Matching and Local Labor Market Size in Mexico

Jorge Pérez Pérez[1] Jorge Meléndez[2] José Nuño-Ledesma[3]

[1]Banco de México
[2]Vancouver School of Economics
[3]University of Guelph

BIS CCA Research Conference on Labor Markets
October 29th, 2025

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

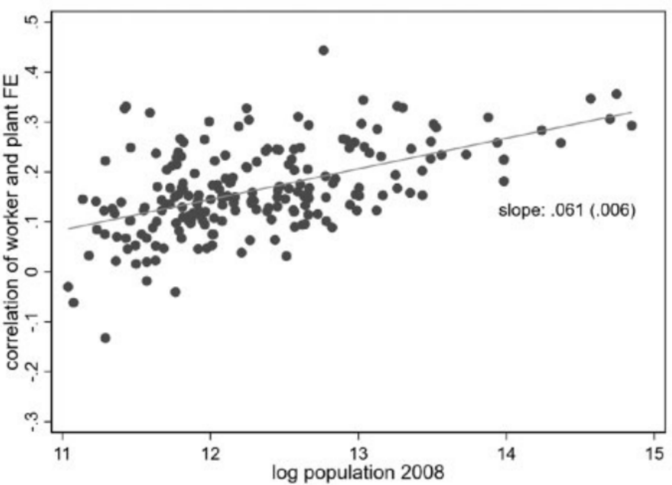
Summary

- **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?**
  - Positive relationship between city 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:

- Negative externalities of agglomeration in developing countries, such as pollution, crime, and congestion (Akbar et al., 2023; Grover et al., 2023)
- Poor transportation infrastructure reduced the scope for matching externalities (Baum-Snow et al., 2017; Ghani et al., 2016)
- Large labor informality (55.4% in Mexico) may hinder development and discourage human capital acquisition (Jedwab et al., 2022; Duranton, 2015)
- 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 (November 2004) and 20.1 million (December 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

	(1)	Real wage (2)	Real wage (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 local labor market level, and regress it on local labor market 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:

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

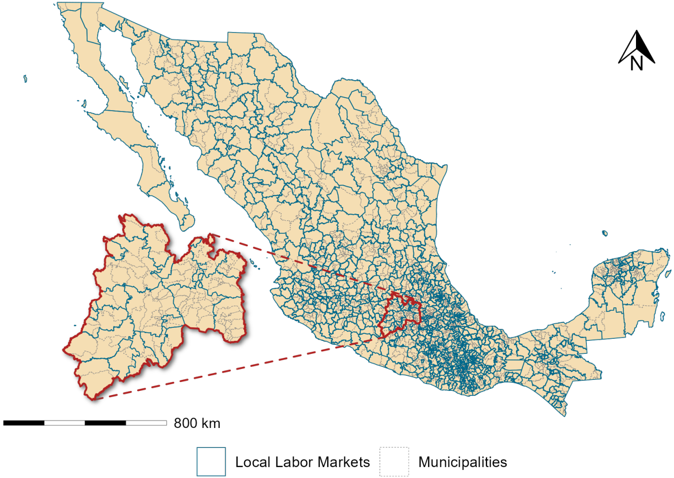
Positive covariance of α_i and $\psi_{\mathbf{J}} \rightarrow$ positive assortative matching, i.e. high quality workers tend to be matched with high quality firms

AKM Model Estimation Results

	Interval 1 2004-2008	Interval 2 2009-2013	Interval 3 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
Matching metric: 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
RMSE	0.195	0.198	0.200
R Squared	0.942	0.942	0.942
Adj. R Squared	0.939	0.940	0.940
Total variance shares			
Worker effects	0.444	0.398	0.369
Workplace effects	0.328	0.359	0.369
2Cov(worker effects, workplace effects)	0.159	0.171	0.193
Remainder	0.069	0.072	0.069

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.

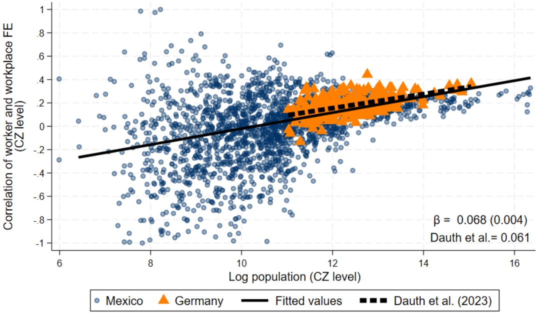
Aside: Local Labor Markets for Mexico



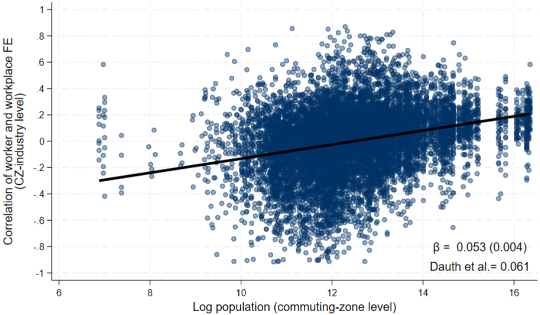
Source: Aldeco et al. (2024)

Assortative Matching and City Size

(a) Commuting Zone



(b) Commuting Zone - Industry



Source: Author's calculations using IMSS and INEGI data. Each panel displays a scatterplot illustrating the relationship between log population and the correlation between estimated worker and workplace effects from AKM models at the commuting zone and commuting zone industry levels. For comparison, panel (a) displays the relationship estimated for Germany by Dauth et al. (2022). For panel (b), we restrict to cells with more than five firms and more than 50 workers. We classify industries according to a 2-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. Clustered standard errors at the commuting-zone level in parentheses.

Assortative Matching and City Size

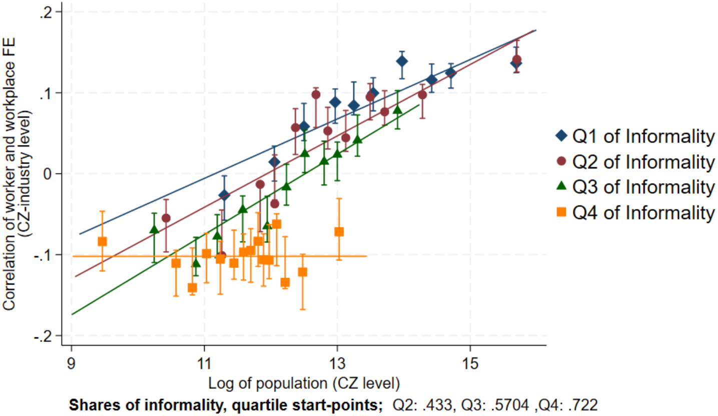
Dependent variable: Matching metric - Correlation of worker and workplace FE		
	(1)	(2)
	CZ	CZ-Industry
A: Baseline Model		
Log Population	0.0679***	0.0531***
	(0.004)	(0.004)
R^2	0.155	0.090
B: Correlation of worker and residual workplace FE		
Log Population	0.0640***	0.0529***
	(0.004)	(0.004)
R^2	0.146	0.090
C: Log population instrumented with population in 1921-1950		
Log Population	0.0592**	0.0420***
	(0.009)	(0.013)
R^2	0.152	0.086
First-stage F	174.786	4.757
N for Panels A-C	1,961	10,118
Mean of dep. variable	-0.014	0.000

Source: Author's calculations using IMSS and INEGI data. "CZ" stands for commuting zone. The regressions pool data from 2004-2008, 2009-2013, and 2014-2018 with interval dummies. Column (2) restricts to cells with over five firms and 50 workers. Panels show the following estimates: A - baseline; B - industry-demeaned workplace fixed effects; C - we instrument population with historical population at the CZ level, relying on historical population estimates from Alix-Garcia and Sellars (2020). Log population instrumented with log population in 1921, 1930, 1940, and 1950 (Table shows different specifications of the historical population IV). Clustered standard errors at the commuting-zone level in parentheses. *: $p<0.1$, **: $p<0.05$, ***: $p<0.01$.

Informality and assortative matching

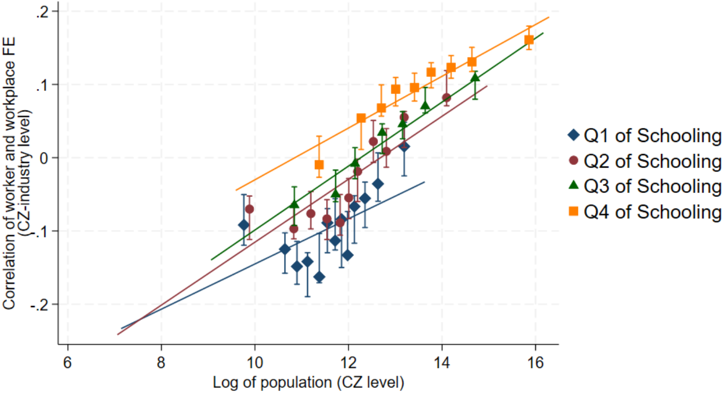
- Increased search costs for firms looking for good matches (Henderson, 1986; Helsley and Strange, 1990; Petrongolo and Pissarides, 2006)
- Lower incentives to acquire specialized human capital (Rotemberg and Saloner, 2000; Wheeler, 2008; Bleakley and Lin, 2012)
- Weaker incentives for formal firms to find good matches

The Role of Informality



Source: Author's calculations using IMSS and INEGI data. The figure displays a binned scatter plot of the log population and our matching metric at the commuting zone-industry level for each quartile of the informality rate. The vertical bars are confidence intervals for the conditional mean of the correlation at each level of (log) population. We used the `binsreg` and `binstest` commands (Cattaneo et al., 2024a,b) with default settings to generate the scatter plots.

The Role of Schooling

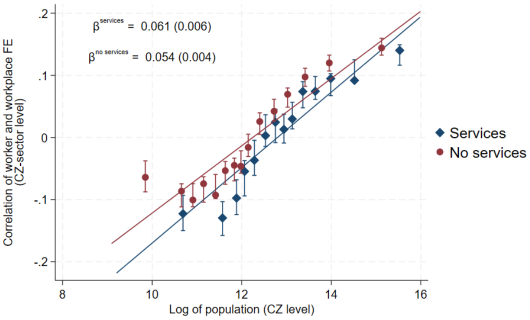


Mean years of schooling, quartile start-points; Q2: 5.856, Q3: 6.8335 ,Q4: 7.676

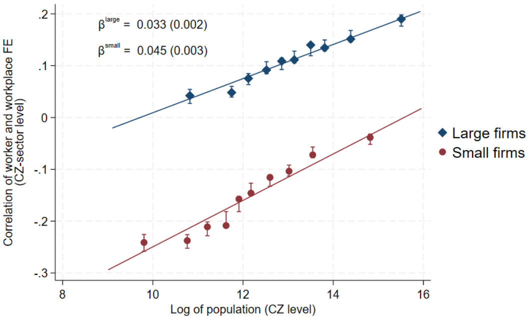
Source: Author's calculations using IMSS and INEGI data. The figure displays a binned scatter plot of the log population and the estimated matching measure at the commuting zone-industry level for each quartile of the mean years of schooling. The vertical bars are confidence intervals for the conditional mean of the correlation at each level of (log) population. We used the `binsreg` and `binstest` commands (Cattaneo et al., 2024a,b) with default settings to generate the scatter plots.

Firm Size and Industry Composition

(a) Industry



(b) Firm Size



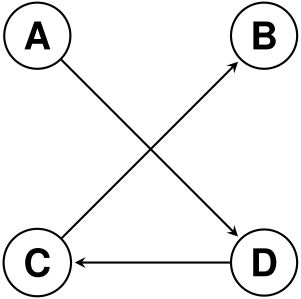
Source: Author's calculations using IMSS and INEGI data. Each panel displays a binned scatter plot of the log population and our matching metric at the commuting zone level. The vertical bars are confidence intervals for the conditional mean of the correlation at each level of (log) population. Panel (a) illustrates the relationship for service and non-service industries, respectively. We classify industries according to a 2-digit NAICS classification. Panel (b) shows the relationship separating large firms (16 or more workers) and small firms (fewer than 16 workers). The top left values display the slope of a linear regression corresponding to the displayed relationship. The regression includes dummies for each time interval. Standard errors in parentheses. To produce the scatter plots, we used the `binsreg` command (Cattaneo et al., 2024a,b) with the default settings. Clustered standard errors at the commuting-zone level in parentheses.

Concluding Remarks

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

`jorgepp@banxico.org.mx`
`www.jorgeperezperez.com`

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.

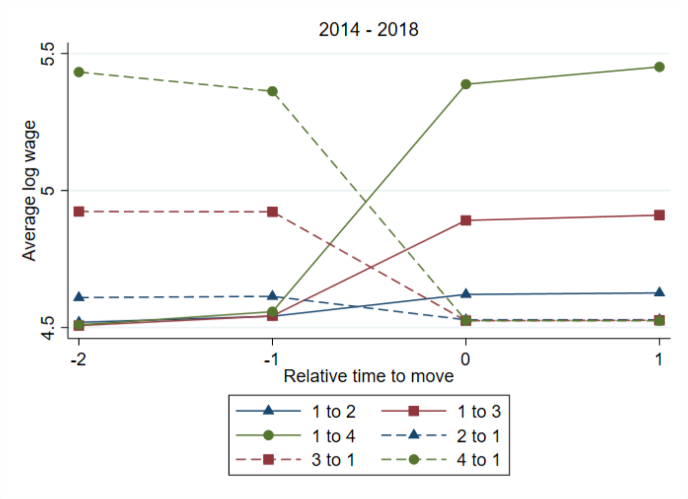
If violated:

- 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. Additional Regressions.

D: Corrected for limited mobility bias		
Log Population	0.0408***	0.0406***
	(0.004)	(0.003)
R^2	0.084	0.070
N	1,926	10,101
Mean of dep. variable	0.197	0.182
E: Dropping the 10% largest and smallest areas		
Log Population	0.0399***	0.0605***
	(0.006)	(0.005)
R^2	0.139	0.075
N	588	7,897
Mean of dep. variable	0.131	0.024

Source: Author's calculations using IMSS and INEGI data. "CZ" stands for commuting zone. The regressions pool data from 2004-2008, 2009-2013, and 2014-2018 with interval dummies. Column (2) restricts to cells with over five firms and 50 workers. Panels show the following estimates: D - Bonhomme et al. (2019)'s limited-mobility bias correction with five workplace clusters, and E - excluding extreme populations. Clustered standard errors at the commuting-zone level in parentheses. *: $p<0.1$, **: $p<0.05$, ***: $p<0.01$.