Matching and Local Labor Market Size in Mexico

Jorge Pérez Pérez[1] Jorge Meléndez[1] José Nuño-Ledesma[2]

[1]Banco de México
[2]University of Guelph

LACEA-LAMES 2024
November 14th, 2024

Disclaimers

- The views and conclusions presented in this document are the exclusive responsibility of the authors and do not necessarily reflect those of Banco de Mexico.
- We acknowledge financial support from Colombia Cientifica – Alianza EFI 60185 contract FP44842- 220-2018, funded by The World Bank through the Scientific Ecosystems, managed by the Colombian Ministry of Science, Technology, and Innovation (MINCIENCIAS).
- The data were accessed through the Econlab at Banco de Mexico. The EconLab collected and processed the data as part of its effort to promote evidence-based research and foster ties between Banco de Mexico's research staff and the academic community. Inquiries regarding the terms under which the data can be accessed should be directed to econlab@banxico.org.mx

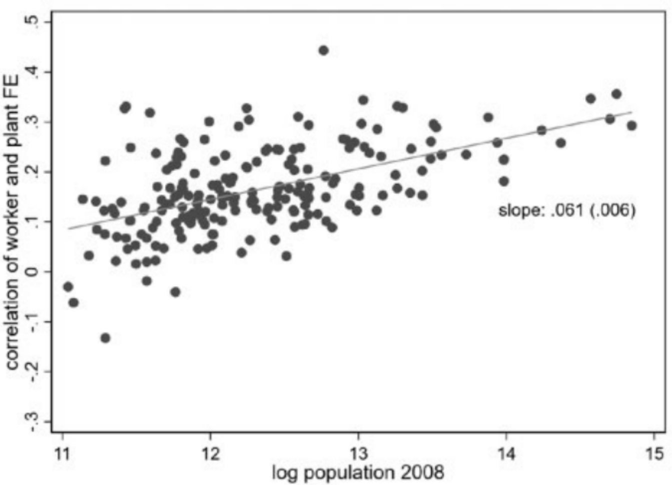
Summary

- **What we do**
  - Measure the extent of agglomeration externalities due to better labor market matching in Mexico.
  - Measure the relationship between city size and assortative matching in Mexico's labor markets.
- **How do we do it?**
  - Use an administrative dataset with the near universe of formal sector workers in Mexico.
  - Estimate models for log-wages with additive worker and workplace effects.
  - Calculate the correlation between worker and workplace effects as a measure of positive assortative matching.
  - Correlate assortative matching and city size.
- **What do we find?**
  - Similar/weaker relationship between market size and matching in Mexico.
  - Labor informality is negatively associated with positive assortative matching in formal labor markets.

Motivation: City Size Wage Gaps

- Workers in larger cities enjoy wage premia even after controlling for worker characteristics (Baum-Snow and Pavan, 2012; Card et al., 2013; Combes and Gobillon, 2015; Gould, 2007)
- One agglomeration force behind these premia is easier worker-workplace matching (D'Costa and Overman, 2014; Dauth et al., 2022; Duranton and Puga, 2004)
  - Positive assortative matching: High-wage workers go to high-wage firms.

Motivation: City Size and Positive Assortative Matching in Germany's Labor Markets



Source: Dauth et al. (2022)

Motivation: Matching Externalities in Developing Countries

Several reasons to think that matching externalities are different in developing economies:

- Large labor informality.
  - May weaken agglomeration forces coming from reduced search costs in large markets (Henderson, 1986; Helsley and Strange, 1990; Petrongolo and Pissarides, 2006).
  - Weakens the advantages of acquiring specialized human capital (Rotemberg and Saloner, 2000; Wheeler, 2008; Bleakley and Lin, 2012).
  - May reduce the productivity advantages of good matches.
- Workplace level determinants more important for wage inequality in developing countries (Bassier, 2023; Diallo et al., 2022; Frías et al., 2022; Pérez Pérez and Nuño-Ledesma, 2024).

Contribution

- **City size wage gaps in developed and developing economies**
  - Developed economies: Baum-Snow and Pavan (2012); Gould (2007); D'Costa and Overman (2014); De la Roca and Puga (2016).
  - Developing economies: Chauvin et al. (2017); Combes et al. (2020); De la Roca et al. (2023); Duranton (2016).
  - **Contribution:** Estimates for Mexico.
- **Matching in labor markets, agglomeration, and informality**
  - Matching in labor markets: Andersson et al. (2007); Baum-Snow and Pavan (2012); Behrens et al. (2014); **Dauth et al. (2022)**
  - Informal and formal labor markets: Ulyssea (2010); Levy Algazi (2018); Ulyssea (2018)
  - **Contribution:** Study these agglomeration forces in a developing country and show that informality weakens them.
- **AKM models for Mexico's labor markets:**
  - Frías et al. (2022); Pérez Pérez and Nuño-Ledesma (2024).

Data

- Monthly Social security records from *Instituto Mexicano del Seguro Social* (IMSS) Nov 2004 - Dec 2018.
- Number of observations within the range of 12.8 million (Nov 2004) and 20.1 million (Dec 2018).
- 83% of private-sector formal workers are affiliated with IMSS (as of 2022).
  - IMSS does not collect information from workers employed by the government or working in the informal economy.
- We restrict our analysis to prime-age men (25-54 years old).

- Informality and education data from censuses and labor surveys.

Data

- Key variables:
  - **Worker ID**: Social security number.
  - **Workplace ID**: *Registro patronal*.
  - **Wage**: Daily taxable income.
  - **Other**: Year of birth, gender.
- Data are bottom-coded (minimum wage) and top-coded (About 12.5 minimum wages).
- The data does not include part-time working status or education variables.
- One employee can be registered as working for more than one employer.
- An employer-employee pair can appear more than once in a month with different income.
- We pair only one job per worker: whichever reports the highest income.

Descriptive Statistics: Workers, Prime-Age men (25-54 y.o.), National Level

		Real wage		
	(1)	(2)	(3)	(4)
	Observations	Mean	Std. dev	Percent censored
2005	73,847,545	394.589	406.167	2.675
2009	80,065,916	394.602	402.992	2.690
2014	96,354,574	394.200	409.212	2.649
2018	110,844,774	401.186	412.367	2.058

Observations correspond to the sum of all the monthly observations in a year. Real wages using prices of July 2018. Percent censored is the percentage of observations with wages exactly equal to the upper wage limit.

Methodology

1. Estimate AKM model with worker and workplace fixed effects.
2. Measure correlation between worker and workplace fixed effects at the city level, and regress it on city size.

AKM Model (Abowd, Kramarz, and Margolis 1999)

$$\ln W_{it} = \alpha_i + \psi_{\mathbf{J}(i,t)} + X'_{it}\beta + r_{it} \tag{1}$$

Where

- W_{it} is the real wage of worker i at period t.
- α_i are worker effects. Factors that are rewarded equally across employers giving rise to a worker-specific wage component.
- $\psi_{\mathbf{J}(i,t)}$ are establishment effects. Proportional wage premium (or discount) that is paid by firm J to all employees.
- x'_{it} is a vector of observable worker characteristics. We include age, age squared, age cube, and a time trend.
- r_{it} is the error term.

AKM Model - Variance and Assortative Matching

The variance of wages can be decomposed as follows:

$$\begin{aligned}\mathrm{Var}\,(\ln W_{it}) = {} & \underbrace{\mathrm{Var}(\alpha_i)}_{\text{workers}} + \underbrace{\mathrm{Var}(\psi_{\mathbf{J}(i,t)})}_{\text{firms}} + \mathrm{Var}(x'_{it}\beta) + \mathrm{Var}(r_{it}) \\ & + 2\,\underbrace{\mathrm{Cov}(\alpha_i, \psi_{\mathbf{J}(i,t)})}_{\text{sorting}} + 2\,\mathrm{Cov}(\psi_{\mathbf{J}(i,t)}, x'_{it}\beta) + 2\,\mathrm{Cov}(\alpha_i, x'_{it}\beta).\end{aligned} \tag{2}$$

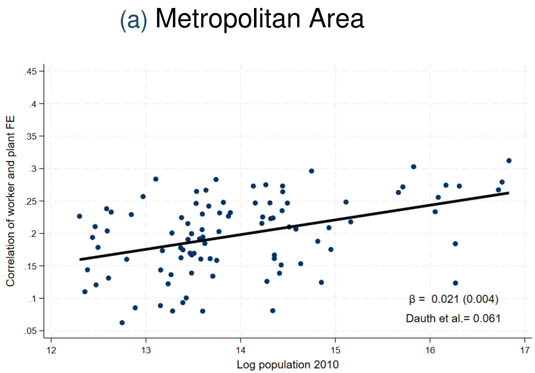
Positive covariance of α_i and $\psi_{\mathbf{J}} \to$ positive assortative matching, i.e. high quality workers tend to be matched with high quality firms.

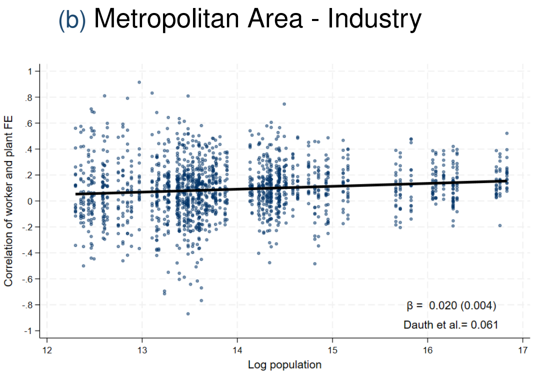
AKM Model Estimation Results

	Interval1 2004-2008	Interval2 2009-2013	Interval3 2014-2018
Worker and workplace parameters			
Number of worker effects	11,363,073	13,083,589	15,512,438
Number of workplace effects	858,480	892,929	1,009,320
Summary of parameter estimates			
St. dev. of worker effects	0.539	0.520	0.503
St. dev. of workplace effects	0.463	0.493	0.503
St. dev. of residuals	0.195	0.198	0.200
Correlation worker/workplace effects	0.208	0.226	0.262
Correlation worker effects/Xb	-0.079	-0.034	-0.067
Correlation workplace effects/Xb	-0.002	0.008	0.003
Goodness of fit			
St. dev. of log wages	0.808	0.823	0.829
R Squared	0.942	0.942	0.942
Adj. R Squared	0.939	0.940	0.940

Source: Authors' calculations using IMSS data. Results from estimating equation (1) via OLS with a pre-conditioned gradient method following Card et al. (2013). Estimations are restricted to prime-aged men (ages 25-54) in the largest connected set per time interval. All the estimations include the following controls: age, age squared, age cube, and a monthly time trend. RMSE is the root mean squared error.

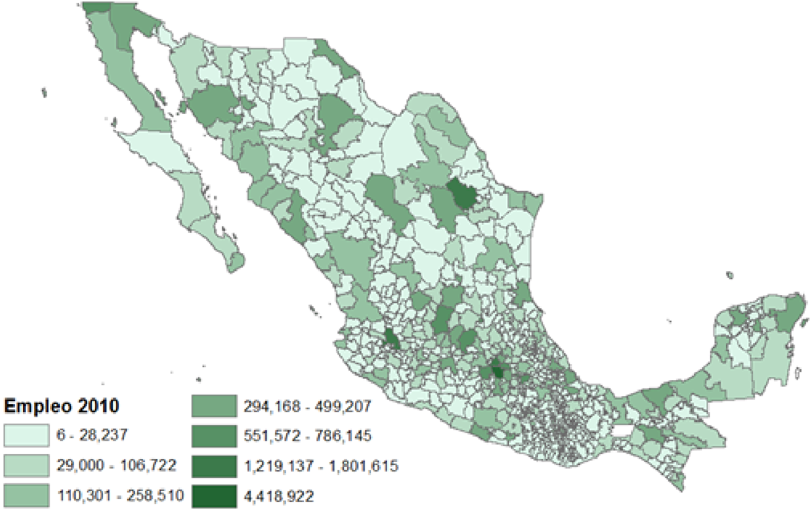
Assortative Matching and City Size





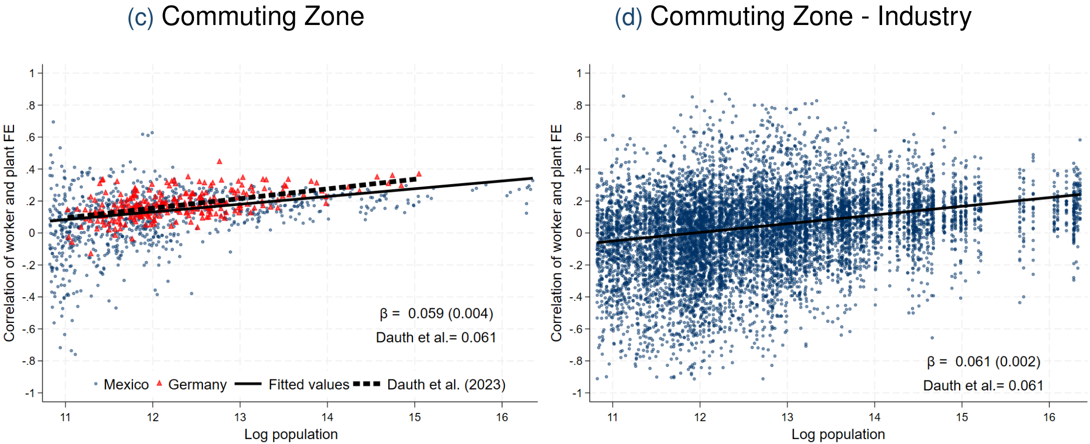
Source: Author's calculations using IMSS data. Each panel shows a scatterplot for the relationship between log population and the correlation between estimated worker and workplace effects from AKM models at different geographical aggregation levels. For comparison, each panel also includes the coefficient for the relationship estimated by Dauth et al. (2022). We classify industries according to a 3-digit NAICS classification. The bottom-right values display the slope of a linear regression corresponding to the displayed relationship. The regressions include dummies for each time interval.

Aside: Local Labor Markets for Mexico



Source: Aldeco et al. (2023)

Assortative Matching and City Size



Source: Author's calculations using IMSS data. Each panel shows a scatterplot for the relationship between log population and the correlation between estimated worker and workplace effects from AKM models at different geographical aggregation levels. Panels c and d restrict to commuting zones with more than 50,000 inhabitants. Panel c also shows the relationship estimated for Germany by Dauth et al. (2022). For panel d, we restrict to cells with more than five firms and more than 50 workers. We classify industries according to a 3-digit NAICS classification. The bottom-right values display the slope of a linear regression corresponding to the displayed relationship. The regressions include dummies for each time interval.

Assortative Matching and City Size

Dependent variable: correlation of worker and plant FE				
	(1)	(2)	(3)	(4)
	Metro	Metro-Industry	CZ	CZ-Industry
A: Baseline Model				
Log Population	0.0205***	0.0202***	0.0589***	0.0614***
	(0.004)	(0.004)	(0.004)	(0.002)
R^2	0.358	0.026	0.167	0.094
B: Correlation of worker and residual workplace FE				
Log Population	0.0184***	0.0202***	0.0041***	0.0034***
	(0.005)	(0.004)	(0.000)	(0.000)
R^2	0.375	0.026	0.136	0.073
C: Log population instrumented with population in 1921-1950				
Log Population	0.0282***	0.0107	0.0400**	0.0523***
	(0.007)	(0.010)	(0.019)	(0.019)
R^2	0.338	0.023	0.154	0.092
First-stage F	39.685	10.148	12.240	3.001
N	97	1,648	877	9,174

Source: Author's calculations using IMSS and census data. The columns show the results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log population at different geographical aggregation levels. All the regressions pool data from the three intervals: 2004-2008, 2009-2013, and 2014-2018, and include dummies by interval. Metro stands for metropolitan area. CZ stands for commuting zone. Columns 3 and 4 are restricted to commuting zones with more than 50,000 inhabitants. For column 4, we restrict to cells with more than five firms and more than 50 workers. Panel A shows baseline estimates. Panel B shows estimates with workplace fixed effects demeaned by industry. Panel C instruments log population in the three intervals with log population in 1921, 1930, 1940; and 1950. First-stage F is the first stage F-statistic. Robust standard errors in parentheses. *: $p<0.1$, **: $p<0.05$, ***: $p<0.01$.

Assortative Matching and City Size. Additional Regressions.

Dependent variable: correlation of worker and plant FE				
	(1) Metro	(2) Metro-Industry	(3) CZ	(4) CZ-Industry
D: Corrected for limited mobility bias				
Log Population	0.0173***	0.0037	0.0454***	0.0456***
	(0.004)	(0.004)	(0.004)	(0.002)
R^2	0.245	0.011	0.117	0.069
N	97	1,647	877	9,163
E: Dropping the 10% largest areas				
Log Population	0.0195***	0.0243***	0.0594***	0.0605***
	(0.006)	(0.005)	(0.005)	(0.002)
R^2	0.366	0.053	0.163	0.075
N	58	986	874	7,897

Source: Author's calculations using IMSS and census data. The columns show the results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log population at different geographical aggregation levels. All the regressions pool data from the three intervals: 2004-2008, 2009-2013, and 2014-2018, and include dummies by interval. Metro stands for metropolitan area. CZ stands for commuting zone. Columns 3 and 4 are restricted to commuting zones with more than 50,000 inhabitants. For column 4, we restrict to cells with more than five firms and more than 50 workers. Panel D shows estimates using the Bonhomme et al. (2019) correction, where workplaces are clustered into five clusters according to within-workplace wage distributions before estimating the AKM model. Panel E shows estimates excluding the 10% largest areas by population. Robust standard errors in parentheses. *: $p<0.1$, **: $p<0.05$, ***: $p<0.01$.

Assortative Matching, Unionization, Informality, and Education

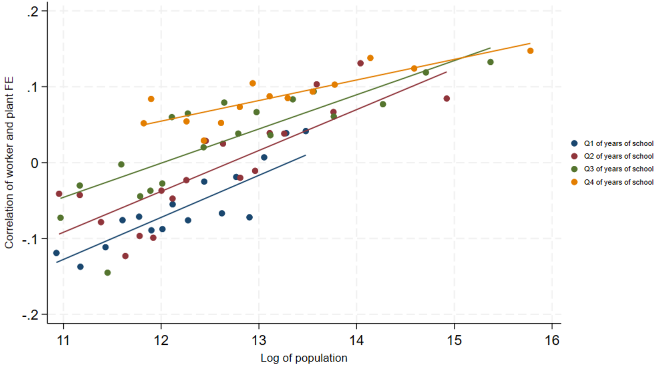
Dependent Variable: Correlation of worker and workplace FE

	(1) City-Industry	(2) City-Industry	(3) CZ	(4) CZ	(5) CZ	(6) CZ-Industry	(7) CZ-Industry	(8) CZ-Industry
Log population	0.0202***	0.0179***	0.0589***	0.0385***	0.2061**	0.0614***	-0.0193	-0.0144
	(0.004)	(0.004)	(0.004)	(0.005)	(0.088)	(0.002)	(0.048)	(0.031)
Informality Rate		-0.3479***		-0.3777***	-0.0177			
		(0.026)		(0.058)	(0.204)			
Unionization Rate		-0.0234						
		(0.035)						
Years of Education				0.0109	-0.0369			
				(0.011)	(0.026)			
CZ FE					Yes		Yes	
CZ-Industry FE								Yes
N	1,648	1,648	877	1,960	1,909	9,174	9,174	8,890
R^2	0.026	0.182	0.167	0.197	0.663	0.094	0.187	0.798

Source: Author's calculations using IMSS, ENOE, and census data. The columns show the results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log population at different geographical aggregation levels and with different specifications. Metro stands for metropolitan area. CZ stands for commuting zone. We restrict to CZs with more than 50,000 inhabitants. For columns 6 to 8, we restrict to cells with more than five firms and more than 50 workers. Robust standard errors in parentheses. *: $p<0.1$, **: $p<0.05$, ***: $p<0.01$.

The Role of Education

Figure: By quartiles of education

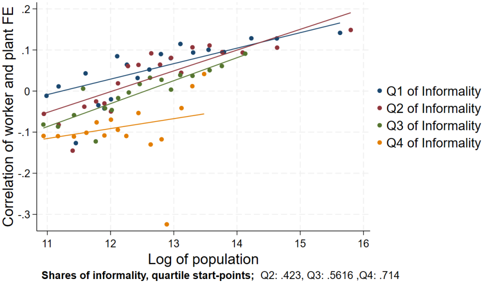


Shares of years of school, quartile start-points; Q2: 5.779, Q3: 6.617 ,Q4: 7.4

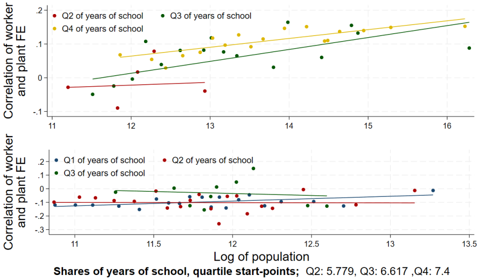
Source: Author's calculations using IMSS data. The figure shows a binned scatterplot for the relationship between log population and the correlation between estimated worker and workplace effects from AKM models at the commuting zone-industry level absorbing period fixed effects. We restrict to CZs with more than 50,000 inhabitants and cells with more than five firms and more than 50 workers.

The Role of Informality

(a) By quartiles of informality



(b) Low informality (top) and high informality (bottom) by quartiles of education



Source: Author's calculations using IMSS data. Each panel shows a binned scatterplot for the relationship between log population and the correlation between estimated worker and workplace effects from AKM models at the commuting zone-industry level absorbing period fixed effects. We restrict to CZs with more than 50,000 inhabitants and cells with more than five firms and more than 50 workers.

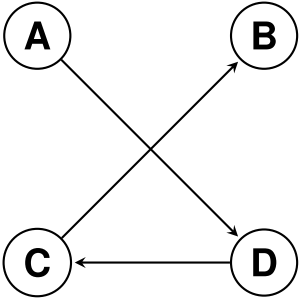
Why does Informality Weaken Matching in the Formal Sector?

- May weaken agglomeration forces coming from reduced search costs in large markets (Henderson, 1986; Helsley and Strange, 1990; Petrongolo and Pissarides, 2006).
  - Harder to find good matches among informal workers.
  - Fewer information about informal workers for formal firms if markets are segmented.
  - Harder to validate credentials of informal workers looking to enter the formal sector.
- Weakens the advantages of acquiring specialized human capital (Rotemberg and Saloner, 2000; Wheeler, 2008; Bleakley and Lin, 2012).
  - If the informal market is the largest employer, there are fewer incentives for workers to acquire human capital that may generate matches with high-paying formal firms.
  - Higher rotation across occupation for informal workers reduces learning-by-doing.

Concluding Remarks

- City size is positively correlated with the intensity of positive assortative matching in Mexico.
- Some evidence of this relationship being weaker in Mexico.
- The presence of informality weakens matching and the advantages of market size for matching in the formal labor market.

Connected Set



Source: Fenizia (2019)

- The firm and worker effects in (1) are separately identified within a "connected set" of firms linked by worker mobility.
- We restrict analysis to the largest connected set in four time intervals: 2004-2008, 2009-2013, and 2014-2018.
- The ratio of observations in largest connected set to all observations ranges between 94.9% and 97.3% (and between 97.5% ad 98.6% individuals).

Connected Set

	All sample				Individuals in largest connected set			
			Log wage				Log wage	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Interval	All obs.	Workers	Mean	Std. dev.	All obs.	Workers	Mean	Std. dev.
Nov 2004-2008	324,468,447	11,835,313	5.627	0.813	311,941,032	11,363,073	5.657	0.808
Ratio: largest connected/all					96.14	96.01	100.53	99.39
2009-2013	431,227,399	13,526,466	5.600	0.826	417,008,147	13,083,589	5.625	0.823
Ratio: largest connected/all					96.70	96.73	100.45	99.65
2014-2018	518,128,252	15,920,775	5.609	0.831	505,015,793	15,512,438	5.628	0.829
Ratio: largest connected/all					97.47	97.44	100.35	99.71
Change from first to last interval			-0.018	0.018			-0.029	0.021

Exchangeability

The assumption that assignment of workers is uncorrelated with the unobserved ability of the worker and the unobserved productivity of the firm.

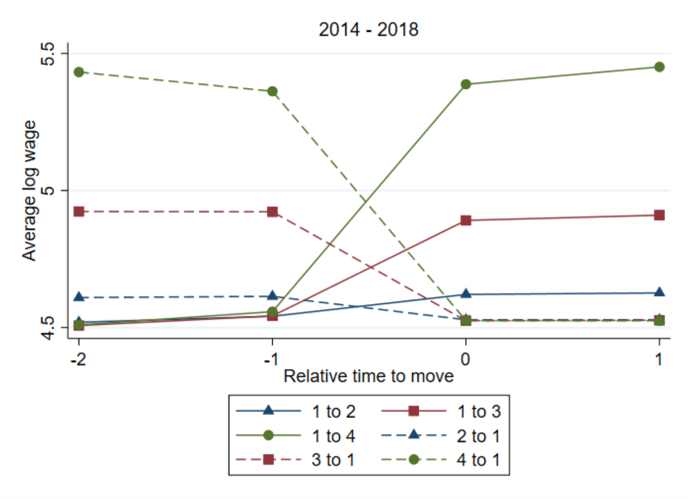
If violated:

- Workers sort into firms based on unobserved characteristics.
- Observed wages might reflect this non-random sorting, biasing our estimates.

If satisfied:

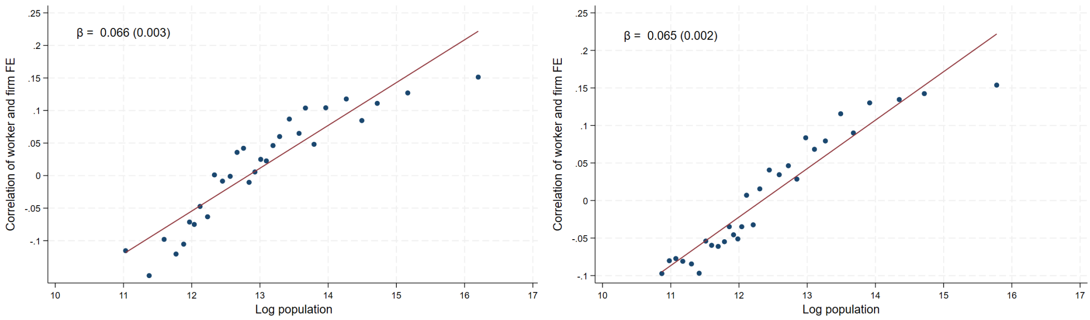
- A worker moving from A to B, should experience a wage change of equivalent magnitude but opposite sign to that experienced by someone moving from B to A.

Exchangeability



Assortative Matching and City Size by sector

(c) Services (d) Non-services



Source: Author's calculations using IMSS data. Each panel shows a binned scatterplot for the relationship between log population and the correlation between estimated worker and workplace effects from AKM models at the commuting zone-industry level. We restrict to CZs with more than 50,000 inhabitants and cells with more than five firms and more than 50 workers.