., high vs. low education (defined as above high school vs. high school or less) in Panel A, male vs. female in Panel B, and white vs. nonwhite in Panel C. Workers in the marginalized group are reweighted using DFL weights by LLM-by-sector.

Table A7: Worker Support of Unions After the Labor Reform (Level Categories)

		Summary of regression results			
	Prob. of any	Relative to Low		Relative to Balanced	
	contribution	High	Mid	Compress	Exacerbate
	(1)	(2)	(3)	(4)	(5)
Union premia		Among Compress		Among High	
Post×Low	-0.083	0.027	0.016	-0.094	-0.075
	(0.036)	(0.020)	(0.024)	(0.021)	(0.021)
Post×High	0.000				
	(0.031)				
Differential bargaining		Among Balanced		Among Mid	
Post×Compress	-0.104	0.083	0.083	-0.104	-0.057
	(0.032)	(0.031)	(0.036)	(0.032)	(0.030)
Post×Exacerbate	-0.057				
	(0.030)				
Interaction terms		Among Exacerbate		Among Low	
Post×Low×Compress	0.067	0.030	0.049	-0.037	-0.022
	(0.043)	(0.019)	(0.021)	(0.030)	(0.029)
Post×Low×Exacerbate	0.035				
	(0.042)				
Post×High×Compress	0.011				
	(0.038)				
Post×High×Exacerbate	-0.019				
	(0.037)				

This table reports the results of the DiD specification from Table 3 in Column (1). The remaining columns summarize the results. Columns (2)-(3) show the worker support effects for high and mid premium unions (relative to low premium unions), respectively, among each of the three categories of differential bargaining. Columns (4)-(5) show the worker support effects for unions that compress and exacerbate wage gaps by education (relative to unions engaging in balanced bargaining), respectively, among each of the three categories of union premia. In both the union premia (low-mid-high) and differential bargaining (compress-balanced-exacerbate) categorical variables we define their categories according to being below -2 log points, between -2 and 2 log points, and above 2 log points in union premia and differential bargaining by education, respectively.

Table A8: Worker Support of Unions After the Labor Reform (Tercile Categories)

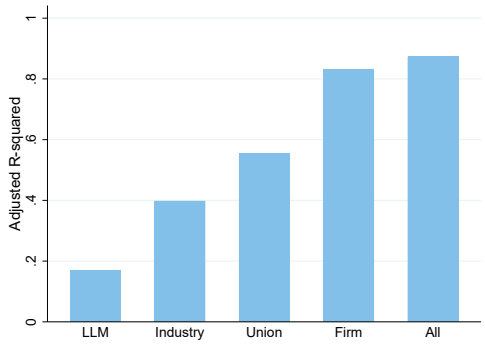
		Summary of regression results			
	Prob. of any	Relative to Low		Relative to Balanced	
	contribution	High	Mid	Compress	Exacerbate
	(1)	(2)	(3)	(4)	(5)
Union premia		Among Compress		Among High	
Post×Low	-0.087	0.017	0.000	-0.083	-0.095
	(0.021)	(0.023)	(0.024)	(0.021)	(0.024)
Post×High	-0.023				
	(0.020)				
Differential bargaining		Among Balanced		Among Mid	
Post×Compress	-0.123	0.064	0.087	-0.123	-0.055
	(0.022)	(0.023)	(0.021)	(0.022)	(0.020)
Post×Exacerbate	-0.055				
	(0.020)				
Interaction terms		Among Exacerbate		Among Low	
Post×Low×Compress	0.087	-0.008	0.055	-0.036	-0.023
	(0.032)	(0.024)	(0.021)	(0.024)	(0.022)
Post×Low×Exacerbate	0.031				
	(0.029)				
Post×High×Compress	0.040				
	(0.030)				
Post×High×Exacerbate	-0.040				
	(0.031)				

This table reports the results the DiD specification from Table 3 in Column (1). The remaining columns summarize the results. Columns (2)-(3) show the worker support effects for high and mid premium unions (relative to low premium unions), respectively, among each of the three categories of differential bargaining. Columns (4)-(5) show the worker support effects for unions that compress and exacerbate wage gaps by education (relative to unions engaging in balanced bargaining), respectively, among each of the three categories of union premia. In both the union premia (low-mid-high) and differential bargaining (compress-balanced-exacerbate) categorical variables, we define their categories according to terciles in the distribution of union premia and differential bargaining by education, respectively.

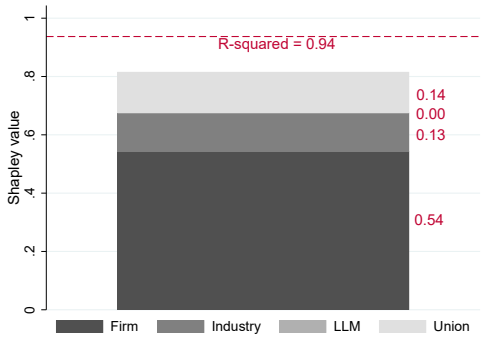
B Appendix Figures

Figure B1: Explained Variation in $\hat{\psi}_j$ by Regressions with Fixed Effects

(a) Adjusted R^2: multi-estab firms



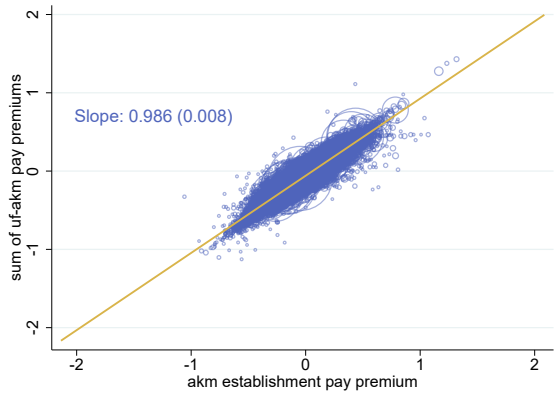
(b) Shapley values: multi-estab firms



These figures show descriptive evidence pertaining to the role of unions in explaining variation in the AKM establishment pay premia. Figure B1a plots the adjusted R-squared from regressing $\hat{\psi}_j$ on separate groups of fixed effects (LLM, Industry, Union, Firm), as well as the entire set of fixed effects (All). To be able to include firm fixed effects, the sample is restricted to establishments in multi-establishment firms, i.e., where there is variation in $\hat{\psi}_j$ within firms. Figure B1b computes the Shapley values for each group of fixed effects when they are all used as regressors for $\hat{\psi}_j$. The horizontal dotted line marks the adjusted R-squared of this regression, and the Shapley values are reported next to each stacked bar. Due to computational limitations when including firm fixed effects, the Shapley value decomposition is geographically restricted to reduce dimensionality. Specifically, Figure B1b pertains to the metropolitan area of Rio de Janeiro, but results are similar across other large metropolitan areas. This explains why the Shapley value of LLM fixed effects is negligible.

Figure B2: Falsification Tests for UF-AKM

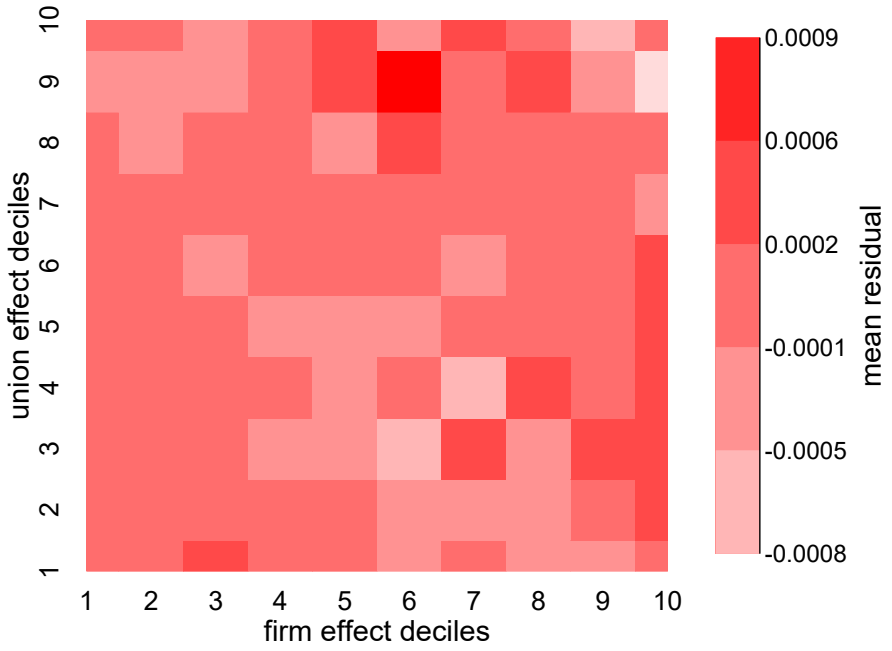
(a) Additivity without establishment effects



(b) Stationarity of union effects

These figures contain falsification tests for additivity ($A1$) and stationarity ($A4$). Figure B2a regresses the sum of the estimates for the firm, union, LLM, and industry fixed effects from the UF-AKM on the corresponding establishment's $\hat{\psi}_j$ from an AKM, instrumenting with a split sample to adjust for attenuation bias due to noise in the estimates. The figure shows the scatterplot of the sum of the UF-AKM components against the IV predicted $\hat{\psi}_j$, along with the fitted line with the slope estimate and its standard error. Figure B2b regresses estimates of $\hat{\eta}_u$ from a late-period UF-AKM (2013-2017) on the $\hat{\eta}_u$ from an early period UF-AKM (2009-2013), instrumenting with the $\hat{\eta}_u$ from a split sample full-period UF-AKM (2009-2017). The figure shows the scatterplot of the late-period $\hat{\eta}_u$ against the IV predicted early-period $\hat{\eta}_u$, along with the fitted line with the slope estimate and its standard error.

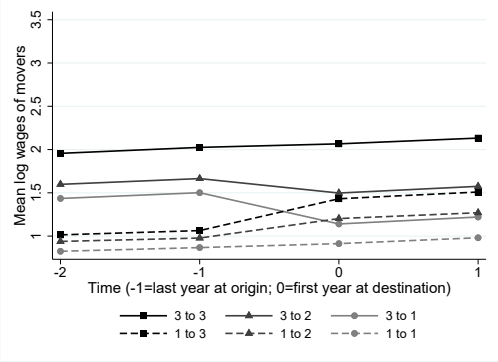
Figure B3: Pattern of Mean Residuals by $\hat{\phi}_{f(j)}$ and $\hat{\eta}_{u(j)}$ Deciles



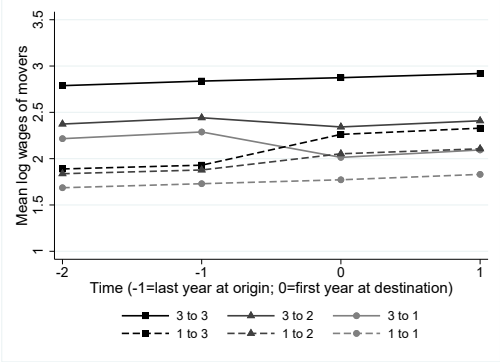
This figure shows a heat map of the mean UF-AKM residuals by cells of $\hat{\phi}_f$ and $\hat{\eta}_u$ deciles. The range of mean residuals is from -0.08 to 0.09 log points and is split into quintiles: darker hues imply higher values (as shown in the legend). Each cell is assigned the hue that corresponds to the quintile containing the value of the mean UF-AKM residual of that cell.

Figure B4: Mover Event Studies by Type of Mover

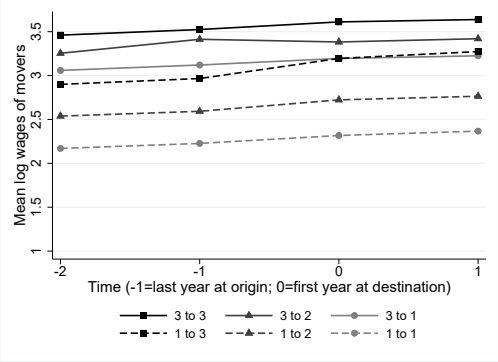
(a) Both movers



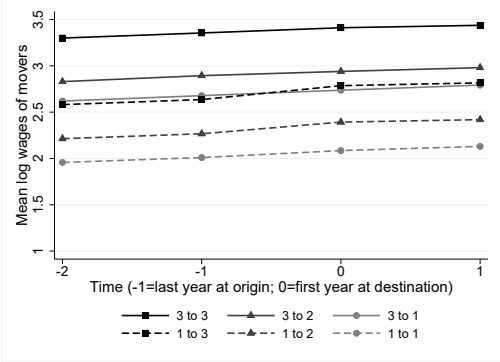
(b) Firm-only movers



(c) Union-only movers



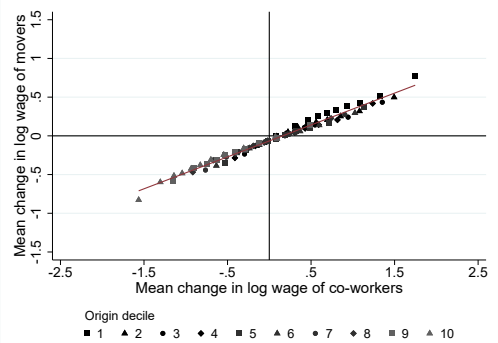
(d) Neither movers



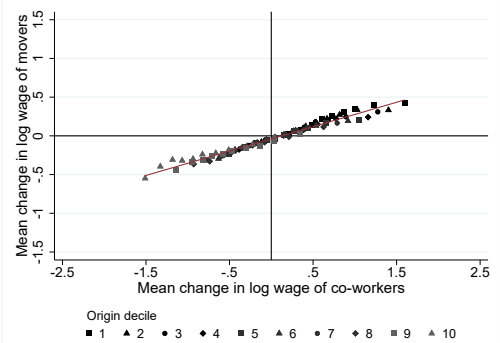
These figures display the evolution of wages for workers who moved from origin establishments in the top and bottom tercile groups to destination establishments in any of the other tercile groups. The movers are defined as workers at establishments who separated from the origin establishment, were reemployed in the destination establishment the next or the following year, and were employed at the origin and destination establishments for 2+ consecutive years. Origin/destination groups are based on terciles of co-worker wages during the calendar year of separation/hiring. Each figure pertains to one of four move types: “both" movers imply a change in firm and union (Figure B4a); “firm-only" movers imply a change in firm but not union (Figure B4b); “union-only" movers imply a change in union but not firm (Figure B4c); “neither" movers imply a change in neither firm nor union (Figure B4d). In the “both” movers figure, wages are net of firm effects using an AKM of “firm-only" movers to reflect our estimation approach (Section 3.2).

Figure B5: Symmetry Test for Moves Up and Down the Job Ladder by Type of Mover

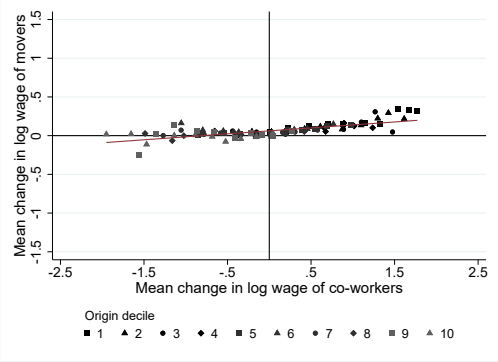
(a) Both movers



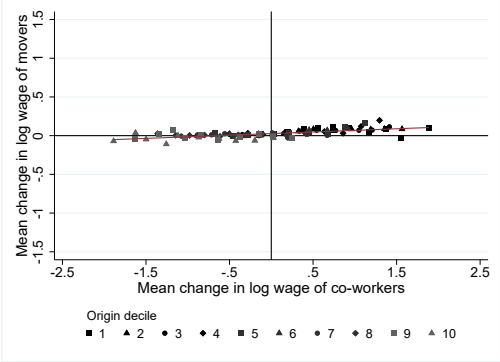
(b) Firm-only movers



(c) Union-only movers



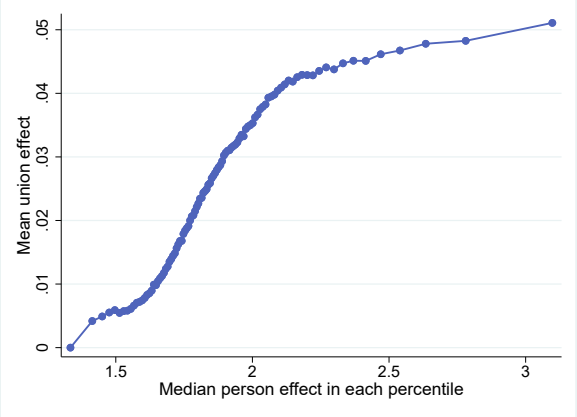
(d) Neither movers



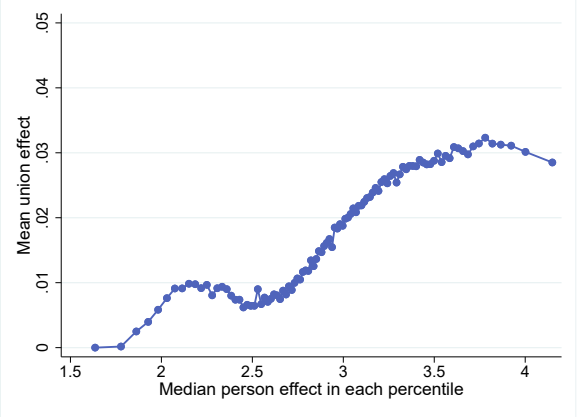
These figures plot the mean change in movers' wages between the years prior to separation and after hiring against the mean change in co-worker wages between origin and destination establishments. Origin and destination establishments are grouped by deciles of co-worker wages; each of the 10x10 dots corresponds to movers from/to given origin/destination deciles. Movers are defined as in Figure B4; and in the "both" movers figure, wages are net of firm effects using an AKM of "firm-only" movers to reflect our estimation approach (see Section 3.2). All wage changes are adjusted for trends based on coefficients from a regression estimated on the sample of stayers, i.e., workers who remain at the same origin establishments over the years around a move. The model includes the education dummies and a quadratic in age fully interacted with these dummies.

Figure B6: Positive Assortative Matching of Workers to Unions

(a) UF-AKM for low skill jobs



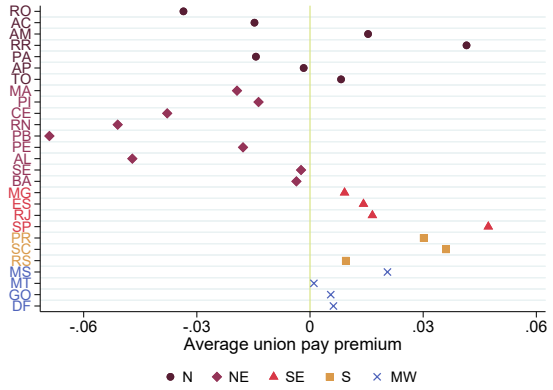
(b) UF-AKM for high skill jobs



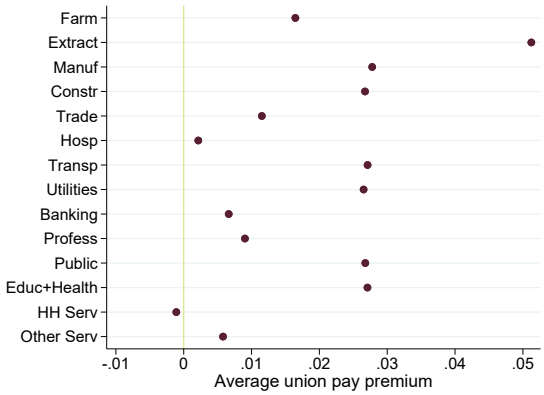
These figures depict positive assortative matching of workers to unions along percentiles of worker skill. Figure B6a uses union effects from a UF-AKM restricted to low skill jobs, i.e., those in occupations whose average AKM person effect is in the bottom tercile. Figure B6b uses union effects from a UF-AKM restricted to high skill jobs, i.e., those in occupations whose average AKM person effect is in the top tercile. To avoid biases in $Cov(\hat{\alpha}_i, \hat{\eta}_{u(i,t)})$ from sampling errors, the person effects (x-axis) are estimated from a union-mover AKM using all jobs. Both figures show the average union effect by percentile bin in the person effect, normalizing the bottom bin to zero.

Figure B7: Differences in Average Union Pay Premia

(a) By state



(b) By sector



These figures show the average union pay premium by state and sector using person-year observations in the baseline UF-AKM sample. The 27 states in Figure B7a are grouped into 5 geographic regions: North (N), Northeast (NE), Southeast (SE), South (S), and Midwest (MW). Out of the 12 sectors in Figure B7b, public administration and household services are dropped from the DiD analysis in Equations (4) due to missing information on union premia in some states.

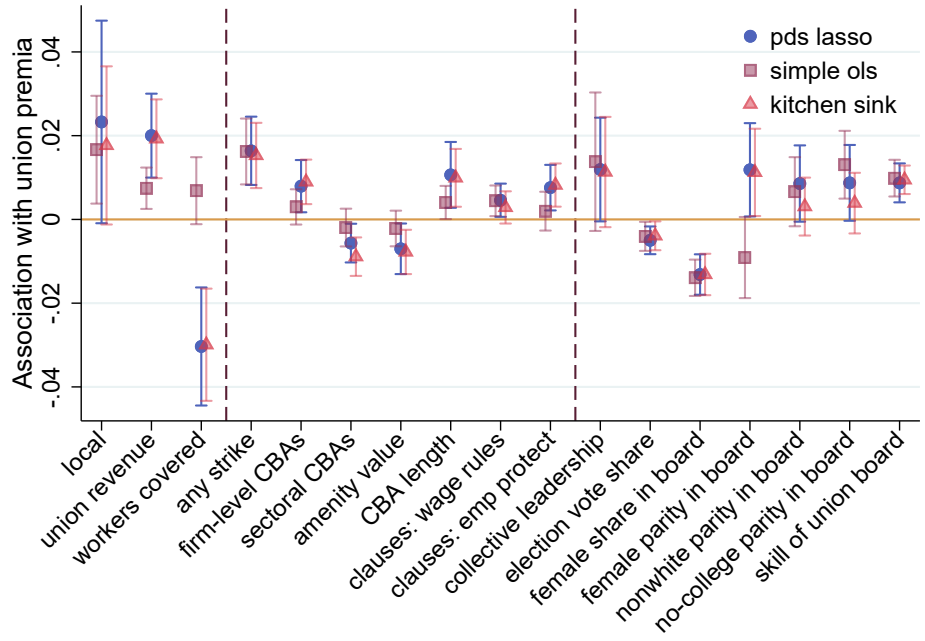
Figure B8: Effect of Reform Across Markets with Varying Union Premia

(a) Impact on employment

(b) Impact on formality

These figures plot the dynamic DiD coefficients from Equation (4) using the PNAD-Contínua data where the treatment dummy $D_{m(i)}$ equals one for labor markets in the top tercile of union pay premia and equals zero for those in the bottom tercile. In Figure B8a, the dependent variable is an indicator for being employed. In Figure B8b, the dependent variable is an indicator for being employed in the formal sector. Pooled DiD estimates with standard errors for the period after the reform and the expiration of extant CBAs (denoted by the vertical dotted lines) are reported in the figure. Regressions use the standard weights provided by PNAD-Contínua, and we also inversely weight the regression by the standard deviation in union pay premia within a labor market. Standard errors are clustered at the state level.

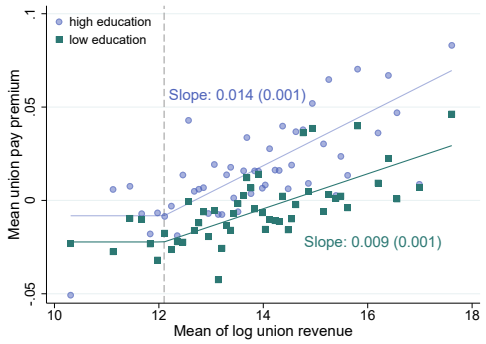
Figure B9: Correlates of Union Pay Premia: Including State and Sector Fixed Effects



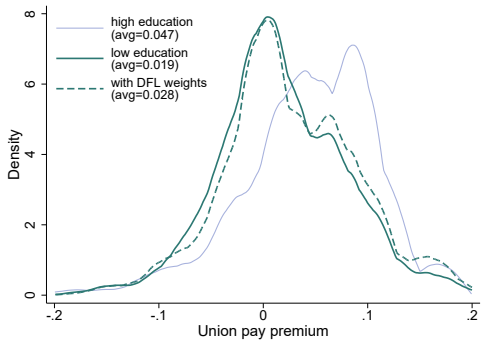
This figure shows the estimates of coefficients correlating union-specific pay premia with union attributes (denoted in the x-axis), accounting for state and sector fixed effects. For each attribute, estimates for three separate analyses are reported: 1) *kitchen sink:* a kitchen-sink regression of the premia on the attribute of interest, adding all other selected attributes as controls; 2) *pds lasso:* a post-double selection LASSO to partially address omitted variables bias with observed attributes; and 3) *simple ols:* a simple bivariate OLS regression of the premia on the attribute of interest. Figure B9 only reports the attributes that have a statistically significant estimate. The dotted vertical lines partition the attribute space into three categories: a) union type; b) union actions; and c) governance and leadership. All regressions are run with unique union observations, weighting by person-year observations.

Figure B10: Union Pay Premia Normalizations and DFL Adjustments by Worker Groups

(a) Normalization by education

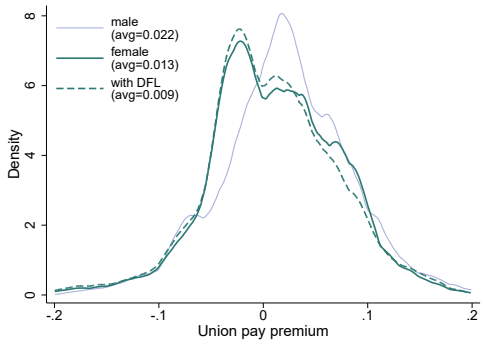


(b) Distributions by education

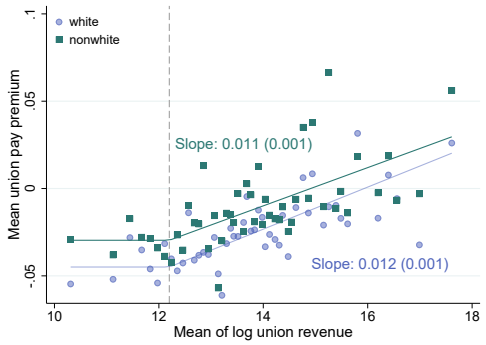


(c) Normalization by gender

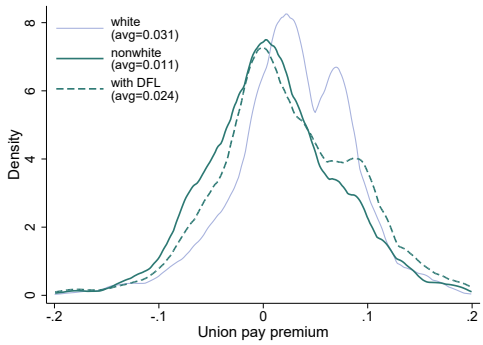
(d) Distributions by gender



(e) Normalization by race

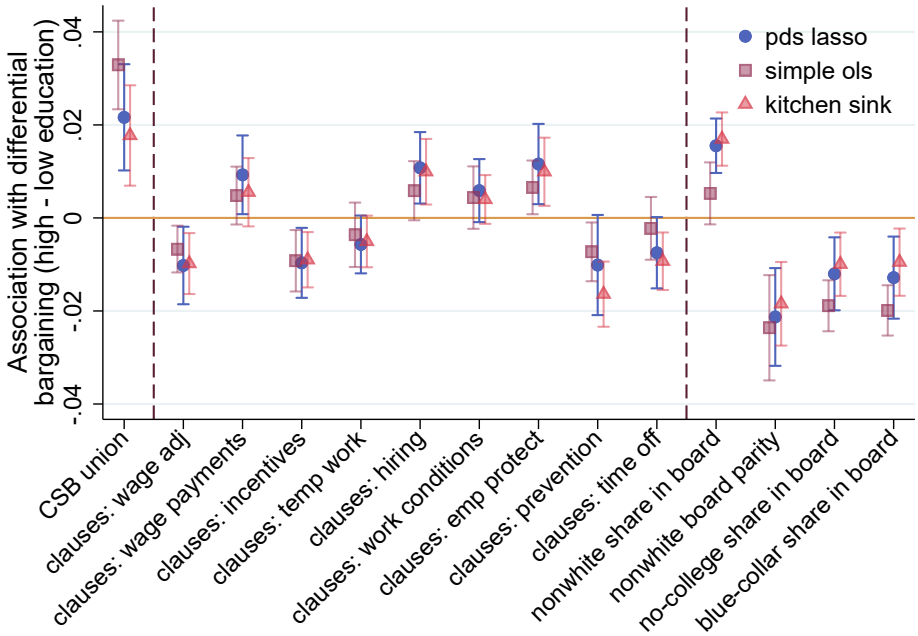


(f) Distributions by race



These figures show the normalization procedure of the union effects, as well as how the distribution of the union pay premiums are affected by the DFL weights for the three pairs of worker groups: high vs. low education (top panel), male vs. female (middle panel), and white vs. nonwhite (bottom panel). The normalization procedures in Figures B10a, B10c, and B10e use the threshold value that jointly minimizes the mean squared error of Equation (3) on both worker groups using 0.10 intervals of mean log union revenue. These figures overlay binned scatter plots of the mean group-specific union effects (estimated from group-specific UF-AKMs) across bins of log union revenue (averaged across 2009-2017 GRCSU files). The fitted lines corresponds to the OLS estimates from Equation (3) for the selected threshold τ (dotted vertical line). The horizontal lines represents the normalizing constants $\hat{\beta}_0$; the slopes conditional on revenues above the threshold $\hat{\beta}_1$ are reported next to the scatterplot. Figures B10b, B10d, and B10f show the kernel densities of the group-specific union pay premia, showing both the unweighted and DFL weighted versions of the marginalized group. The corresponding average union pay premia are reported in the legend.

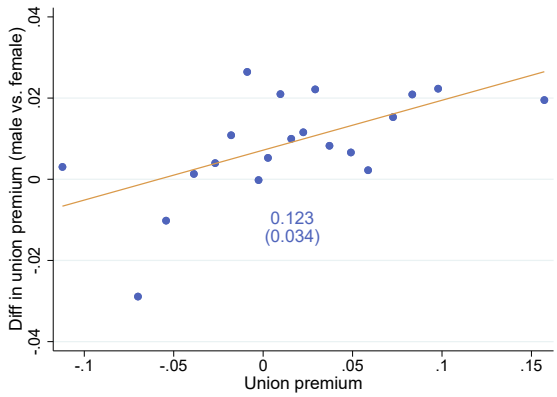
Figure B11: Correlates of Differential Bargaining by High vs. Low Educated Workers



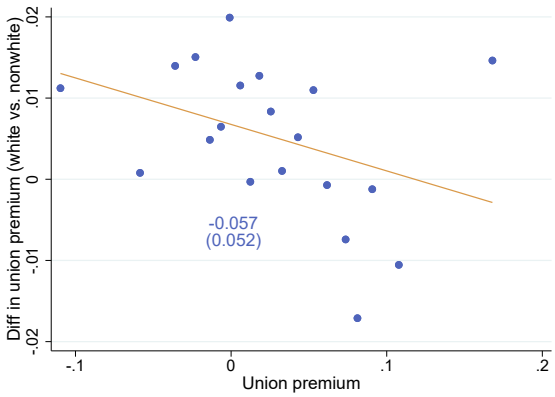
This figure shows the estimates of coefficients correlating union-specific bargaining differentials for high- versus low-educated workers with union attributes (denoted in the x-axis). For each attribute, estimates for three separate analyses are reported: 1) *simple ols:* a simple bivariate OLS regression of the premia on the attribute of interest; 2) *kitchen sink:* a kitchen-sink regression of the premia on the attribute of interest, adding all other selected attributes as controls; and 3) *pds lasso:* a post-double selection LASSO to partially address omitted variables bias with observed attributes. Figure B11 only reports the attributes that have a statistically significant estimate. The dotted vertical lines partition the attribute space into three categories: 1) union type; 2) union actions; and 3) governance and leadership. All regressions are run with unique union observations, weighting by DFL-reweighted low education person-year observations.

Figure B12: Potential Tradeoffs Between Union-Driven Wage Gains and Compression

(a) Gender gaps



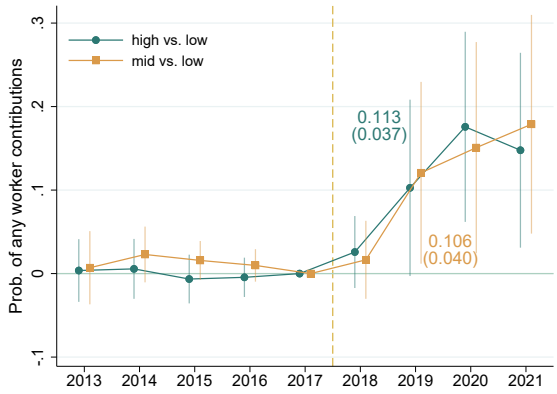
(b) Racial gaps



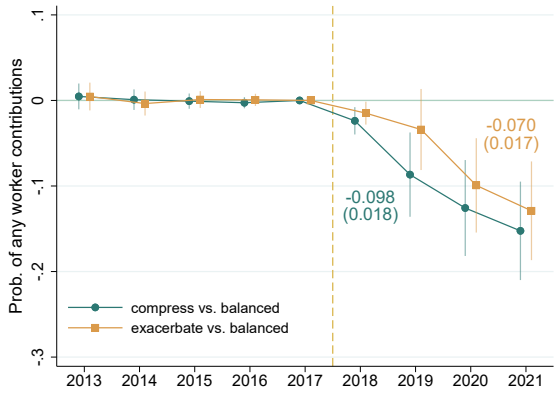
These figures depict binned scatterplots of the differential bargaining effects by gender (Figure B12a) and race (Figure B12b) on the union pay premium, including the linear prediction from an OLS. The estimates for the slope coefficients and their standard errors are reported next to each scatterplot. Samples are restricted to the dual connected set from the UF-AKMs of by gender and race, respectively, applying DFL weights by LLM-by-sector to the marginalized group.

Figure B13: Differentials in Worker Support After the Labor Reform

(a) By union premia



(b) By differential bargaining



These figures plot the dynamic DiD coefficients from $y_{ut} = \sum_{j \neq 0} \beta_j (D_u \times \delta_{t=j}) + \mu_u + \delta_t + \gamma X_{ut} + \varepsilon_{ut}$ using an indicator for receiving any unionization contribution as a dependent variable. In Figure B13a, we condition on unions that engage in balanced bargaining and plot the coefficients for two regressions: one where the treatment dummy D_u captures high versus low premium unions and the other where D_u captures mid versus low premium unions. In Figure B13b, we condition on unions with high-to-mid pay premia and plot the coefficients for two regressions: one where the treatment dummy D_u captures compressing versus balanced bargaining and the other where D_u captures exacerbating versus balanced bargaining. The pooled DiD estimates and their standard errors are reported in each figure. Because we condition the samples based on the results in Table 3, these pooled estimates are different from those reported there. Time-varying controls include state-by-year and sector-by-year fixed effects. Vertical dotted lines denote the implementation of the labor reform. Standard errors are clustered at the union level.

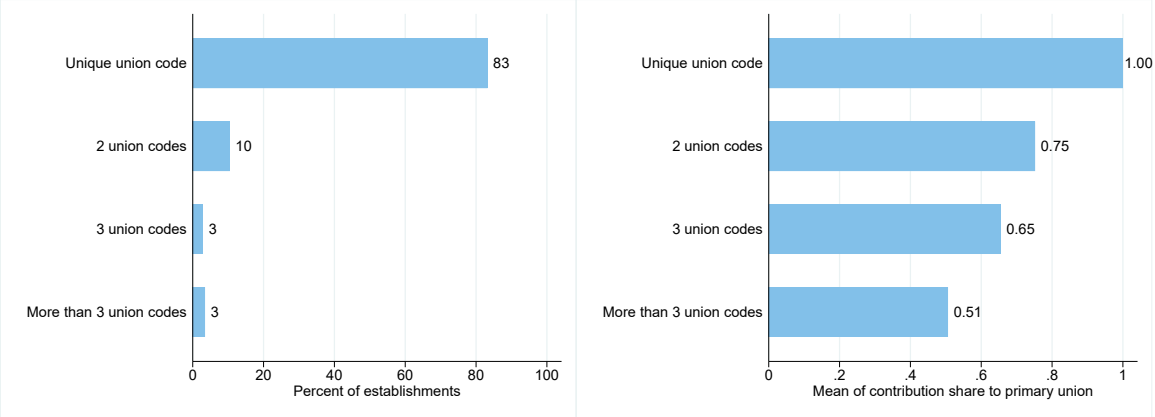
C Data Appendix

C.1 Guia de Recolhimento de Contribuição Sindical Urbana (GRCSU)

Our main task is to assign a primary union to establishments in RAIS. The GRCSU data include the amounts paid by each establishment to the unions representing their workers—excluding the agricultural sector. We import the annual GRCSU text files and remove payments to employer associations and the government. We then collapse the sum of the payments made to the establishment-union-year level (2006-2021).

Figure C1: Establishments Linked to Unions in GRCSU

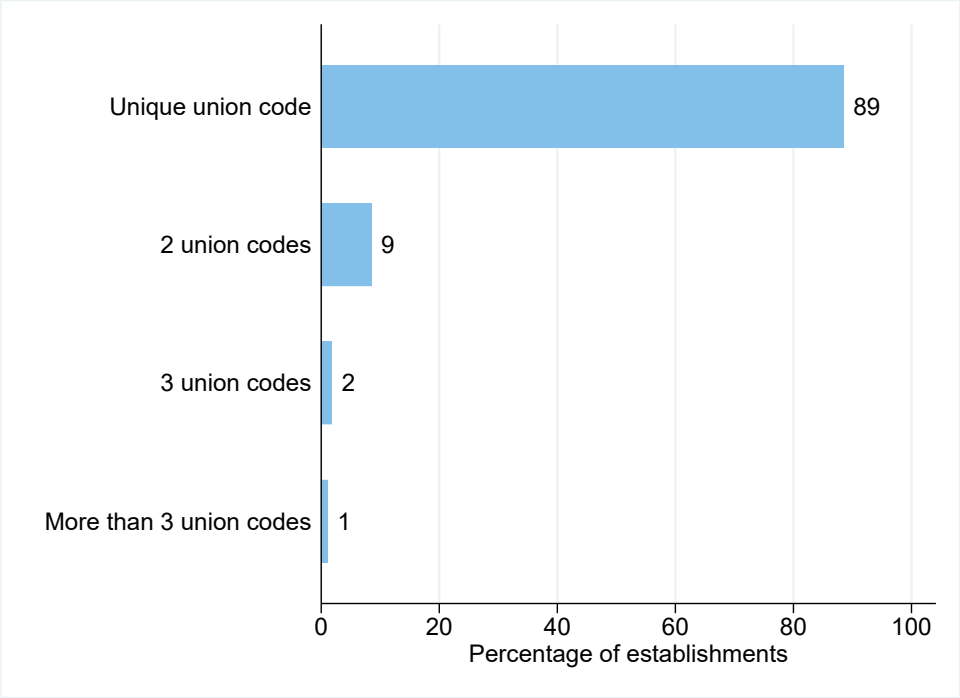
(a) Establishment-to-union mapping (b) Contribution share to primary union



These figures explore how reasonable assigning a primary union to each establishment is. Figure C1a shows the percent of establishments in GRCSU by the number of unions they make contributions to over the entire sample period. Figure C1b depicts the contribution share that the primary union receives over the entire sample period (averaged across establishments) by the number of unions associated with the establishment.

Depicted in Figure C1a is the percent of establishments in GRCSU by the number of unions they make contributions to over the entire sample period. The figure shows that 83% of establishments only make contributions to a single union, while 10% make contributions to only 2 unions. We then define the primary union to be the one receiving the largest share of the contributions an establishment makes over the sample period. Figure C1b depicts the contribution share that the primary union receives over the entire sample period (averaged across establishments) by the number of unions associated with the establishment. The figure shows that the primary union, on average, receives the vast majority of contributions even when the establishment-to-union mapping is not unique.

Figure C2: Establishments Linked to Year-Specific Primary Unions



This figure explores how reasonable it is to assume that the primary union of each establishment is time invariant. Figure C2 shows the percent of establishments in the GRCSU data by number of primary unions across years when the assignment is determined within each year of the sample period.

We then check whether it is reasonable to assume that the primary union of each establishment is time invariant. Hence, we define the year-specific primary union to be the one receiving the largest share of the contributions an establishment makes in a given year. Depicted in Figure C2 is the percent of establishments in the GRCSU data by number of primary unions across years. The figure shows that 89% of establishments never switch primary union, while 9% experience only one change in primary union over the sample period.

In order to increase the amount of establishments in RAIS with an assigned primary union, we implement the following algorithm. (1) For each establishment, assign it the union to which it makes the majority of contributions over the sample period. (2) If no assignment is made, for each remaining establishment, assign it the union to which its parent firm makes the majority of contributions over the sample period. (3) If no assignment is made, for each remaining establishment, assign it the union with which it has the longest relationship in terms of negotiated firm-level CBAs in *Sistema Mediador*. Among

the establishments with a primary union, 76.1% are assigned according to (1), 23.7% are assigned according to (2), and 0.2% are assigned according to (3).

C.2 Relação Anual de Informações Sociais (RAIS)

Linked employer-employee data from RAIS allows us to track workers and their hourly earnings across jobs. We follow Gerard et al. (2021) to construct a worker-year panel with the 2009-2017 RAIS data. Specifically, for each state-year file, we keep workers employed on December 31st whose tenure is greater than one month. Log hourly wages are constructed by taking the logarithm of the real value of December earnings (using the December 2015 CPI) divided by the monthly contracted hours. We create unique person-year observations by choosing the job with the highest contracted hours, breaking ties by selecting the job with the highest log hourly wage.

The selected unique person-year observations for each state are then stacked across 2009-2017 into a single state file. Each establishment is assigned its modal legal classification, municipality, and industry code. Each worker is assigned its modal gender, race, date of birth, and education. We then keep observations belonging to the private sector based on the legal classification of each establishment, and workers who are hired on open-ended nonfarm contracts that are paid on a monthly basis.

The remaining observations in each state file are then stacked across states into a single master file. The entire employment history of an individual is removed when one of the following four conditions is satisfied. First, the worker has a repeated person-year observation across states. Second, the nominal value of the reported December wage is below the federal minimum wage for that month. Third, the log hourly wage is above the 99th percentile of the state- and year-specific wage distribution. Fourth, the log hourly wage changes by more than 100 log points between consecutive observations. Based on these person-year observations, the modal assignments from the previous paragraph are applied again, before restricting to workers age 25 to 54.

Table C1: Unionization by Industry × Local Labor Market Cells

Industry	Local labor market	% of union's coverage	% of cell covered by union
Union of Workers in Pharmacy, Cosmetics, Dentistry, and Health Products			
Retail of pharmaceuticals, cosmetics, and medical/optical items	Manaus	68	58
Wholesale trade of non-food consumer products	Manaus	27	32
Union of Employees in Commerce of Manaus			
Non-specialized retail trade	Manaus	19	91
Retail of pharmaceuticals, cosmetics, and medical/optical items	Manaus	4	36
Union of Professional Journalists of Mato Grosso do Sul			
Editing for books, newspapers, magazines, and other publications	Campo Grande	73	57
Printing activities	Campo Grande	4	4
Union of Workers in the Graphic Industries of Mato Grosso do Sul			
Printing activities	Campo Grande	24	89
Editing for books, newspapers, magazines, and other publications	Campo Grande	11	35
Union of Employees in Security and Surveillance Companies in Cuiabá			
Surveillance, private security, and cash transport activities	Cuiabá	82	94
Retail trade of IT and communication equipment; household items	Cuiabá	7	11
Union of Workers Employed in Buying, Selling, Leasing, and Managing Real Estate and Brokerage			
Combined services for building support	Blumenau	45	63
Real estate activities under contract or commission	Blumenau	24	72

This table includes a list of unions, along with a list of two industry × local labor market cells where they represent workers. For each cell, Table C1 reports the percentage of the union's constituent workers in that cell, as well as the percentage of the workforce in that cell that is represented by the union. Highlighted in gray are cells that are represented by more than one of the unions shown in the table. Industry refers to 3-digit industry codes (CNAE) and local labor markets refer to micro-regions.

Finally, the above sample is restricted to establishments that have an assigned primary union. Table C1 includes a list of unions, along with a list of two industry × local labor market cells where they represent workers. The concordance between the union names and the reported industry-LLM cells provide a sense check for the primary union assignment algorithm. Importantly, the table illustrates that the category-geography cells where unions enjoy a monopoly do not map to industry-LLM cells. For example, no single union represents all unionized workers in an industry-LLM cell. In fact, the median share of workers covered by a given union in an industry-LLM cell is 57%, and the average union represents workers in establishments across 13 industries and 7 LLMs.

C.3 Key Variables

Units of observation

- **Worker.** An employed individual. Only identified in RAIS with a unique PIS code.

- **Establishment.** A single physical location where economic activity (like the production of goods or services) takes place. Identified by the 14-digit CNPJ code.
- **Firm.** An economic organization that produces goods or services for sale. It can consist of one or more establishments that operate under unified management. Identified by the first 8 digits of each establishment's 14-digit CNPJ code.
- **Union.** An organization formed to represent workers' collective interests in negotiations with employers. Identified by a unique union code (or *codigo sindical*) that has a corresponding CNPJ code.

Worker/job characteristics

- **Log earnings (RAIS).** Logarithm of the real value of December earnings divided by the monthly contracted hours, using workers in RAIS who were employed throughout the month of December (see details in previous section).
- **Log earnings (PNAD).** Logarithm of the real value of monthly earnings (winsorized at 1 and 99%) using workers in PNAD with age 25-54 who report being employed in the private sector.
- **Education.** A high- vs. low-education dichotomy based on the worker's modal education in RAIS. High education is defined as above high school, while low education is defined as high school or less.
- **Gender.** A male vs. female dichotomy based on the worker's modal gender in RAIS.
- **Race.** A white vs. nonwhite dichotomy based on the worker's modal race in RAIS. White workers are those classified as white (*branco*), while nonwhite workers are those classified as mixed (*pardo*) or black (*preto*).
- **Collar.** A white- vs. blue-collar dichotomy based on the modal CBO code of a worker's job, i.e., IBGE's classification code for occupations. White-collar (blue-collar) workers have modal a CBO starting with 0-4 (5-9).
- **Job skill.** A high- vs. low-skill dichotomy based on whether the average AKM person effect (at the three-digit CBO occupational code level) is in the top vs. bottom tercile of this distribution, respectively.

Establishment/firm characteristics

- **Industry.** A group of establishments that produce similar or closely related goods or services. Identified by three-digit CNAE codes, i.e., 285 detailed groupings based on IBGE's classification code for economic activities.
- **Local labor market (LLM).** A specific geographic area where workers and employers are constrained by similar economic conditions. Identified by microregions, i.e., neighboring municipalities grouped into 543 units by IBGE.
- **Sector.** A broader classification that groups multiple industries based on a shared area of the economy. Identified by CNAE codes aggregated to the *seção* level, i.e., 15 broad groupings based on IBGE's classification code for economic activities.

Union characteristics

- **Union revenue.** The total real value of unionization contributions that a union receives in a given year, based on information in GRCSU. When used for the normalization of union premia and as an attribute associated with the union pay premia, we take the average of the log union revenue across years. When used as an outcome to proxy for worker support, we use the inverse hyperbolic sine of union revenue in each year, as well as an indicator for whether any contribution was made.
- **Local union.** A union representing workers of a given category in a set of municipalities. This is in contrast to federations (confederations), who represent workers at the state (national) level. This information is provided in CNES.
- **Industrial union.** A union whose category reflects the establishment's main economic activity. This is in contrast to occupation-based (or craft) unions, where the category is determined by the type of job performed (i.e., *categoria diferenciada* or *profissionais liberais*). This information is provided in CNES.
- **Municipal coverage.** A union whose geographic coverage is restricted to a single municipality or a set of municipalities. This information is provided in CNES.
- **Union central.** National level, umbrella organizations that help coordinate the activities of unions and lobby for political favor. While union centrals do not directly participate in collective bargaining, they are indirectly involved by coordinating

union priorities *across* worker categories. There are 9 union centrals in Brazil. Information on an individual union's affiliation with a specific union centrals is provided in CNES.

- **Workers covered.** The logarithm of the total person-year observations in RAIS assigned to a union divided by the years observed in the data.
- **Union density.** The mean number of members reported in union elections (provided in CNES) divided by the workers in RAIS assigned to a union.
- **Strike.** An indicator for whether a union engaged in a strike. Information is taken from SAG, which contains establishments that experienced a strike. We merge these establishments to RAIS using a fuzzy merge based on the name, industry, and location of the establishment, and assign the strike to the establishment's primary union.
- **CBAs: count, length, and clauses.** The firm-level and sectoral CBAs in SM are assigned to establishments following Corradini et al. (2025). We record the existence of any CBA, as well as the total count of CBAs by type. We measure CBA length as cross-year average of tokens (words) in the CBAs that cover workers at an establishment. We measure CBA clauses as the cross-year average of clauses (overall or by subgroup) in the CBAs that cover workers at an establishment.
- **Amenity value of CBA.** Borrowing the wage-equivalent value of clause subgroups for firm-level CBAs provided in Lagos (2024), we take the cross-year average of the amenity value of the CBAs that cover workers at an establishment.
- **Collective leadership.** A union whose union leadership is formed by a collective that shares power (*colegiada*). This is in contrast to a presidential leadership, where a single individual is the head of the union. This information is provided in CNES.
- **Elected by assembly.** A union that requires a general assembly of its constituents to take place prior to an election. This information is provided in CNES.
- **Election participation.** The average participation rate across a union's elections, calculated by taking the number of votes cast and dividing by the number of workers affiliated with the union. This information is provided in CNES.

- **Election vote share.** The average share of votes of the winning slate across a union's elections, calculated by taking the number of votes cast for the winning slate and dividing by the total number of votes cast. This information is provided in CNES.
- **Share in board.** The average share of some characteristic among board members across a union's elections, where information on characteristics (gender, race, education, occupation) are obtained by linking the union leaders to RAIS.
- **Parity in board.** An indicator for whether the share of some characteristic among board members is at least as large as than among the workers covered.
- **Skill of union board.** The average of the mean AKM person effect among union board members across a union's elections.

D UF-AKM Appendix

D.1 Economic model

The baseline UF-AKM specification in Equation (2) results from an economic model combining 1) a posting model motivated by workers' heterogeneous preferences over firms (Card et al., 2018); and 2) a right-to-manage model of union bargaining over posted wages (Nickell and Andrews, 1983). This framework deviates from that in Lagos (2024) by distinguishing between firms and establishments, focusing solely on wages rather than amenity-wage tradeoffs, and imposing a proportionality assumption for bargaining power.

Posting model Suppose there are $\mathcal{J}$ establishments in a market, where each establishment $j \in \{1, ..., \mathcal{J}\}$ belongs to some firm $f(j)$ that posts wages $w_{f(j)}$ that workers observe at no cost. Workers in this market have heterogeneous preferences over firms. Specifically, for each worker i, the indirect utility of working at firm $f(j)$ is

$$u_{i,f(j)} = \beta \log(w_{f(j)} - b) + a_{f(j)} + \varepsilon_{i,f(j)}, \tag{D1}$$

where b is the workers' reference wage from their outside option, $a_{f(j)}$ are non-wage amenities that generate vertical differentiation across firms, and $\varepsilon_{i,f(j)}$ refers to the idiosyncratic preferences from working at firm $f(j)$.

Assuming that the $\{\varepsilon_{i,f(j)}\}$ are independent draws from a Type I Extreme Value distribution and the number of firms is very large (McFadden, 1973), workers' choice probabilities are closely approximated by exponential probabilities.[1] Hence, the firm-specific labor supply functions are approximated by

$$\log(L_{f(j)}) = \log(\lambda) + \beta \log(w_{f(j)} - b) + a_{f(j)}. \tag{D2}$$

Firms generate revenue in a perfectly competitive product market (whose price is used as the numeraire) by producing output $Y_{f(j)} = T_{f(j)} f(L_{f(j)})$, where $T_{f(j)}$ represents a firm-specific productivity shifter and $f(\cdot)$ is the production function—assumed to be a twice continuously differentiable and homogeneous of degree $\alpha \in (0, 1)$. The market is further simplified by ignoring capital and intermediate inputs.

[1] The exponential probabilities are $p_{f(j)} \approx \lambda \exp(\beta \log(w_{f(j)} - b) + a_{f(j)})$, where λ is a constant common across all firms in the market.

The firm's problem is to post the wage that minimizes production costs given labor supply in Equation (D2). The posted wage is common to all workers since firms cannot discriminate on the basis of idiosyncratic preferences $\{\varepsilon_{i,f(j)}\}$. The optimal choice is given by the following cost-minimization problem

$$\min_{w} w_{f(j)} L(w_{f(j)}) \quad \text{s.t. } T_{f(j)} f(L(w_{f(j)})) \geq \bar{Y}. \tag{D3}$$

The first order conditions imply that the wage set by each "monopsonistic" firm is

$$w^M_{f(j)} = T_{f(j)} f'(L_{f(j)}) \mu_{f(j)} \left(\frac{e^L_{wf(j)}}{1 + e^L_{wf(j)}} \right) \tag{D4}$$

where $e^L_{wf(j)} = \frac{\beta w_{f(j)}}{w_{f(j)} - b}$ represents the firm-specific labor supply elasticity with respect to wages. The $\mu_{f(j)}$ term captures the marginal cost of production, which the firm will equate to marginal revenue at its optimal choice for $Y_{f(j)}$. Hence, wages are marked down from the marginal revenue product of labor by a Lerner markdown of $e^L_{wf(j)}/(1 + e^L_{wf(j)})$.

As in Card et al. (2018), assume a linear technology $f(L_{f(j)}) = \theta L_{f(j)}$ and price-taking employers in the output market to specify the marginal revenue product of labor: $T_{f(j)} f'(L_{f(j)}) \mu_{f(j)} = T_{f(j)} P_{f(j)} \theta$. To simplify further, assume that reference wage is proportional to productivity ($b = \bar{b}\theta$). Rearranging terms results in

$$w^M_{f(j)} = \left(\frac{\theta \bar{b}}{1 + \beta} \right) \left(1 + \beta R_{f(j)} \right), \tag{D5}$$

where $R_{f(j)} = T_{f(j)} P_{f(j)} / \bar{b}$ captures the proportional gap in marginal labor productivity at firm $f(j)$ relative to the outside option. From Equations (D5) it's clear to see that taking logs would give a linearly additive expression for log wages that involves a fixed worker component and a fixed firm component as in Abowd et al. (1999).

Right-to-manage model Suppose that workers at each establishment j are represented by some union $u(j)$. Following the setting in Brazil, unions can represent workers at establishments in multiple firms and multi-establishment firms need not have the same union at each of their establishments. The role of these unions is to constrain employers' wage setting power by negotiating over wages at the establishment level.

Wage negotiations occur as in a right-to-manage model (Nickell and Andrews, 1983). That is, unions and employers engage in Nash bargaining over wages, knowing that em-

ployment is determined by either the labor supply or labor demand curve—whichever is binding at the negotiated wage. The solution to the Nash bargaining problem is

$$w_j^* = \text{argmax}[\pi(w_j) - \pi(g)]^{1-\gamma_j}[(v(w_j) - v(g))L(w_j)]^{\gamma_j}, \tag{D6}$$

where $\gamma_j \in [0, 1]$ is the bargaining power of the union relative to j; w_j is the negotiated wage; g is the wage under an impasse; $\pi(\cdot)$ is the employer's profit function; $v(\cdot)$ is the union's indirect utility per worker; and $L(\cdot)$ pins down employment.

The solution produces a contract curve for wages starting at the impasse point g which increases with the union's relative bargaining power γ_j. At $\gamma_j = 0$ the union does not constrain wage setting and we get the monopsony wage, i.e., $w_j^* = g = w_{f(j)}^M$. Denote $G_{u(j)}(\gamma_j)$ as the proportional change in the monopsony wage produced by the Nash bargaining solution at some γ_j. In particular, this means that $G_{u(j)}(0) = 1$.

We assume that the bargaining power of any given union relative to each of its corresponding establishments is such that it produces the same proportional change to the monopsony wage. In other words, $G_{u(j)}(\gamma_j) = G_{u(j)} \geq 1$ for all establishments j whose workers are unionized by $u(j)$. Under this proportionality assumption for bargaining power, the wage set at the establishment level is given by

$$w_j^* = w_{f(j)}^M G_{u(j)} = \left(\frac{\theta \bar{b}}{1+\beta}\right)\left(1 + \beta R_{f(j)}\right) G_{u(j)}. \tag{D7}$$

Taking logs of Equation (D7) results in an expression for wages that maps to the UF-AKM in Equation (2), i.e., log wages are additive in a fixed worker component, a fixed firm component, and a fixed union component. Hence, in light of this economic model, the union premium captures $\log(G_{u(j)})$ which is increasing in the union's bargaining power. To the extent that the proportionality assumption does not hold in reality, the union premium captures some average wage effect that is increasing in each γ_j of the union.

Although this model is useful to understand the economic object that the union premia in the UF-AKM capture, it does not provide a micro-foundation for the heterogeneity in these premia. In order to do so, one would have to specify the source of γ_j, which could stem from local economic conditions, the financial stability of the union, and worker support, among others. Moreover, the model abstracts from a union's agency in leveraging its power by assuming a simple objective function. For example, to incorporate negative union premia into the model, one could consider unions that act as if $\gamma_j = 0$ and

instead use their leverage to hinder productivity at the establishment, i.e., $T_j = T_{f(j)}G_{u(j)}$ for $G_{u(j)} < 1$. The empirical evidence we provide in Sections 4 and 5 are useful building blocks for formalizing a model of heterogeneous union behavior.

D.2 Identification of ATEs

Consider wage changes among movers from establishment from ℓ to k

$$\mathbb{E}[y_{i2} - y_{i1}|j(i,1) = \ell, j(i,2) = k].$$

Using our potential outcomes notation, this is equivalent to

$$\mathbb{E}[y_{i2}(d_\ell, d_k) - y_{i1}(d_\ell, d_k)|D_{i,1} = d_\ell, D_{i,2} = d_k].$$

Using assumption $A1$, we rewrite this as

$$\begin{aligned}&\mathbb{E}[y^f_{i2}(f(d_\ell), f(d_k)) - y^f_{i1}(f(d_\ell), f(d_k))|D_{i,1} = d_\ell, D_{i,2} = d_k] \\ &+ \mathbb{E}[y^u_{i2}(u(d_\ell), u(d_k)) - y^u_{i1}(u(d_\ell), u(d_k))|D_{i,1} = d_\ell, D_{i,2} = d_k].\end{aligned}$$

Simplifying with assumption $A2$ gives

$$\begin{aligned}&\mathbb{E}[y^f_{i2}(f(d_k)) - y^f_{i1}(f(d_\ell))|D_{i,1} = d_\ell, D_{i,2} = d_k] \\ &+ \mathbb{E}[y^u_{i2}(u(d_k)) - y^u_{i1}(u(d_\ell))|D_{i,1} = d_\ell, D_{i,2} = d_k].\end{aligned}$$

We then add and subtract potential outcomes terms at $t = 2$ for establishment ℓ

$$\begin{aligned}&\mathbb{E}[y^f_{i2}(f(d_k)) - y^f_{i2}(f(d_\ell)) + y^f_{i2}(f(d_\ell)) - y^f_{i1}(f(d_k))|D_{i,1} = d_\ell, D_{i,2} = d_k] \\ &+ \mathbb{E}[y^u_{i2}(u(d_k)) - y^u_{i2}(u(d_\ell)) + y^u_{i2}(u(d_\ell)) - y^u_{i1}(u(d_k))|D_{i,1} = d_\ell, D_{i,2} = d_k],\end{aligned}$$

in order to implement $A3$ and remove potential outcomes terms at $t = 1$

$$\begin{aligned}&\mathbb{E}[y^f_{i2}(f(d_k)) - y^f_{i2}(f(d_\ell))|D_{i,1} = d_\ell, D_{i,2} = d_k] \\ &+ \mathbb{E}[y^u_{i2}(u(d_k)) - y^u_{i2}(u(d_\ell))|D_{i,1} = d_\ell, D_{i,2} = d_k].\end{aligned}$$

Given $A5$, we can drop the conditional expectations to get

$$\mathbb{E}[y^f_{i2}(f(d_k)) - y^f_{i2}(f(d_\ell))] + \mathbb{E}[y^u_{i2}(u(d_k)) - y^u_{i2}(u(d_\ell))].$$

Lastly, $A4$ allows us to generalize the potential outcome functions to any t and $A2$ allows us to rewrite this as our ATEs of interest

$$\underbrace{\mathbb{E}[y^f_{it}(f(d_k), f(d_k)) - y^f_{it}(f(d_\ell), f(d_\ell))]}_{\phi_{f(k)} - \phi_{f(\ell)}} + \underbrace{\mathbb{E}[y^u_{it}(u(d_k), u(d_k)) - y^u_{it}(u(d_\ell), u(d_\ell))]}_{\eta_{u(k)} - \eta_{u(\ell)}}.$$

This implies that wage changes among job switchers identify the ATE of working for firm $f(k)$ relative to firm $f(\ell)$ and being represented by union $u(k)$ relative to union $u(\ell)$. More generally, under $A1$-$A5$, realized wage changes can be written as

$$\Delta y_i = \Delta D_i' \underbrace{\psi}_{[\phi+\eta]} + \overbrace{(y_{i2}(j(i,2)) - \underbrace{\psi_{j(i,2)}}_{[\phi_{f(j(i,2))}+\eta_{u(j(i,2))}]}) - (y_{i1}(j(i,1)) - \underbrace{\psi_{j(i,1)}}_{[\phi_{f(j(i,1))}+\eta_{u(j(i,1))}]})}^{\text{unobserved "idiosyncratic" treatment effects heterogeneity} \equiv \Delta\varepsilon_i}$$

and fitting OLS to the equation above identifies the ATEs of firms $\{\phi_{f(1)}, \phi_{f(2)}, ..., \phi_{f(\mathcal{J})}\}$ and unions $\{\eta_{u(1)}, \eta_{u(12}, ..., \eta_{u(\mathcal{J})}\}$ up to normalizing constants. When $T = 2$, OLS in first differences equals OLS in levels with worker fixed effects, i.e., the UF-AKM specification provided in Equation (2).

D.3 Limited mobility bias

To explain the source of limited mobility bias in the UF-AKM, we start with a two-period UF-AKM without time-varying covariates, i.e.,

$$y_{it} = \alpha_i + \phi_{f(i,t)} + \eta_{u(i,t)} + \varepsilon_{it} \quad t \in \{1, 2\}.$$

Considering firm-only moves $\mathcal{F}$ (with cardinality N_f), the relative firm effects can be approximated by

$$\hat{\phi}_{f(k)} - \hat{\phi}_{f(\ell)} \approx \frac{1}{N_f} \sum_{i \in \mathcal{F}} (y_{i2} - y_{i1}).$$

Considering both moves $\mathcal{B}$ (with cardinality N_b) for the unions in the corresponding connected set of firm-only moves, the relative union effects can be approximated by

$$\hat{\eta}_{u(k)} - \hat{\eta}_{u(\ell)} = \frac{1}{N_b} \sum_{i \in \mathcal{B}} [(y_{i2} - y_{i1}) - (\hat{\phi}_{f(k)} - \hat{\phi}_{f(\ell)})].$$

In finite samples, there is a sampling error that remains. In the case of the firm effects

$$\begin{aligned} \hat{\phi}_{f(k)} - \hat{\phi}_{f(\ell)} \approx & \phi_{f(k)} - \phi_{f(\ell)} + \frac{1}{N_f} \sum_{i \in \mathcal{F}} (\varepsilon_{i2} - \varepsilon_{i1}) \\ \implies & \hat{\phi}_f \approx \phi_f + \hat{e}_{\mathcal{F}}. \end{aligned}$$

In the case of the union effects

$$\begin{aligned} \hat{\eta}_{u(k)} - \hat{\eta}_{u(\ell)} \approx & \eta_{u(k)} - \eta_{u(\ell)} + \frac{1}{N_b} \sum_{i \in \mathcal{B}} \left[(\varepsilon_{i2} - \varepsilon_{i1}) - \frac{1}{N_f} \sum_{i \in \mathcal{F}} (\varepsilon_{i2} - \varepsilon_{i1}) \right] \\ \implies & \hat{\eta}_u \approx \eta_u + \hat{e}_{\mathcal{B}} - \hat{e}_{\mathcal{F}}. \end{aligned}$$

And in the case of the person effects

$$\begin{aligned} \hat{\alpha}_i \approx & \alpha_i + \frac{1}{T} \sum_t \left[(\phi_{f(i,t)} - \hat{\phi}_{f(i,t)}) + (\eta_{u(i,t)} - \hat{\eta}_{u(i,t)}) + \varepsilon_{it} \right] \\ \implies & \hat{\alpha}_i \approx \alpha_i - (\hat{e}_{\mathcal{F}}) - (\hat{e}_{\mathcal{B}} - \hat{e}_{\mathcal{F}}) = \alpha_i - \hat{e}_{\mathcal{B}}. \end{aligned}$$

If there is a small number of moves relative to the number of fixed effects, the sampling errors can generate substantial bias in the variance and covariance terms. In our setting, $Var(\hat{e})_{\mathcal{F}} > Var(\hat{e}_{\mathcal{B}})$ because the ratio of both moves per union fixed effects is much larger than that of firm-only moves per firm fixed effect. Hence, the variance of the union effects has the largest potential for upward bias, i.e., $Var(\hat{\eta}_u) \approx Var(\eta_u) + Var(\hat{e}_{\mathcal{F}}) + Var(\hat{e}_{\mathcal{B}})$. Similarly, the union-to-firm covariance term has the largest potential for downward bias, i.e., $Cov(\hat{\phi}_f, \hat{\eta}_u) \approx Cov(\phi_f, \eta_u) - Var(\hat{e}_{\mathcal{F}})$. Since unions are assigned to workplaces independently from firm-premia in our setting, we expect $Cov(\phi_f, \eta_u) = 0$. Hence, the magnitude of $Cov(\hat{\phi}_f, \hat{\eta}_u)$ can be interpreted as $Var(\hat{e}_{\mathcal{F}})$, thereby capturing the extent of limited mobility bias. Table 1 shows that $Cov(\hat{\phi}_f, \hat{\eta}_u)$ explains only -0.5% of the variance in earnings, assuaging concerns about limited mobility bias in our UF-AKM.