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)

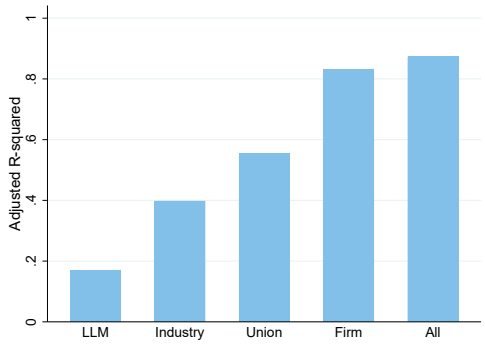
		Summary of regression results			
	Prob. of any	Relative to Low		Relative to Balanced	
	contribution	High	Mid	Compress	Exacerbate
	(1)	(2)	(3)	(4)	(5)
Union premia		Among Compress		Among High	
Post×Low	-0.087	0.017	0.000	-0.083	-0.095
	(0.021)	(0.023)	(0.024)	(0.021)	(0.024)
Post×High	-0.023				
	(0.020)				
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.

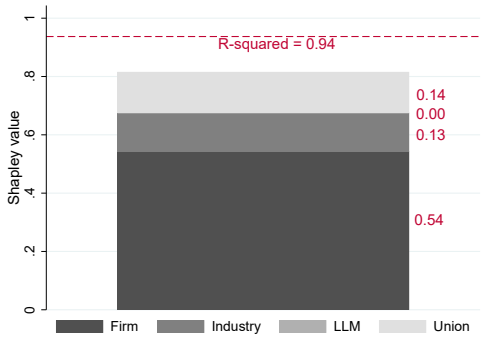
B Appendix Figures

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

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



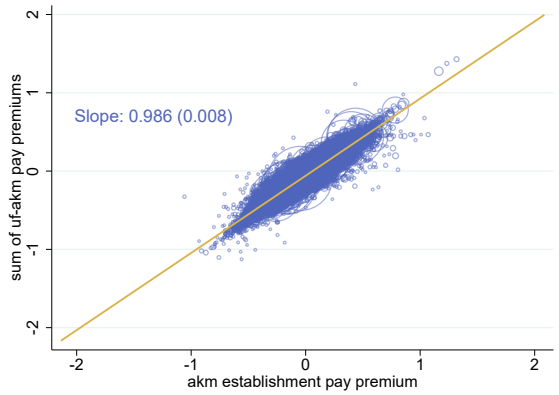
(b) Shapley values: multi-estab firms



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

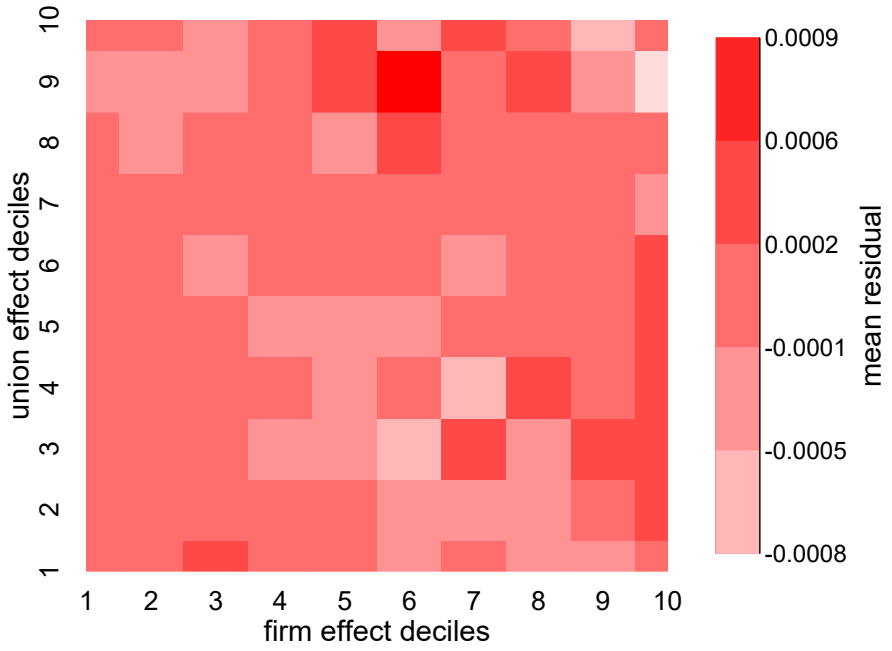
(a) Additivity without establishment effects



(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.

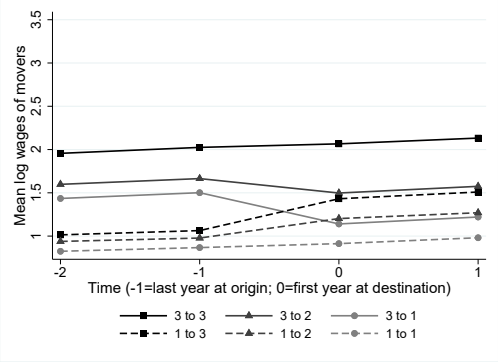
Figure B3: Pattern of Mean Residuals by $\hat{\phi}_{f(j)}$ and $\hat{\eta}_{u(j)}$ Deciles



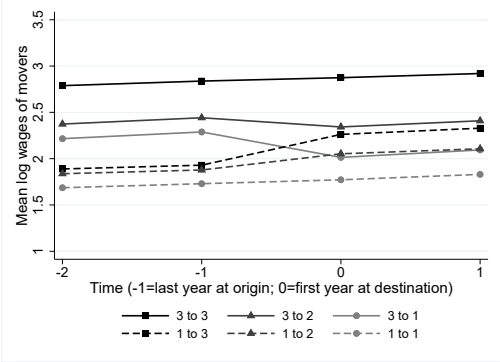
This figure shows a heat map of the mean UF-AKM residuals by cells of $\hat{\phi}_f$ and $\hat{\eta}_u$ deciles. The range of mean residuals is from -0.08 to 0.09 log points and is split into quintiles: darker hues imply higher values (as shown in the legend). Each cell is assigned the hue that corresponds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

Figure B4: Mover Event Studies by Type of Mover

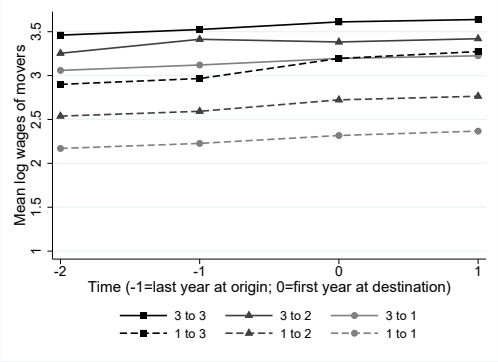
(a) Both movers



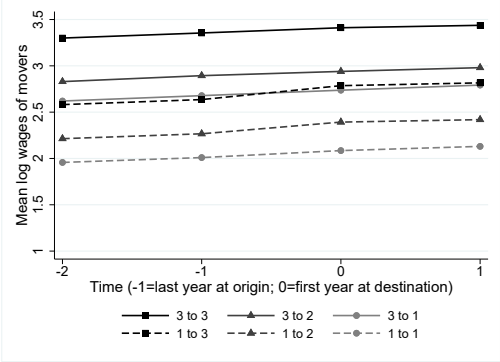
(b) Firm-only movers



(c) Union-only movers



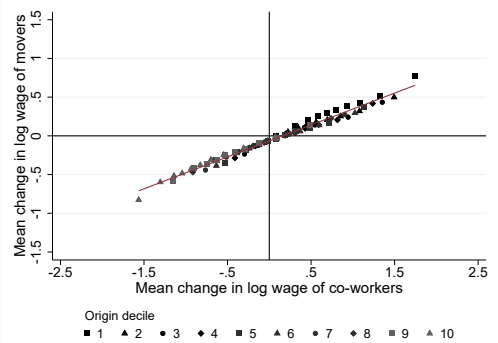
(d) Neither movers



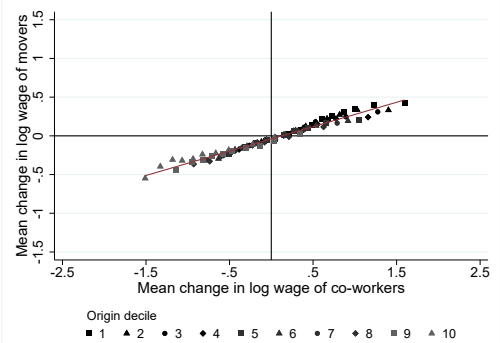
These figures display the evolution of wages for workers who moved from origin establishments in the top and bottom tercile groups to destination establishments in any of the other tercile groups. The movers are defined as workers at establishments who separated from the origin establishment, were reemployed 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

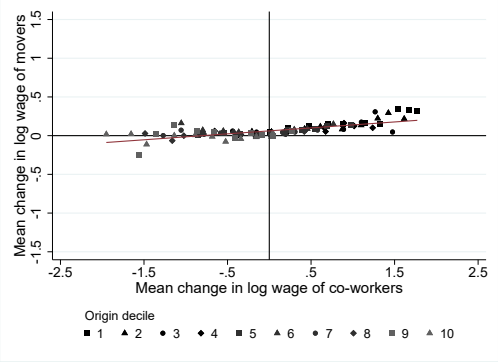
(a) Both movers



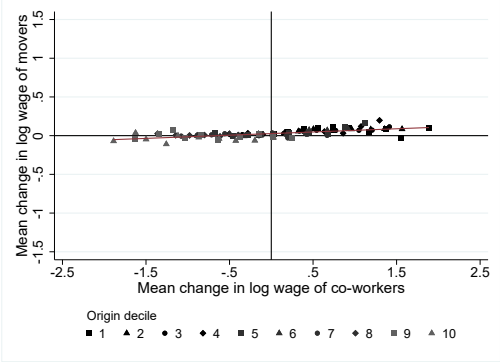
(b) Firm-only movers



(c) Union-only movers



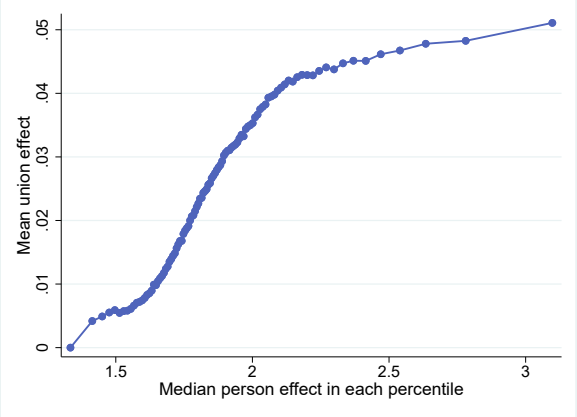
(d) Neither movers



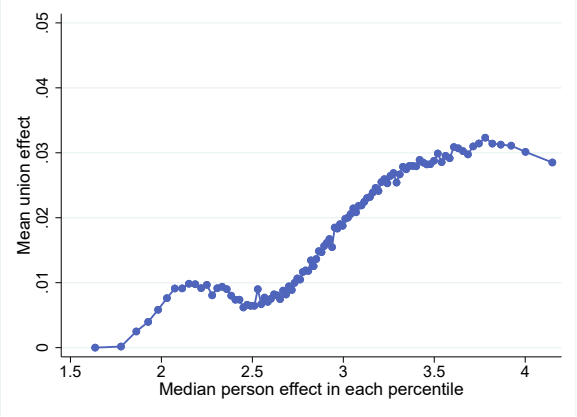
These figures plot the mean change in movers' wages between the years prior to separation and after hiring against the mean change in co-worker wages between origin and destination establishments. Origin and destination establishments are grouped by deciles of co-worker wages; each of the 10x10 dots 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



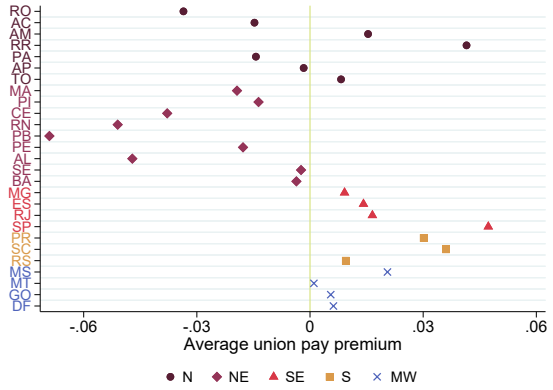
(b) UF-AKM for high skill jobs



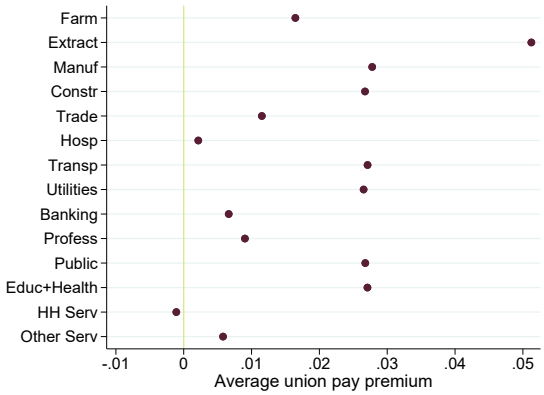
These figures depict positive assortative matching of workers to unions along percentiles of worker skill. Figure B6a uses union effects from a UF-AKM restricted to low skill jobs, i.e., those in occupations whose average AKM person effect is in the bottom tercile. Figure B6b uses union effects from 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.

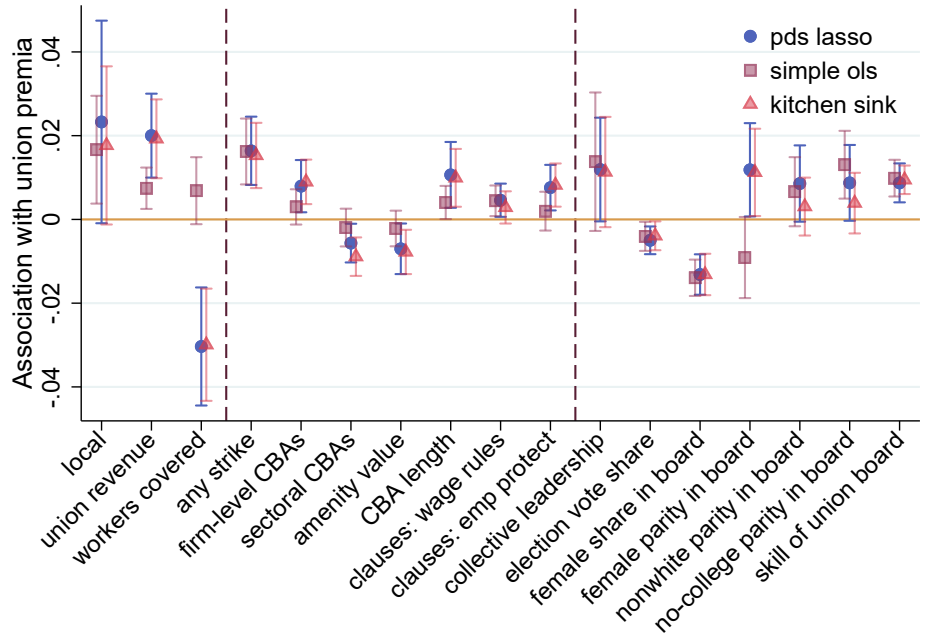
Figure B8: Effect of Reform Across Markets with Varying Union Premia

(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.

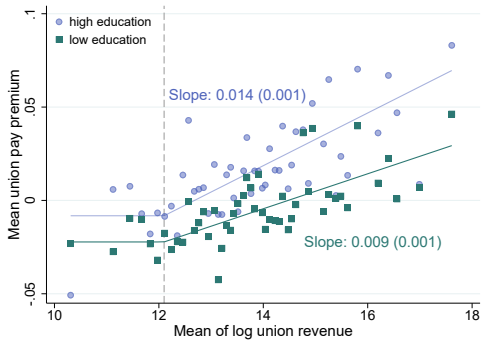
Figure B9: Correlates of Union Pay Premia: Including State and Sector Fixed Effects



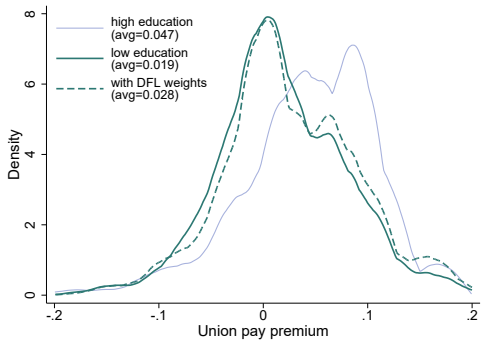
This figure shows the estimates of coefficients correlating union-specific pay premia with union attributes (denoted in the x-axis), accounting for state and sector fixed effects. For each attribute, estimates for three separate analyses are reported: 1) *kitchen sink:* a kitchen-sink regression of the 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

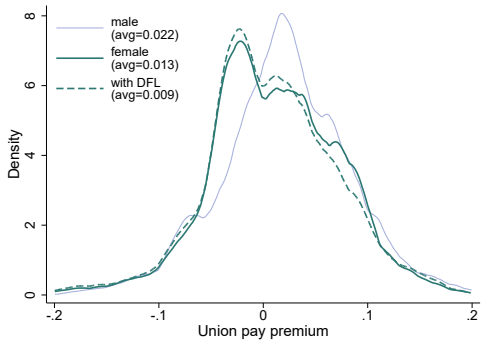


(b) Distributions by education

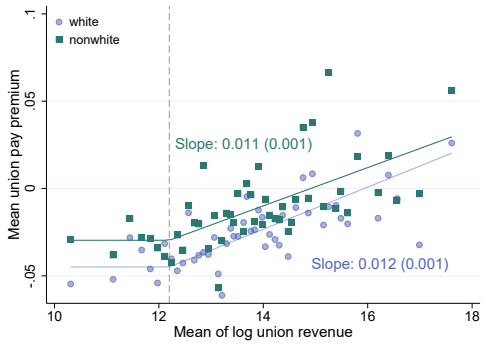


(c) Normalization by gender

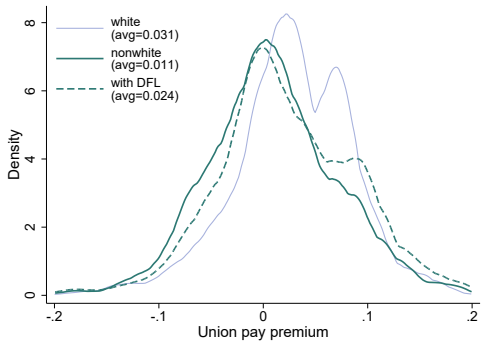
(d) Distributions by gender



(e) Normalization by race

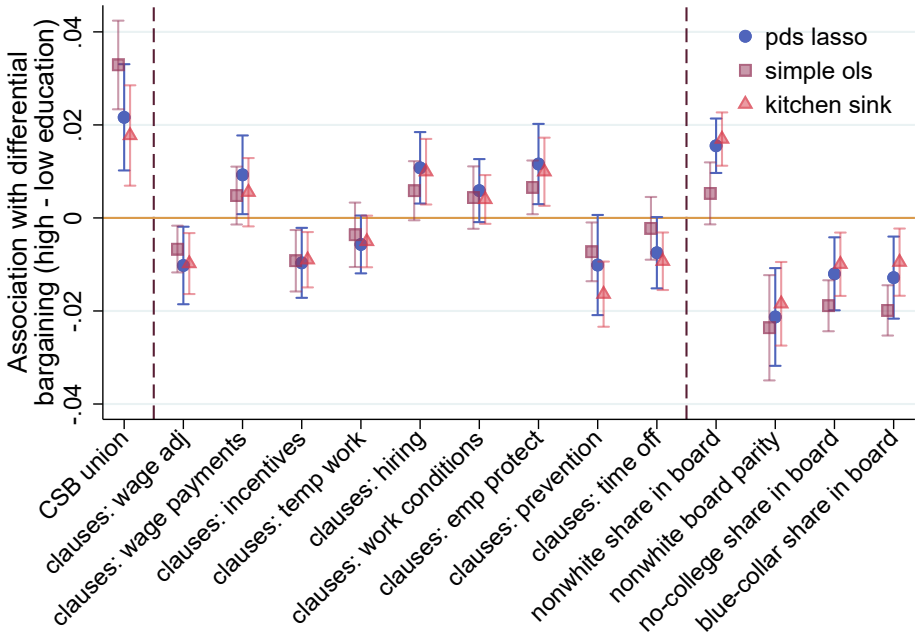


(f) Distributions by race



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.

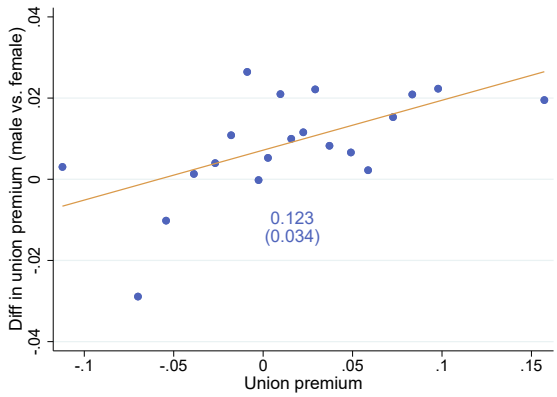
Figure B11: Correlates of Differential Bargaining by High vs. Low Educated Workers



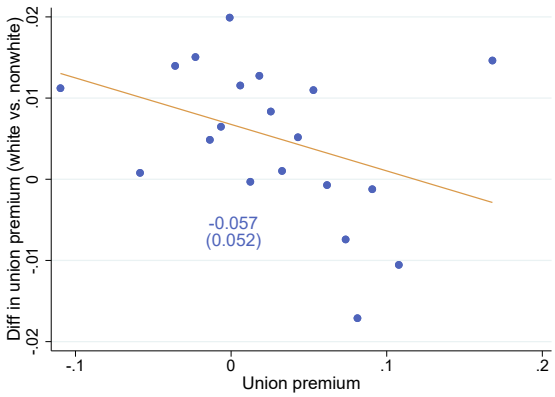
This figure shows the estimates of coefficients correlating union-specific bargaining differentials for high- versus low-educated workers with union attributes (denoted in the x-axis). For each attribute, estimates for three separate analyses are reported: 1) *simple ols:* a simple bivariate OLS regression 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



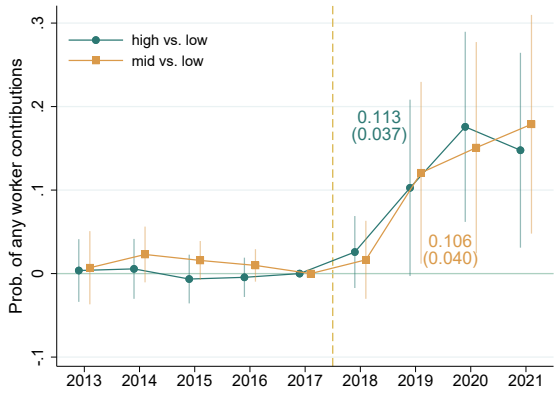
(b) Racial gaps



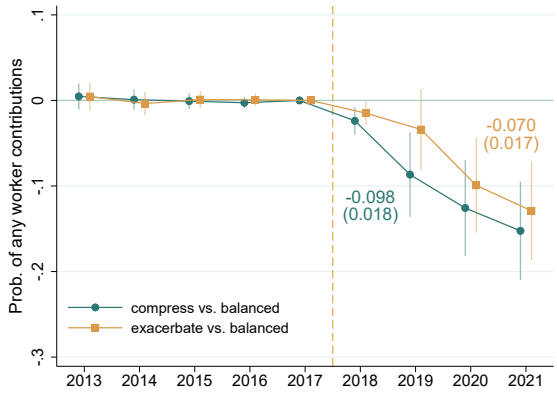
These figures depict binned scatterplots of the differential bargaining effects by gender (Figure B12a) and race (Figure B12b) on the union pay premium, including the linear prediction from an OLS. The estimates for the slope coefficients and their standard errors are reported next to each scatterplot. 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.

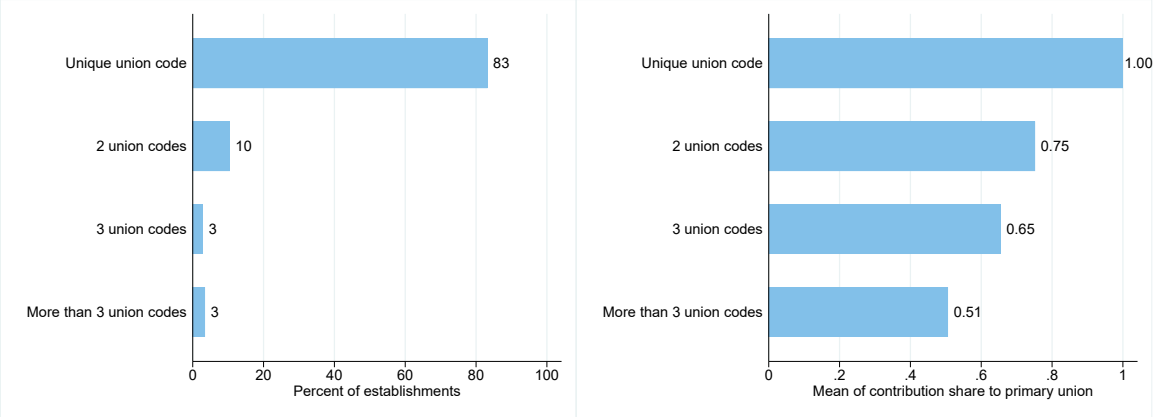
C Data Appendix

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

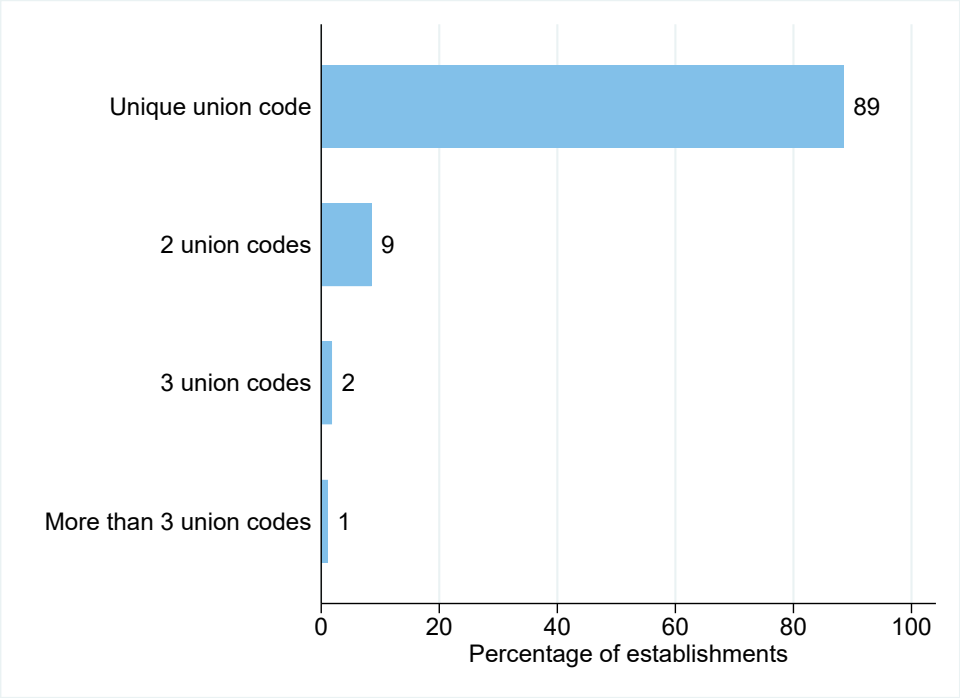
(a) Establishment-to-union mapping (b) Contribution share to primary union



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.