Matching and Local Labor Market Size in Mexico

Jorge Pérez Pérez* Jorge Meléndez† José G. Nuño-Ledesma‡

July 6, 2024§

Link to most recent version

Abstract

We explore how the size of local labor markets (LLM) influences the quality of matching between workers and firms in Mexico, a middle income country. Using a matched employer-employee dataset that covers nearly all formal workers in the country, we estimate econometric models of log-wages with additive worker and workplace fixed effects, which we leverage to construct a measure of assortative matching. We then correlate our metric for sorting to the population of two types of LLMs: metropolitan areas and commuting zones. We find that the impact of LLM size is relatively modest in Mexico. Doubling the size of a metropolitan area enhances assortative matching by just half of what would be expected from previous estimates for Germany. At the commuting zone level, the effect is significantly stronger but still below the German estimation. The reduced LLM-size effect is primarily associated with differences in the proportion of employment in the service industry and the preponderance of small firms between Mexico and Germany, while not strongly linked to the relative prevalence of informality and union coverage.

Keywords: Assortative Matching; City size; Local Labor Markets; Mexico; Wage determination

JEL Codes: J21, J31, O54, R23

§We thank Santiago Hermo, Alan Ker, Nicolás Sidicaro, Martin Trombetta, David Wiczer, and to seminar audiences at Michigan State University, the Urban Economics Association, ITAM, AAEA, and the Canadian Economics Society. Karla Neri, Guillermo Mondragón, René Nieto, Diego Mayorga, Marco González, and Carolina Crispín provided great research assistance. We are grateful to Jennifer Alix-Garcia for help with historical population data, and to Wolfgang Dauth for help with German local labor market data. The views and conclusions presented in this document are the exclusive responsibility of the authors and do not necessarily reflect those of Banco de México. We acknowledge financial support from Colombia Cientifica – Alianza EFI #60185 contract #FP44842- 220-2018, funded by The World Bank's Scientific Ecosystems, and managed by the Colombian Ministry of Science, Technology, and Innovation. The data were accessed through Banco de México's Econlab, which collected and processed the data as part of its effort to foster ties between Banco de México'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.

*Corresponding author. Banco de México. General Directorate of Economic Research. 18 Av. 5 de Mayo. Colonia Centro, Mexico City, México 06000. Email: jorgepp@banxico.org.mx.

†Banco de México. General Directorate of Economic Research. 18 Av. 5 de Mayo. Colonia Centro, Mexico City, México 06000. Email: jorge.melendez@banxico.org.mx.

‡University of Guelph. Department of Food, Agricultural and Resource Economics. J.D. MacLachlan Building 50 Stone Road East Guelph, Ontario, Canada N1G 2W1. Email: jnuno@uoguelph.ca.

1 Introduction

Wage disparities across local labor markets (LLM) within a country have a greater impact on overall earnings inequality than the differences in remunerations among workers residing in the same LLM (Baum-Snow and Pavan, 2012; Combes and Gobillon, 2015; Gould, 2007). One of the most important contributors to wage gaps across localities is the easier worker-workplace matching that takes place in larger LLMs compared to smaller cities (D'Costa and Overman, 2014; Dauth et al., 2022; Duranton and Puga, 2004). Intuitively, the better matching process fuels a virtuous cycle wherein better matches enhance productivity, which raises salaries, attracting better candidates to the area, adding to local population (Ahlfeldt and Pietrostefani, 2019). Thus, a complete understanding of any country's labor market dynamics must include a careful analysis of the way enhanced matching due to city-size agglomeration externalities contributes overall wage dynamics. We study the relationship between the size of local labor markets and assortative in Mexico.

Most of the literature on matching in local markets examines developed economies (Dauth et al., 2022). We study the link between LLM size and the quality of worker-workplace matching in Mexico, a middle-income country with high levels of economic informality and significant regional variation. There is a lack of research on how the worker-workplace matching process functions in developing economies, where labor markets are afflicted with structural issues which can hinder the pairing process. We expect the LLM-size-matching link to be different in Mexico compared to developed economies because some features of its labor market dynamics have been documented to be surprisingly distinct. For example, workplace-level wage determinants are significantly more important in explaining wage differentials in Mexico and their relevance has only increased in recent years (Pérez Pérez and Nuño-Ledesma, 2022), an empirical regularity shared with other developing economies (Bassier, 2023; Diallo et al., 2022; Frías et al., 2022).

The literature on agglomeration in developing countries reveals that while wage elasticities with respect to urban density are comparable to those observed in developed economies, these gains are often insufficient to outweigh dis-utilities such as pollution and crime common in developing countries Grover et al. (2023).

To evaluate the relationship between LLM-size and worker-workplace sorting, we first construct a metric of assortative matching. To this end, we leverage data from social security records comprising the near-universe of formal workers in Mexico from 2004 to 2018. These data have recently been used, among other goals, to survey recent dynamics in Mexico's income inequality, to investigate the relationship between international trade and wage premia in Mexico, and to decompose total Mexican wage dispersion into worker and workplace components (Puggioni et al., 2022; Frías et al., 2022; Pérez Pérez and Nuño-Ledesma, 2022). With these records, we estimate models of log-wages with additive worker and workplace fixed effects using the approach first popularized by Abowd et al. (1999). To proxy assortative matching, we use the covariance of worker and workplace fixed effects at the metropolitan area and commuting zone levels.

We proceed by taking our measure of assortative matching and correlating it with population at the metropolitan area and commuting zone levels. To provide context to our estimations, we contrast them against equivalent metrics obtained from research studying labor markets in Germany. We find that the association between LLM-size and the degree of positive assortative matching in formal labor markets is weaker in Mexico. For example, at the metropolitan area level, we estimate the average covariance between worker and workplace fixed effects in Mexico at the city level hovers around 0.021, compared to 0.061 in Germany (Dauth et al., 2022). At the commuting zone level, the slope of the relationship between log city size and the correlation coefficient between workplace and worker fixed effect estimates is 0.068, much closer to the the slope of 0.069 estimated by Dauth et al. (2022). Estimates controlling for additional fixed effects show generally weaker slopes, implying that agglomeration forces due to better matching in labor markets are weaker in Mexico.

When we turn to the reasons behind this difference, we find that the size of the informal labor market is negatively correlated with the strength of assortative matching in the formal labor market. We also find that unions have a negligible effect on the relationship between city size and matching. Nevertheless, our estimates suggest that neither of these variables can account for the diminished strength of matching externalities in Mexico. Instead, we argue that the industry composition in Mexico and the preponderance of small firms reduce matching externalities.

We contribute to three strands of literature. First, we provide an additional example to the set of studies documenting the presence of city-size wage gaps not only in developed countries as discussed by Baum-Snow and Pavan (2012); Gould (2007); D'Costa and Overman (2014), and De la Roca, Jorge and Puga, Diego (2016), but also in developing economies (Chauvin et al., 2017; Combes et al., 2020; De la Roca et al., 2023; Duranton, 2016). Our estimates for Mexico, a middle-income country with high labor informality, validate these previous estimates. Second, we contribute to the literature on matching in labor markets and agglomeration forces (Andersson et al., 2007; Baum-Snow and Pavan, 2012; Behrens et al., 2014; Dauth et al., 2022), by showing that agglomeration forces are somewhat weaker in Mexico and that labor informality reduces positive assortative matching in formal labor markets. In doing so, we also contribute to the literature on informal labor markets Ulyssea (2018) and their relationship with formal ones (Levy Algazi, 2018; Ulyssea, 2010). Lastly, we contribute to the small but growing set of studies, such as Frías et al. (2022) and Pérez Pérez and Nuño-Ledesma (2022), that adapt multi-dimensional fixed effects models to the Mexican case.

The rest of the paper proceeds as follows. Section 2 details the data and the Mexican context. Section 3 details our AKM model estimation. Section 4 estimates the relationship between assortative matching and city and local-labor-market-level covariates and estimates the relationship between matching, city size, and informality in Mexico. Section 6 concludes.

2 Data

Our data consists of social security records from the Mexican Social Security Institute (IMSS by its acronym in Spanish). The Institute is the Mexican government's main social security, pension, and public health administrator. Salaried workers in the private sector are required to register with IMSS by law. By the government's estimation, 83% of all formal workers are affiliated with IMSS (Pérez Pérez and Nuño-Ledesma, 2022). The records at our disposal report monthly observations from November 2004 to December 2018. This last month of the data contained information for about 20.1 million workers.[1] Notwithstanding its wide coverage, our data have some limitations. A different agency manages records for public employees, so they do not appear in our data. Daily earnings above 25 "units of measurement and update" (approximately 140 USD) are censored. Naturally, information from informal workers (persons working in the shadow economy) is absent from IMSS data. Informality is widespread in Mexico; according to the country's National Survey of Occupation and Employment (ENOE by its acronym in Spanish), 55% of all workers operate in

[1] See online appendix A for details on variable construction and descriptive statistics.

the informal economy Banco de México (2023). When estimating the AKM models, the dependent variable of interest is the natural logarithm of real daily taxable income.[2] Taxable income may include remunerations made to the worker other than wage. The dataset also includes information on gender, age, and registration date to IMSS. The dataset does not include information on education or hours worked. Our primary sample of interest is prime-age men (25-54 years old) who have likely concluded their education. Thus, their estimated worker effect should include wage variance attributable to education. Regarding employer information, the data includes workplace and economic sector identifiers.

We complement IMSS data with city-level characteristics from Mexican censuses and intercensal surveys for 2005, 2010, and 2020. We construct city-level labor informality rates using data from Mexico's labor survey, ENOE. Regarding local-labor-market-level characteristics, we rely on the definition of local labor markets (and the associated dataset) used by Aldeco et al. (2023), which divides Mexico into 777 local labor markets.[3]

3 Constructing a Measure of Assortative Matching

Our empirical approach is to first build a measure of assortative matching to then correlate it to the size of local labor markets. Our measure of sorting is the covariance of worker and workplace fixed effects retrieved from models of log-wages following the methodology introduced by Abowd et al. (1999) and popularized by works like Card et al. (2013) and Card et al. (2018) (frequently referred to as "AKM models"). In AKM models, log wages are modeled as a function of additive worker and workplace fixed effects:

$$\ln(wage_{it}) = \alpha_i + \psi_{J(i,t)} + X_{it}'\beta + r_{it}. \tag{1}$$

Here, $wage_{it}$ is the real wage of worker i at time t. The vector of fixed effects $\alpha_\mathbf{i}$ captures the influence of all time-invariant worker characteristics. Similarly, the vector of fixed effects $\psi_{J(i,t)}$ collects time-invariant factors at the workplace level for workplace J where worker i was employed

[2] As Dauth et al. (2022) note, using nominal or real wages only changes the scale of the firms' fixed effects.

[3] Aldeco et al. (2023) calculate commuting zones using the methodology in Fowler and Jensen (2020). B in the Appendix provides detail about the commuting zone construction.

at time t. The vector X_{it} includes control variables, which in our estimations include functions of age and time-interval trends. We estimate equation (1) by OLS with a preconditioned iterative gradient method (Card et al., 2013).[4] We generate estimates of the model for three discrete time segments: 2004-2008, 2009-2013, and 2014-2018. In AKM models, worker mobility identifies firm and worker effects. As pointed out by Abowd et al. (1999), these effects can be disentangled by worker mobility –generated when workers change employers– that creates a network of directly or indirectly connected workplaces. We restrict estimation to the largest "connected set" of workplaces in each time interval across which workers change jobs at least once (Abowd et al., 1999).[5]

We show summary statistics of our estimations in Table 1. The AKM models explain about 94% of the variance of log wages each period. The variance of worker effects decreased over time, while the variance of worker effects and the covariance between worker and workplace effects increased over time. The correlation coefficient between worker and workplace effects is between 0.21 and 0.25, much lower than that found for Germany by Dauth et al. (2022) of 0.64 in 2008-2014.

We follow Card et al. (2013) and use our estimated model to decompose the variance of wages into the shares attributed to each component:

$$\begin{aligned}\mathrm{Var}\left(\ln \mathrm{wage}_{it}\right) = &\underbrace{\mathrm{Var}(\alpha_i)}_{\text{workers}} + \underbrace{\mathrm{Var}(\psi_{\mathbf{J}(i,t)})}_{\text{workplaces}} + \mathrm{Var}(x_{it}'\beta) + \mathrm{Var}(r_{it}) \\ &+ 2\underbrace{\mathrm{Cov}(\alpha_i, \psi_{\mathbf{J}(i,t)})}_{\text{Assortative Matching}} + 2\mathrm{Cov}(\psi_{\mathbf{J}(i,t)}, x_{it}'\beta) + 2\mathrm{Cov}(\alpha_i, x_{it}'\beta).\end{aligned} \tag{2}$$

The last rows of Table 1 show the results of this decomposition. Worker effects explained about 42% of wage variance in 2004-2008, and their contribution to total variance has decreased over time. In contrast, the contribution of workplace effects and sorting has been increasing over time. Sorting explains 15 % of wage variance in 2004-2008, whereas it explains about 19% of wage

[4] We provide validation exercises for the linearity in equation (1) and the uncorrelatedness of the error term r_{it} and the fixed effects and covariates in online appendix section C. Specifically, we follow Card et al. (2013) and show that firms are exchangeable: for a worker, moving from firm A to firm B has approximately the same effect on wages as moving from firm B to firm A. Pérez Pérez and Nuño-Ledesma (2022) also provide evidence of the validity of AKM models in this sample.

[5] Online appendix table A.1 shows that descriptive statistics are similar in the full and connected set samples.

Table 1: 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
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.

variance in 2014-2018. This pattern suggests that assortative matching is increasingly important in determining wage variance in Mexico over time.[6]

4 Assortative Matching and Labor Market Size

We proceed to analyze how the strength of positive assortative matching varies with city size across Mexican cities. To this end, we regress our measure of assortative matching against the logarithm of the population at the Local Labor Market level.[7] Recall that our measure of matching is the correlation coefficient between worker and workplace fixed effects retrieved from the AKM models estimated above. With respect to LLM population, we conduct our analysis at two geographical aggregations: metropolitan areas (INEGI, 2018) and commuting zones as defined by Aldeco et al. (2023). Metropolitan areas only include the largest cities in the country, while commuting zones are a more refined category, including population centers in addition to the largest cities. To delimit the analysis to specific labor markets further, We also conduct our analysis at the metropolitan area-industry and commuting zone-industry levels.[8]

When conducting our econometric analysis, we pool the estimated correlation coefficients between worker and workplace fixed effects in a single AKM model and include binary variables indicating the period accounted for in the regression.[9]

We show the results of our estimations in Table 2. Panel A, column 1 shows that the correlation between labor market size and assortative matching for Mexico's metropolitan areas is positive and statistically significant at both the metropolitan area and commuting zone levels. At the metropolitan area (columns 1 and 2), the link between matching and labor market size is much weaker than

[6]A pervasive issue in these decompositions is the bias in the plug-in OLS estimates of the variance components of equation (2). Kline et al. (2020) show that even if OLS estimates of the fixed effects are unbiased, estimates of their variance may be biased, usually attenuating estimates of the covariance between worker and workplace fixed effects. Bonhomme et al. (2019) show that the bias in these variance components worsens when there is limited mobility of workers across workplaces. To address these issues, in Appendix section D, we repeat the decomposition using alternative estimators for the variance components of equation 2. The results of these decomposition exercises with limited mobility bias corrections are similar to the baseline results.

[7]In Appendix Table E.1, we show that we obtain similar results using log-employment as a proxy variable of labor market size.

[8]We use a three-digit NAICS industry classification. Our data includes 22 industries. For the estimates at the commuting zone by industry level, we exclude cells with fewer than 50 workers or fewer than five firms.

[9]Appendix table E.2 shows that we obtain similar results if we estimate these regressions for each time interval separately.

that found in Dauth et al. (2022) for Germany (0.061). In fact, doubling the metropolitan area size increases the correlation coefficient between worker and workplace fixed effects by 2.0 p.p. This correlation is weak even for the country's largest urbanized areas, indicating weak agglomeration externalities in metropolitan areas.

Columns 3, and 4 of panel A show estimates relying on commuting zones as the definition of LLM.[10] At this geographical level, the estimated association between population and matching is significantly stronger than the estimated for metropolitan areas. Doubling the population size of a commuting zone or a commuting zone-industry cell increases assortative matching by 6.8 and 5.3 p.p., correspondingly which is still half of the association found by Dauth et al. (2022). Figure 2 shows the relationship we obtained and compares each to the comparable German estimate reported by Dauth et al. (2022).

In the remaining panels B to D Table 2, we show estimates from alternative specifications, which we conducted to gauge the robustness of our findings. Panel B demeans the estimated workplace fixed effects by industry. By doing this, the estimated correlations between city size and matching control for different industry compositions across cities. Using these "residual" fixed effects, the association between city size and matching is weaker than in previous panels. This pattern of results indicates that part of the effect seen in the estimates with unadjusted fixed effects was due to industries with better labor market matching located in larger cities.

Panel C instruments current population with population in 1950 to avoid reverse causality between labor market matching and population growth. Compared to Panel A, when we leverage this instrumental variable, we obtain larger estimates at the metropolitan area, but weaker at the commuting zone level.

In panel D we report results from an econometric model adjusting for limited mobility bias as proposed by Bonhomme et al. (2019). For these estimates, we cluster firms into five clusters using 20 percentiles of the within-firm distribution of wages as clustering variables. Then, we re-estimate the AKM model with this reduced workplace effect number of parameters and recalculate the correlations between worker and workplace effects. With this correction, the estimated correlations are smaller, suggesting that limited mobility bias is not dampening our estimates of the strength of

[10] For the commuting zone by industry estimates, we restrict to cells with more than five firms and more than 50 workers, following Dauth et al. (2022).

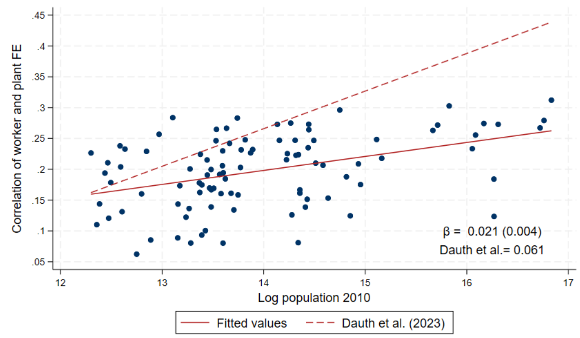
Table 2: City Size and Assortative Matching in Mexico's Formal Labor Markets

Dependent variable: correlation of worker and plant FE				
	(1) Metro	(2) Metro-Industry	(3) CZ	(4) CZ-Industry
A: Baseline Model				
Log Population	0.0205***	0.0202***	0.0679***	0.0531***
	(0.004)	(0.004)	(0.004)	(0.002)
R^2	0.358	0.026	0.155	0.090
B: Correlation of worker and residual workplace FE				
Log Population	0.0184***	0.0202***	0.0035	0.0030***
	(0.005)	(0.004)	(0.000)	(0.000)
R^2	0.375	0.026	0.131	0.074
C: Log population instrumented with population in 1950				
Log Population	0.0322***	0.0112	0.0577***	0.0493***
	(0.007)	(0.011)	(0.008)	(0.010)
R^2	0.312	0.024	0.152	0.067
First-stage F	21.468	7.235	621.692	34.729
N for Panels A-C	97	1,648	1,961	10,118
D: Corrected for limited mobility bias				
Log Population	0.0173***	0.0037	0.0447***	0.0406***
	(0.004)	(0.004)	(0.003)	(0.001)
R^2	0.245	0.011	0.113	0.070
N	97	1,647	1,776	10,101
E: Dropping the 10% largest and 10% smallest areas				
Log Population	0.0195***	0.0243***	0.0399***	0.0605***
	(0.006)	(0.005)	(0.004)	(0.002)
R^2	0.366	0.053	0.139	0.075
N	58	986	588	7,897

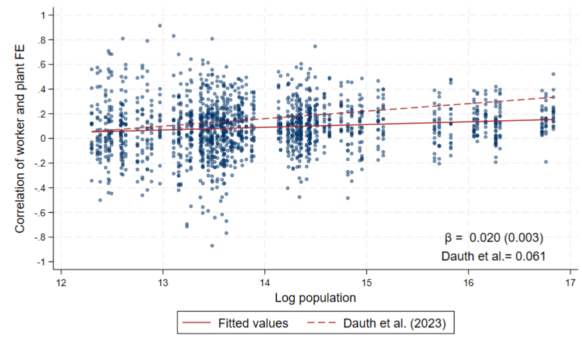
Source: Author's calculations using IMSS and census data. The columns show the results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log population at different geographical aggregation levels. All the regressions pool data from the three intervals: 2004-2008, 2009-2013, and 2014-2018, and include dummies by interval. Metro stands for metropolitan area. CZ stands for commuting zone. 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 1950 as reported by Alix-Garcia and Sellars (2020).. First-stage F is the first stage F-statistic. 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 and smallest areas by population. Robust standard errors in parentheses. *: $p<0.1$, **: $p<0.05$, ***: $p<0.01$.

Figure 1: City Size and Assortative Matching in Mexico's Formal Labor Markets

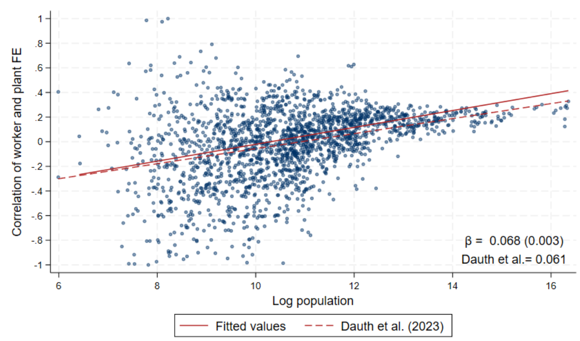
(a) Metropolitan Area



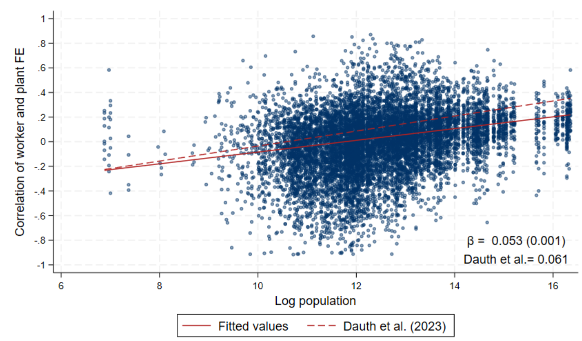
(b) Metropolitan Area - Industry



(c) Commuting Zone



(d) Commuting Zone - Industry



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 displays a dashed line with the implied relationship if the coefficient for the relationship were the same as that estimated 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.

matching externalities.

Last, panel E removes the largest and smallest cities from the sample. The results are similar in this restricted sample, implying the small association we find is not due to the smallest cities or Mexico City's influence.

Overall, at the metropolitan area, the city-size advantage for matching in the labor market is weaker in Mexico compared to the results of Dauth et al. (2022). The estimated relationship is closer, albeit still weaker, when looking at commuting zones. We now explore possible explanations for this difference.

5 Why are Matching Agglomeration Externalities Weak in Mexico?

There are many differences between labor markets in developed and less advanced economies which may debilitate the agglomeration benefits of larger local labor markets when it comes to worker-workplace matching. We concentrate on two of the main differences between the labor markets of Mexico and Germany, which has been the object of previous analyses similar in spirit to ours.

First, labor market informality is high in Mexico. According to Mexico's ENOE, 55% of employed workers labor in the informal economy as of 2022. Informal workers may not pay income taxes, lack labor stability, and do not have access to social security through their employers. The presence of a large informal labor market may depress assortative matching processes and city-size wage advantages for multiple reasons. From the labor supply side, informal labor markets may attract workers who would otherwise join the formal sector and match with a firm there. Although informal workers are less productive in general (Levy Algazi, 2018), if they were to join low-productivity formal firms instead of informal firms, this would enhance assortative matching compared to a scenario where low-productivity informal firms are either hiring productive workers or not hiring.[11] From the labor demand side, informality may dampen formal firms' productivity

[11] It is also possible that informality strengthens matching in the formal sector because of its role as a stepping stone in the labor market. Many workers in Mexico enter the labor force through the informal market and only obtain formal jobs later in life. If these workers are better informed about their abilities at this stage, they may search for firms where they are a better match. However, since our results point towards reduced matching externalities in Mexico,

because of competition. It may also encourage formal firms to have some informal workers, who may be less productive but charge lower wages and may be poor matches for the firm.

Another substantial difference between Mexico's and Germany's labor markets is the limited power of unions in Mexico. In 2018, the collective bargaining coverage (percentage of employees with a legal right to bargain) in Mexico and Germany was 10% and 54%, correspondingly (OECD, 2023). Card et al. (2013) argue that the declining power of unions in Germany increased the importance of assortative matching in determining wage inequality. However, if this were the case for Mexico, we would expect that Mexico's labor markets would have stronger assortative matching.

We test the relationship between informality, unionization, and matching in formal labor markets in columns 1 to 4 of Table 3. There, we regress assortative matching on city size as in section 4, but we also control for the labor market informality rate and the unionization rate. We do so for different geographical aggregations. At the metropolitan-area-industry level, we find that the informality rate has a negative effect on formal labor market assortative matching. A 10 % increase in the informality rate reduces the correlation between worker and firm effects in 3.49 p.p. In turn, increasing the unionization rate does not affect the correlation. Further, controlling for these variables reduces the estimated size of the city-size-matching relationship. At the commuting zone level, where we lack unionization information, we still find a negative relationship between informality and matching.

We follow a more agnostic approach in column 5 of Table 3. We take advantage of having repeated estimates for three different time intervals and estimate regressions with commuting-zone fixed effects (in addition to dummies for each interval). This way, we are comparing the evolution of assortative matching in commuting zones where the population grew faster to commuting zones where it grew slower instead of comparing commuting zones in the cross-section. This comparison renders the coefficient on informality insignificant, but the population coefficient remains statistically significant. In columns 6 to 8 at the commuting-zone-industry level, controlling for additional fixed effects renders the relationship between city size and matching insignificant. This finding contrasts with the results of Dauth et al. (2022), which find that the relationship between

this stepping-stone role must no be the primary mechanism through for informality affecting formal labor market matching.

Table 3: City Size and Assortative Matching in Mexico's Formal Labor Markets: Additional Controls

Dependent Variable: Correlation of worker and workplace FE

	(1) City-Industry	(2) City-Industry	(3) CZ	(4) CZ	(5) CZ	(7) CZ-Industry	(8) CZ-Industry	CZ-Industry
Log population	0.0202***	0.0178***	0.0679***	0.0394***	0.1996**	0.0531***	-0.0232	-0.0238
	(0.004)	(0.004)	(0.004)	(0.005)	(0.088)	(0.002)	(0.044)	(0.029)
Informality Rate		-0.3493***		-0.4303***	0.0360			
		(0.025)		(0.036)	(0.201)			
Unionization Rate		-0.0278						
		(0.032)						
CZ FE					Yes		Yes	
CZ-Industry FE								Yes
N	1,648	1,647	1,961	1,961	1,910	10,118	10,118	9,746
R^2	0.026	0.182	0.155	0.197	0.662	0.090	0.206	0.792

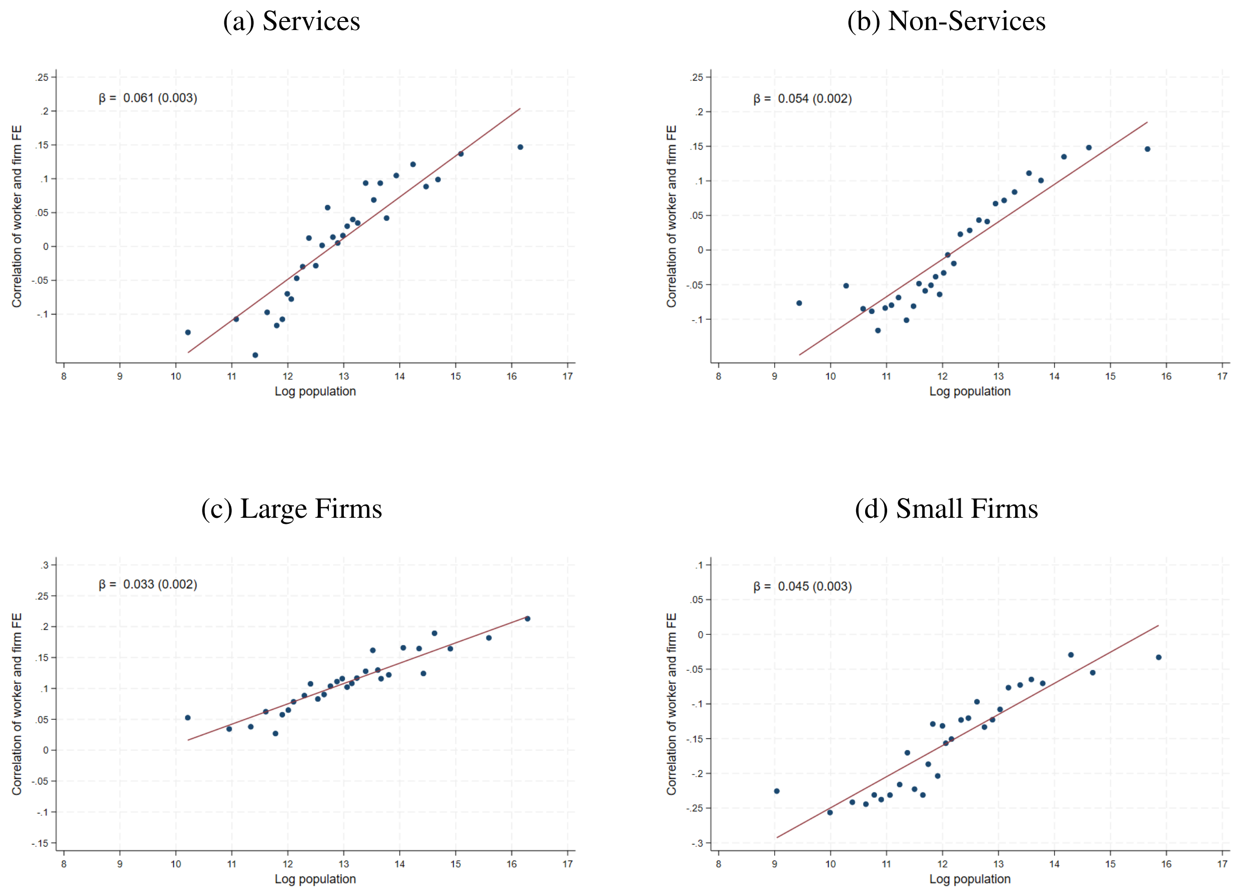
Source: Author's calculations using IMSS, ENOE, and census data. The columns show the results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log population at different geographical aggregation levels and with different specifications. Metro stands for metropolitan area. CZ stands for commuting zone. 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.

city size and matching holds after controlling for additional fixed effects. These findings highlight that agglomeration externalities in Mexico's formal labor markets are weaker than what was found by Dauth et al. (2022).

Since Mexico's labor market features are not behind the weaker matching externalities, we now explore the role of other features of the Mexican economy that are not specific to the labor market but affect the economy as a whole: industry composition and firm size. Compared to Germany, where 71% of employment is in services (Bank, 2023), Mexico has a larger share of manufacturing employment and employment in other sectors. If matching externalities were stronger in the service industry, then part of the reason behind the weaker matching could be the lower share of service jobs. Figure 2, panels A and B show binned scatterplots displaying the relationship between city size and labor market matching at the commuting zone level for service and non-service sectors. The relationship between city size and matching in the services industries is steeper.

We estimate regressions of population and matching by sector at the commuting-zone level and then obtain weighted averages of the slopes separating industries into services and others. For

Figure 2: City Size and Assortative Matching in Mexico's Formal Labor Markets: Industry and Firm Size Differences



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 cells with more than five firms and more than 50 workers. Panels a and b show the relationship for service and non-service industries, respectively. We classify industries according to a 3-digit NAICS classification. Panels c and d show the relationship separating large firms (more than 16 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.

services, the weighted average slope is 0.061. In contrast, the average slope for non-service sectors is 0.054. These findings suggest that agglomeration forces are stronger in the services sectors. Still, these do not explain the shortfall in these slopes compared to Dauth et al. (2022).

Another endemic feature of Mexico's economy is a firm-size distribution overtly concentrated on small firms compared to developed countries and countries with similar development levels (Levy Algazi, 2018). Previous evidence shows that agglomeration effects are more substantial for larger firms and that agglomeration positively affects firm size (Li et al., 2012). Eeckhout and Kircher (2018) show that in a labor market matching model, firm size and worker quality sorting are positively correlate under some complementarity conditions.

Figure 2, panels c and d, shows the relationship between city size and assortative matching at the commuting zone level, separating the sample into large firms (more than 16 workers) and small firms (fewer than 16 workers), following the firm size classification of Levy Algazi (2018). The differences between these relationships for small and large firms are striking. For small firms, the correlation between log population and assortative matching is negative for a large share of commuting zones, and the increase in this correlation as the population grows is moderate. For large firms, the picture is quite different. Assortative matching is positive throughout commuting zones. These findings validate the hypothesis that labor market matching externalities are at play mainly for large firms.

The evidence in Figure 2 suggests that industrial composition and firm size account for the stronger matching externalities found in Dauth et al. (2022) compared to those for Mexico.

6 Concluding Remarks

Using administrative Mexican social security records, we provide a descriptive analysis of the underlying mechanisms of wage variation and worker-firm sorting patterns in the country. We show that city size is positively correlated with the intensity of positive assortative matching. However, the correlation is weaker than estimations previously reported by similar exercises conducted in Germany. The results persist even when we consider limited worker mobility, different industry compositions across cities, and a potential reverse causality between labor market matching and population growth. Our findings suggest that agglomeration externalities, which often amplify positive assortative matching in larger cities, are a frail force in Mexico's labor market. The softer association between city size and matching points to differences in the underlying labor market structures and economic conditions unique to economies with large informal sectors, a small share of workers in service industries, and weak collective bargaining coverage.

Declaration of Generative AI and AI-assisted technologies in the writing process

Statement: During the preparation of this work the authors used ChatGPT 3.5 and ChatGPT 4.0 in order to check spelling, improve clarity and format BibTeX entries. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

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

Ahlfeldt, G. M. and Pietrostefani, E. (2019). The economic effects of density: A synthesis. *Journal of Urban Economics*, 111:93–107.

Aldeco, L., Calderón, M., Chiquiar, D., Hanson, G., Pérez Pérez, J., and Velásquez, C. (2023). Mercados de trabajo locales en México: Definición, bases de datos y análisis descriptivo 1990-2020. Technical report, Banco de México.

Alix-Garcia, J. and Sellars, E. A. (2020). Locational Fundamentals, Trade, and the Changing Urban Landscape of Mexico. *Journal of Urban Economics*, 116:103213.

Andersson, F., Burgess, S., and Lane, J. I. (2007). Cities, Matching and the Productivity Gains of Agglomeration. *Journal of Urban Economics*, 61:112–128.

Banco de México (2023). Informe Trimestral, Julio-Septiembre 2023.

Bank, W. (2023). Employment in services (% of total employment). Accessed: 2023-06-10.

Bassier, I. (2023). Firms and Inequality When Unemployment is High. *Journal of Development Economics*, 161:103029.

Baum-Snow, N. and Pavan, R. (2012). Understanding the City Size Wage Gap. *The Review of Economic Studies*, 79(1):88–127.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy*, 122(3):507–553.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, 87(3):699–739.

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

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). What is Different About Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98:17–49.

Combes, P.-P., Démurger, S., Li, S., and Wang, J. (2020). Unequal Migration and Urbanisation Gains in China. *Journal of Development Economics*, 142:102328.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of regional and urban economics*, volume 5, pages 247–348. Elsevier.

Dauth, W., Findeisen, S., Moretti, E., and Suedekum, J. (2022). Matching in Cities. *Journal of the European Economic Association*, 20(4):1478–1521.

D'Costa, S. and Overman, H. G. (2014). The Urban Wage Growth Premium: Sorting or Learning? *Regional Science and Urban Economics*, 48:168–179.

De la Roca, J., Parkhomenko, A., and Velásquez-Cabrera, D. (2023). Skill Allocation and Urban Amenities in the Developing World. Working paper.

De la Roca, Jorge and Puga, Diego (2016). Learning by Working in Big Cities. *The Review of Economic Studies*, 84(1):106–142.

Diallo, Y., Sarr, I., and Diagne, I. (2022). Role of Firms in Wage Dispersion: Evidence from a Developing Country. *PEDL Research Papers*.

Duranton, G. (2016). Agglomeration Effects in Colombia. *Journal of Regional Science*, 56(2):210–238.

Duranton, G. and Puga, D. (2004). Micro-foundations of Urban Agglomeration Economies. In *Handbook of regional and urban economics*, volume 4, pages 2063–2117. Elsevier.

Eeckhout, J. and Kircher, P. (2018). Assortative Matching with Large Firms. *Econometrica*, 86(1):85–132.

Fowler, C. S. and Jensen, L. (2020). Bridging the Gap Between Geographic Concept and the Data We Have: The Case of Labor Markets in the USA. *Environment and Planning A: Economy and Space*, 52(7):1395–1414.

Frías, J. A., Kaplan, D. S., Verhoogen, E., and Alfaro-Serrano, D. (2022). Exports and Wage Premia: Evidence from Mexican Employer-Employee Data. *The Review of Economics and Statistics*, pages 1–45.

Gould, E. D. (2007). Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium. *The Review of Economic Studies*, 74(2):477–506.

Grover, A., Lall, S., and Timmis, J. (2023). Agglomeration economies in developing countries: A meta-analysis. *Regional Science and Urban Economics*, 101:103901.

INEGI (2018). Delimitación de las zonas metropolitanas de México 2015.

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

Levy Algazi, S. (2018). *Under-rewarded Efforts: The Elusive Quest for Prosperity in Mexico*. Inter-American Development Bank.

Li, D., Lu, Y., and Wu, M. (2012). Industrial Agglomeration and Firm Size: Evidence from China. *Regional Science and Urban Economics*, 42(1-2):135–143.

OECD (2023). Collective bargaining coverage. Accessed: 2023-06-10.

Pérez Pérez, J. and Nuño-Ledesma, J. (2022). The Contribution of Workers, Workplaces, and Sorting to Wage Inequality in Mexico. Documento de Trabajo WP 2022-012, Alianza EFI.

Puggioni, D., Calderón, M., Cebreros Zurita, A., Fernández Bujanda, L., Inguanzo González, J. A., and Jaume, D. (2022). Inequality, Income Dynamics, and Worker Transitions: The case of Mexico. *Quantitative Economics*, 13(4):1669–1705.

Ulyssea, G. (2010). Regulation of Entry, Labor Market Institutions and the Informal Sector. *Journal of Development Economics*, 91(1):87–99.

Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8):2015–2047.

Online Appendix - Not for Publication

A Data and Descriptive Statistics

We use IMSS data previously analyzed in Pérez Pérez and Nuño-Ledesma (2022). IMSS refers to the Instituto Mexicano del Seguro Social, a Mexican government organization that handles public health, pension management, and social security. Private sector employees are required by law to register with IMSS. According to Mexico's labor survey, about 83% of the formal workforce in 2022 were registered with IMSS. Self-employed individuals have the option to register with IMSS, granting them access to certain aspects of the social security system. Typically, self-employed workers register with the equivalent of one legal minimum salary. Self-employed records make up approximately 0.1% of the overall IMSS database. In cases where a worker reports multiple jobs, we retain the job with the highest reported wage. Only 2.5% of workers reported having more than one job in December 2018.

The IMSS social security information is available on a monthly basis. We analyze records spanning from November 2004 to December 2018. Our analysis concludes in 2018 due to significant changes in Mexico's labor market resulting from the COVID-19 pandemic and substantial increases in the minimum wage during the period from 2019 to 2022. The database initially consists of 12.8 million workers in November 2004 and grows to 20.1 million workers by December 2018. Our main variable of interest is the daily taxable income, which encompasses various forms of compensation, excluding additional non-taxable payments such as paid vacations and bonuses. Wages exceeding 25 UMAs (units of measure and update) are capped. In 2018, this cap stood at 2,015 MXN daily, approximately 102 USD.

IMSS identifies workplaces through an employer registry number, known as "registro patronal," The registry number pertains to an employer rather than a physical location. This means that workers who operate in the same plant can have multiple employers, each identified by their registro patronal. For the purposes of our analysis, we utilize anonymized identifiers for the registro patronal, which prevents the precise identification of individual workers or firms within the dataset. The dataset we received from the Mexican Central Bank's EconLab already contains these masked identifiers, and this anonymization process does not affect our econometric analysis or its results. Our study does not explicitly report plant effects as estimated in previous research that utilizes the AKM methodology.

Table A.1, reproduced from Pérez Pérez and Nuño-Ledesma (2022), displays a summary of the IMSS data. We divide our data into three time intervals: 2004-2008, 2009-2013, and 2014-2018. Within each year, our sample includes a substantial number of wage observations, ranging from 73 to 113 million for men aged 25-54. In column (2) of the table, we observe a 0.7% decrease in the average real daily wage for prime-age men between 2009 and 2014 when compared to 2005. However, this decline is followed by a 1.5% increase by 2018. Column (3) illustrates a widening gap in earned wages between 2005 and 2018.

Table A.1: Descriptive Statistics: Prime-age Men, 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

Source: Authors' calculations using IMSS data. Observations correspond to the sum of all the monthly observations in a year. Real wages are daily taxable income registered in IMSS, expressed in real terms using prices from July 2018. Percent censored is the percentage of observations with wages exactly equal to the upper wage limit of 25 minimum wages or units of measure and update per year.

Abowd et al. (1999) show that AKM estimates identify fixed effects for workers and workplaces within a "connected set" of workplaces where there is a shared pool of workers who switch jobs at least once. Our estimates use the largest connected set within each time interval. A workplace is part of the connected set if at least one of its workers has worked or will work in a different workplace during the given time interval. Direct connections between every pair of workplaces are not necessary for a connected set to exist.

Table A.2, reproduced from Pérez Pérez and Nuño-Ledesma (2022), shows the number of worker-month observations for prime-age men that had more than one job, the number of individuals, and the average and standard deviation of log wages. In each interval, our database consists of 158 to 297 million worker-month observations representing 5 to 9 million individuals. The standard deviation of salaries slightly increased from 0.77 in the 2004-2008 interval to 0.79 in the 2014-2018 interval. Average real wages exhibited a consistent upward trend throughout the sample. Columns (5) to (8) of Table A.2 display the corresponding descriptive statistics for the largest connected set of prime-age male workers. This largest connected set encompasses at least 94% of all worker-year observations and 97% of all individuals within a given interval. Average wages within the connected set are slightly higher compared to the overall sample, while standard deviations are marginally smaller. Given the substantial size of the connected set relative to the entire sample, the similar mean salaries, standard deviations, and comparable trends in average wages and salary dispersion, our focus on this connected group does not involve a significant loss of detail.

Table A.2: Descriptive Statistics - Overall Sample and Workers in the Largest Connected Set

	All sample				Individuals in largest connected set			
			Log wage				Log wage	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Interval	All obs.	Individuals	Mean	Std. dev.	All obs.	Individuals	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

Source: Authors' calculations using IMSS microdata. Statistics for men 25 to 54 years old who had more than one job, i.e. were employed in more than one workplace. Log wage is the log of daily taxable income registered in IMSS, expressed in real terms using prices from July 2018. "Ratio: largest connected/all" is the ratio of the corresponding statistic in the largest connected set to its counterpart in the full sample.

We complement the IMSS data with city-level and commuting-zone level covariates. To calculate informality and unionization rates, we use the Occupation and Employment Survey of the National Institute of Statistics and Geography (INEGI). The informality rate is the percentage of workers who do not have social security benefits over the total number of workers. The unionization rate is the percentage of workers who state they belong to a union over the total number of formal workers.We calculate quarterly rates from 2004 to 2018 for the 43 biggest Mexican cities and then averages at city levels for estimation periods (2004-2008, 2009- 2013, 2014-2018). We also calculate the informality rate at the commuting-zone level, using the 2005, 2015, and 2020 Population and Housing Censuses to calculate informality at this geographic level.

B Details about Commuting Zone Construction

We use commuting zones for Mexico calculated by Aldeco et al. (2023). We provide a summary of their methodology to build commuting zones here. Using residence-workplace data from the Mexican census of 2010, Aldeco et al. (2023) group municipalities according to their similarity in commuting patterns, using the same methodology as Fowler and Jensen (2020) for the US. First, they calculate a commuting dissimilarity index for each pair of municipalities i, j:

$$D_{ij} = 1 - \frac{f_{ij} + f_{ji}}{\min\left(\sum_l f_{il}, \sum_l f_{lj}\right)} \tag{B.1}$$

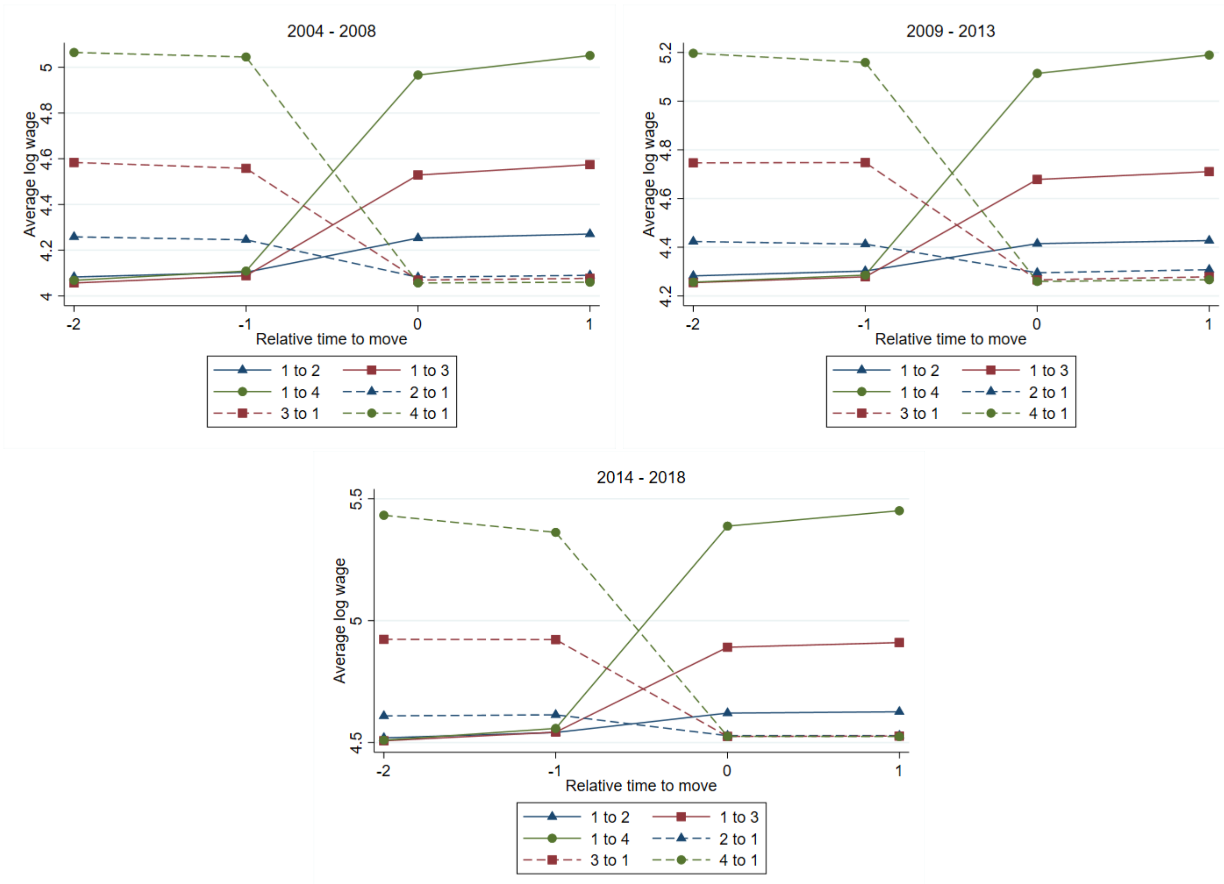
This index grows larger as the share of workers that commute between municipalities i and j becomes smaller. After building the index, they cluster the municipalities based on this index using a hierarchical clustering algorithm.

C Exchangeability

We reproduce the evidence of exchangeability shown in Pérez Pérez and Nuño-Ledesma (2022) in this section. According to Card et al. (2013), if the residual term in equation (1) is uncorrelated with the variables on the right-hand side, workers who move from workplace A to workplace B should, on average, experience a wage change opposite in sign to workers moving in the opposite direction. To explore this in our dataset, we follow Card et al. (2013) and present an event study in Figure C.1, adapted from Pérez Pérez and Nuño-Ledesma (2022). The plot illustrates the average wages of workers who changed jobs during each time interval of our analysis period. These workers may transition from ”low-wage” to ”high-wage” workplaces or vice versa. We categorize workplaces based on the quartile of the average wage of their co-workers in the initial job and the corresponding quartile in the final job. We then calculate average wages before and after the job switch for each category. Our analysis excludes observations from establishments with only one worker and focuses on ”direct” moves, meaning moves without an unemployment spell between jobs.

The Figure reveals that different mobility groups, classified by the average wage of co-workers, have distinct average wage levels before and after a job move. Prior to the move, average wages in the quartile of origin exhibit a monotonic variation with respect to the destination quartile. For instance, workers moving from quartile 4 (highest average co-worker wage) to quartile 1 (lowest mean co-worker wage) have higher average wages before the job switch compared to those moving from quartile 3 to 1, and so on. Additionally, the absolute change in average wages when transitioning from one quartile to another is equivalent in magnitude to the variation associated with the opposite change. This symmetry aligns with an additive wage model that incorporates worker and workplace fixed effects, similar to the one we estimate.

Figure C.1: Exchangeability: Average Log Wage Around Movement by Quartile of Average co-workers' Wages in the Origin and Destination Workplace



Source: Authors' calculations using IMSS data. The graph shows average worker wages for workers who move between an origin workplace to destination workplace, from two months before the move to one month after the move. The lines group workers according to the quartiles of average co-worker wages in the origin and destination workplaces. The panels correspond to different time intervals. We exclude observations from establishments with only one worker. We keep only "direct" moves without an unemployment spell in the transition between jobs.

D Variance Decomposition with Limited Mobility Bias Corrections

In this section we show that the results of variance decompositions are similar in the baseline estimates and in estimates with limited mobility bias corrections. Table D.1 show variance decompositions with different corrections.

Columns 1 to 3 show the baseline estimates. Columns 4 to 6 show estimates from a model clustering firms into groups following Bonhomme et al. (2019). We calculate twenty percentiles of the within-firm distribution of wages. Then, we cluster firms into five clusters using the percentile values as variables to cluster. Last, we reestimate the AKM model using firm cluster indicators instead of firm indicators, and recalculate the variance decomposition. The results show slightly higher variance shares attributed to assortative matching.

Columns 7 to 9 show results of the variance decomposition using a leave-one-out variance components estimators from Kline et al. (2020). In this case, the variance shares are similar to those from the baseline estimates.

Table D.1: Variance Decompositions with Different Corrections for Limited Mobility Bias

	Baseline			Bonhomme et al. (2019)			Kline et al. (2020)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Interval 1 2004-2008	Interval 2 2009-2013	Interval 3 2014-2018	Interval 1 2004-2008	Interval 2 2009-2013	Interval 3 2014-2018	Interval 1 2004-2008	Interval 2 2009-2013	Interval 3 2014-2018
Variance and covariance									
Total variance of log wages	0.65	0.68	0.69	0.65	0.68	0.69	0.64	0.67	0.67
Variance of worker effects	0.29	0.27	0.25	0.28	0.26	0.25	0.29	0.27	0.25
Variance of workplace effects	0.21	0.24	0.25	0.16	0.19	0.20	0.20	0.23	0.25
2 Cov(worker effects, workplace effects)	0.10	0.12	0.13	0.16	0.17	0.18	0.10	0.12	0.13
Variance shares									
Variance of worker effects	0.44	0.40	0.37	0.43	0.39	0.37	0.46	0.41	0.37
Variance of workplace effects	0.33	0.36	0.37	0.25	0.28	0.29	0.32	0.35	0.37
2 Cov(person effects, firm effects)	0.16	0.18	0.20	0.25	0.25	0.26	0.17	0.19	0.21

Source: Authors' calculations using IMSS data. The columns display the results of variance decompositions following equation (2) of the main text. Columns (1) to (3) display the baseline estimates using unadjustes estimated worker and workplace fixed effects. Columns (4) to (6) display 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. Columns (7) to (9) show estimates using the Kline et al. (2020) correction, where the variance components are calculated using leave-one-out estimators over the connected set with observations from January, May and September for each year.

E Additional Estimates

Table E.1: Estimates with Log Employment as Independent Variable

	(1) Metro	(2) Metro-Industry	(3) CZ	(4) CZ-Industry
Dependent variable: correlation of worker and plant FE				
Log employment	0.0228***	0.0295***	0.0500***	0.0621***
	(0.004)	(0.004)	(0.002)	(0.001)
R^2	0.354	0.065	0.245	0.153
N	97	1,648	1,961	10,118

Source: Author's calculations using IMSS and census data. The columns show results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log employment, 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. For column 4, we restrict to cells with more than 50 workers and more than 5 firms. Robust standard errors in parentheses. *: p<0.1, **: p<0.05, ***: p<0.01.

Table E.2: City Size and Assortative Matching in Mexico's Formal Labor Markets: Estimates by Interval

Dependent variable: correlation of worker and plant FE				
	(1)	(2)	(3)	(4)
	Metro	Metro-Industry	CZ	CZ-Industry
A: 2004-2008				
Log Population	0.0199**	0.0136*	0.0689***	0.0569***
	(0.010)	(0.007)	(0.007)	(0.003)
N	32	543	614	3209
R^2	0.139	0.005	0.147	0.085
B: 2009-2013				
Log Population	0.0202***	0.0163**	0.0663**	0.0554***
	(0.007)	(0.008)	(0.007)	(0.003)
N	32	544	668	3387
R^2	0.178	0.007	0.130	0.081
C: 2014-2018				
Log Population	0.0215***	0.0313***	0.0685***	0.0477***
	(0.005)	(0.007)	(0.006)	(0.003)
N	33	561	679	3,522
R^2	0.256	0.029	0.166	0.073

Source: Author's calculations using IMSS and census data. The columns show results of regressions of the correlation coefficient between worker and workplace effects from AKM model estimates and log employment, at different geographical aggregation levels. Each panel corresponds to a different time interval. Metro stands for metropolitan area. CZ stands for commuting zone. For column 4, we restrict to cells with more than 50 workers and more than 5 firms. Robust standard errors in parentheses. *: p<0.1, **: p<0.05, ***: p<0.01.

Appendix References

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

Aldeco, L., Calderón, M., Chiquiar, D., Hanson, G., Pérez Pérez, J., and Velásquez, C. (2023). Mercados de trabajo locales en México: Definición, bases de datos y análisis descriptivo 1990-2020. Technical report, Banco de México

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, 87(3):699–739

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967–1015

Fowler, C. S. and Jensen, L. (2020). Bridging the Gap Between Geographic Concept and the Data We Have: The Case of Labor Markets in the USA. *Environment and Planning A: Economy and Space*, 52(7):1395–1414

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

Pérez Pérez, J. and Nuño-Ledesma, J. (2022). The Contribution of Workers, Workplaces, and Sorting to Wage Inequality in Mexico. Documento de Trabajo WP 2022-012, Alianza EFI