Labor Market Power Across Cities

Claudio Luccioletti*

October 17, 2024

Abstract

Workers in larger cities are paid higher wages. The city-size wage premium may reflect the productivity gains from agglomeration or sorting of more productive workers in densely populated areas. However, local labor markets in large cities have more firms and are expected to be more competitive, which could also generate part of the urban earnings premium. I quantify the importance of this channel with rich administrative data for Spain using a spatial equilibrium model to guide the empirical strategy. To address the identification challenge posed by labor market power and wages moving endogenously with unobserved local productivity shocks, I first control for firms' revenues per worker and market-level trends. I then develop a new instrumental variable that leverages quasi-experimental variation in monopsony power stemming from changes over time in the size of local public firms. I conclude that 20–30% of the city-size wage premium can be attributed to differences in labor market power across locations.

Keywords: Labor Market Power; City Sizes; Wage Premium.

JEL Classification: R10; J42; R23.

*Bank of Italy, claudio.luccioletti@cemfi.edu.es. I am very grateful for comments and suggestions from Dmitry Arkhangelsky, Samuel Bentolila, Paula Bustos, Micole De Vera, Sebastián Fanelli, Nezih Guner, Zijian He (discussant), Javier Quintana González, Monica Martinez-Bravo, Felix Montag (discussant), Enrique Moral-Benito, Josep Pijoan-Mas, Diego Puga, Virginia Sánchez-Marcos, Tate Twinam (discussant), and seminar participants at the 2021 and 2022 North American Meetings of the Urban Economics Association, at the 2022 EEA-ESEM Congress, at the 2022 SAEe, and at the 2023 RIEF conference. I acknowledge financial support from Grant PRE2020-094296 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future", the María de Maeztu Unit of Excellence (CEMFI MDM-2016-0684), funded by MCIN/AEI/10.13039/501100011033, and CEMFI.

1 Introduction

It has long been observed that people who live in large cities earn higher wages than those living in smaller towns. To explain the *city-size wage premium*, a broad empirical literature has attempted to identify the productivity advantages of highly populated urban areas. On the one hand, *agglomeration economies* allow firms and workers to be more productive in larger cities. On the other hand, big cities attract and retain more talented workers and entrepreneurs. However, wage differentials do not fully reflect productivity differences when labor market are imperfectly competitive and employers pay workers less than their marginal product. Local labor markets in larger urban areas tend to host more firms, and so are expected to display higher levels of competition on average. Because firms operating in competitive labor markets are forced to share more profits with workers by raising their wages, this mechanism has the potential to explain part of the city-size wage premium.

In this paper, I quantify the fraction of the urban earnings premium that can be attributed to differences in labor market power between small and large cities. I start by building a simple Rosen-Roback spatial equilibrium model in which wages in each city depend on local productivity and local labor market power of firms. Productive cities attract many competitor firms in equilibrium, whereas less productive locations host fewer employers who, unchecked by competition, exert labor market power over their employees. Workers choose where to live taking economic and noneconomic factors into account. If they are more mobile across locations (for example, owing to a more elastic housing supply), they have a larger set of job positions to choose from, which effectively limits employers' market power in low-productivity cities.

I then estimate the impact of labor market power on wages using matched employer-employee data for Spain and an empirical strategy derived from the equilibrium relations of the model. As emphasized in the model, labor market power and wages may move endogenously with unobserved productivity shocks. For instance, positive productivity shocks can increase wages and, by inducing competitors' entry, reduce labor market power. To achieve identification, I follow two complementary approaches. First, I control for unobserved labor market productivity using market-level trends and a rich set of fixed effects. Additionally, I use balance sheet information for the quasi-universe of Spanish firms to control for revenue productivity at the local labor market level.

As for the second strategy, I propose a new instrumental variable that exploits changes in the size of local public firms to provide exogenous variation in labor market power. In some local labor markets, public and private firms are competitors hiring from the same pool of workers. Therefore, idiosyncratic movements in the size of public employers (e.g., due to a policy change at a higher administrative level that is unrelated to local economic conditions) can influence workers' wages in the private sector by affecting the level of competition among potential recruiters. I motivate the exogeneity assumption of the instrument by showing that public firms' contribution to labor market concentration is not related to local revenue productivity when I focus on health- and education-related markets, which are the industries to which I restrict attention in the IV analysis.

The estimated impact of labor market concentration on wages that I obtain from both strategies is comparable in magnitude and consistent with prior empirical studies. Transitioning from fully unconcentrated labor markets to the situation with the highest level of monopsony power (i.e., the case of a single employer operating in the market) is associated with a causal reduction in wages of around $7-14\%$. Given the differences in the degree of labor market competition between small and big cities, these estimates imply that monopsony power can explain approximately $20-30\%$ of the city-size wage premium.

This study is closely related to the large literature on the determinants of the urban wage premium (Glaeser and Maré, 2001). The existence of agglomeration economies (De la Roca and Puga, 2017, Duranton and Puga, 2004) and sorting of more productive workers and firms to large cities (Behrens et al., 2014) are the explanations that are typically put forward to rationalize the urban premium in earnings. However, these papers generally assume that labor markets are perfectly competitive, thus ruling out any possible explanation related to differences in monopsony power between small and large cities. Hirsch et al. (2022) is an exception. Using German administrative data, they find that differences in labor market imperfections between urban areas of different size explain approximately 40% of the city-size wage premium. In their empirical analysis, they use data on hires coming from non-employment (as opposed to employment) as an instrument for labor market frictions. However, this variable is likely to be correlated with local unemployment and, consequently, with unobserved productivity in the city. Therefore, this strategy might not be able to fully separate the effect of local labor market power on urban wages from the influence of agglomeration.

Another related paper is Azkarate-Askasua and Zerecero (2022). Using a structural model calibrated to the French economy, they conclude that employers' labor market power accounts for around a third of the observed urban-rural wage gap. For identification, they use national mass layoff shocks affecting employment shares of firms competing in the same local labor market as the treated establishment. I instead propose a new instrument for employment concentration based on changes in the size of local public firms, which is found to be unrelated to local economic conditions. This connects my work to Guillouzouic et al. (2022), who show that the public sector exerts monopsony power in the labor market. Moreover, I explicitly control for revenue productivity and market-level trends.

This paper also builds on the growing body of empirical and theoretical work examining the impact of monopsony power on workers' outcomes (Manning, 2011, Sokolova and Sorensen, 2021). In particular, it is closely linked to Manning (2010), an early paper that emphasizes the connection between labor market power and city size with a model in which the labor supply elasticity is endogenous to the number of firms in the market. In his framework, large cities host many firms that face a very elastic labor supply and high competition in the labor market. There exists a broad literature that measures monopsony power by estimating reduced-form labor supply elasticities for individual firms, using both empirical methods (Bassier et al., 2022) and structural approaches from industrial organization (Azar et al., 2022). Moreover, Bamford (2021) and Datta (2022) estimate spatial models that account for local monopsony power, the former leveraging within worker-region variation, and the latter using two instruments based on firm-specific wage floors and irregularities in job advertisements.

I follow a complementary approach and, as in much of the recent literature on labor market power (e.g., Arnold, 2022, Azar and Vives, 2021), I assume that firms compete *à la Cournot* for workers. In the model, the labor supply elasticity is constant, while variations in labor market power across space only come from changes in labor market concentration, as measured by the employment Herfindahl-Hirschman Index (HHI). This paper introduces a new instrument for HHI, adding to a recent literature that exploits merger-induced variation (Arnold, 2022, Benmelech et al., 2022) or shift-share shocks based on national firms' hiring growth (Schubert et al., 2024) as IVs. Quantitative spatial models that impose a similar oligopsony structure include Berger et al. (2022) and Azkarate-Askasua and Zerecero (2022).

While several recent empirical studies estimate a strong negative relationship between HHI and wages across local labor markets (e.g., Azar et al., 2020 and Lipsius, 2018), two issues complicate a causal interpretation. On the one hand, the observed correlation between HHI and earnings could be spuriously determined by unobserved factors affecting both variables (Berry et al., 2019). On the other hand, to make HHI operational, one has to define what a local labor market is, but the literature has not settled on a single satisfactory definition.[1] In this paper, I address the first issue by exploiting only the sources of variation in HHI that are not driven by unobserved factors (e.g., productivity) that may endogenously affect wages. Regarding the second issue, I provide a data-driven definition of local labor markets (Nimczik, 2020), which clusters subindustries into different markets based on worker flows. Hence, a local market is a collection of subindustries in a city such that when workers change jobs, they tend to stay within the given set of subindustries in that location. This approach connects

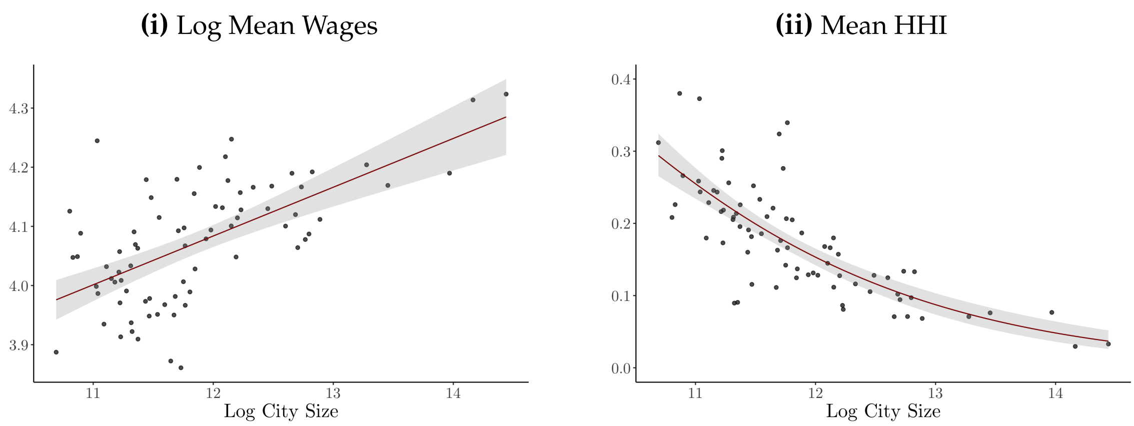
[1] These papers usually define local labor markets as combinations of commuting zones and occupations (e.g., Azar et al., 2020) or of commuting zones and industries (e.g., Benmelech et al., 2022).

with a broader strand of literature that moves beyond traditional labor market definitions by using data-driven techniques (Caldwell and Danieli, 2024, Jarosch et al., 2024, Manning and Petrongolo, 2017, Nimczik, 2020).

1.1 Wages and Labor Market Power Across Cities

Large Spanish cities pay substantially higher wages on average. As panel (i) of Figure 1 shows, the difference in mean earnings offered in small and large cities is of approximately 0.3 log points. Employment concentration is also lower in larger cities, as shown in panel (ii). Concentration is measured using the employment Herfindahl-Hirschman Index, $\text{HHI}_m = \sum_{f=1}^{N_m} s_f^2$, where s_f is the employment share of firm f operating in market m and N_m is the number of competing firms. The HHI is bounded between 0, indicating fully unconcentrated labor markets (with atomist firms and $s_f = 0$), and 1, which occurs when a single monopsonists operates in the market. The evidence shows that labor markets in large cities such as Madrid or Barcelona have very low concentration levels, whereas smaller cities like Utrera show significantly higher concentration, with an average HHI of around 0.3.

Figure 1: Wages and HHI across cities of different size

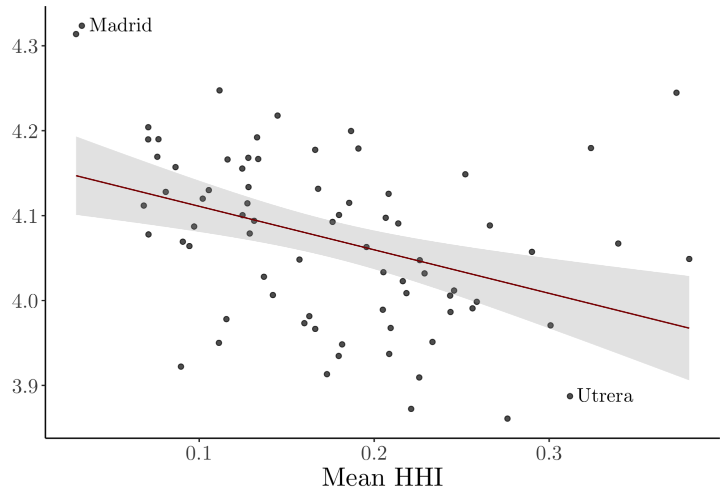


Note: These figures plot market mean wages (panel (i)) and HHI (panel (ii)) as functions of the size of the city where the markets are located. Mean wages and employment HHI are computed for local labor markets and averaged across time at the city level (market employment weights are used). Labor markets are clusters of subindustries within cities, estimated to minimize cross-cluster worker flows (source: MCVL, years 2005-2019). City size is population within 10km of the average resident (De la Roca and Puga, 2017).

Finally, Figure 2 shows the correlation between wages and HHI. Wages are higher in cities where local labor markets tend to be less concentrated. This relationship could be causal (low levels of labor market power put upward pressure on wages) or spurious. In Section 2, I outline a simple model that highlights a series of relevant (causal and spurious) channels through which labor market power and local wages are connected.

The paper is organized as follows. Section 2 presents the model's economy. Section 3 discusses the research design. Section 4 provides estimates of the extent of the urban wage premium that can be attributed to monopsony power in the labor market. Section 5 concludes.

Figure 2: Log Mean Wages



Note: This figure plots mean wages as a function of mean HHI in the same city. Mean wages and employment HHI are computed for local labor markets and averaged across time at the city level (market employment weights are used). Labor markets are clusters of subindustries within cities, estimated to minimize cross-cluster worker flows (source: MCVL, years 2005-2019).

2 Model

The model presented in this section extends the stylized Rosen-Roback framework described in Moretti (2011) by allowing for imperfect competition in the labor market. Workers and firms are mobile and choose to locate in the city that gives them higher utility and profit. Firms employ workers living in the city where they operate and have monopsony power in the local labor market. The model rationalizes the evidence presented in Section 1.1 (Figures 1 and 2) as a spatial equilibrium in which high-productivity big cities display low labor market power, low-productivity small cities display high labor market power, and firms and workers have no incentive to move.[2]

[2] Although labor markets in the model are imperfectly competitive, I assume that there is no product market power and that there are no unions. I control for the influence of both variables in the empirical analysis of Section 4.

2.1 Workers

There are two cities, s (*small*) and b (*big*). Local labor markets, indexed by m, are given by combinations of cities $c \in \{s, b\}$ and industries $k \in \{1, ..., K\}$. The indirect utility of individual i living in city c and working in industry k is given by

$$U_{ick} = \log(W_{ck}) - r_{ck} + b_c + e_{ick},$$

where W_{ck} denotes wages in the local labor market, r_{ck} measures housing costs, b_c indicates local amenities, and e_{ick} measures idiosyncratic preferences for city c for workers operating in industry k. Workers cannot change their industry k, but are free to move between cities. This allows them to remain in industry k while migrating to a different city.

Idiosyncratic preferences of workers for the *small* relative to the *big* city are uniformly distributed as

$$e_{isk} - e_{ibk} \sim U[-z, z].$$

Parameter z governs the importance of idiosyncratic preferences in workers' decisions to be located in a certain city, big or small. If z is low, idiosyncratic preferences for cities are less important, and workers are more willing to migrate to arbitrage away differences in real wages and amenities across cities.[3] As z increases, workers become less mobile, as they have a higher idiosyncratic taste for the city they are currently living in. As a result, they are less likely to out-migrate from a city even if that city's economic outcomes or amenities worsen.

Housing supply faced by workers in industry k living in city c is given by

$$r_{ck} = r + \kappa \log(L_{ck}),$$

where L_{ck} denotes the number of workers and it is assumed that each worker consumes one housing unit. Parameter $\kappa > 0$ denotes the housing supply elasticity.[4]

Each worker i chooses city $c \in \{s, b\}$ depending on whether U_{isk} or U_{ibk} is higher. Therefore, the number of workers in each city is determined endogenously. If a city pays higher wages, it attracts a larger number of workers. Because housing costs increase with population, that city also becomes less attractive: housing prices act as a congestion force.

From the indifference condition of the marginal worker ($U_{isk} = U_{ibk}$), we can derive

[3] Workers are perfectly mobile if $z = 0$.

[4] For simplicity, we assume perfect residential segregation along skill lines within a city. Therefore, rents in city c faced by a worker in industry k do not depend on the number of workers in a different industry k' living in the same city.

the local labor supply in city b and industry k as

$$\log(W_{bk}) = \underbrace{g_{sk} - b_b}_{\log(\beta_{b(s)k})} + \underbrace{(z+\kappa)}_{\eta^{-1}} \log(L_{bk}), \tag{1}$$

where

$$g_{sk} = \log(W_{sk}) + b_s - (z+\kappa)\log(L_{sk})$$

measures the attractiveness of city s for workers in industry k, $\log(\beta_{b(s)k})$ is the local labor supply intercept, and η^{-1} is the inverse labor supply elasticity.[5] Equation (1) states that workers in city b accept lower wages if the *big* city has better amenities, but want to receive higher compensation if the city is large and/or if the outside option – that is, the *small* city – is attractive. Labor supply in city s is symmetric.

Labor supply elasticity η measures workers' willingness to migrate. If the labor supply is highly elastic (low η^{-1}), then the housing congestion forces and idiosyncratic preferences for specific locations are less important, such that small increases in wages W_{ck} attract a large influx of migrants and translate into large changes in the number of workers L_{ck}. Therefore, elasticity η^{-1} plays a key role in the analysis of labor market power, as it governs workers' willingness to move out of cities with high monopsony power to find jobs in cities that pay them at a more competitive rate. If workers are highly mobile, their credible threat to leaving the city restricts employers' ability to set wages below the marginal product of labor, limiting firms' monopsony power.

2.2 Firms

Firms operate in industry k in one of the two cities, and employ workers in that local labor market to produce a good that is freely traded with the other city. Like workers, they cannot change their industry k and are free to relocate between cities. The price of the final good is normalized to one. While there is perfect competition in the final goods market, the market for labor, which is the unique input of production, is imperfectly competitive. In particular, firms compete *à la Cournot* for all workers in a local labor market, and they internalize that the labor supply is upward sloping and given by expression (1).

Firms operating in city c and industry k have a Cobb-Douglas production function

$$Q_{fck} = A_{fck} l_f^{\theta}, \quad \theta \leq 1,$$

where A_{fck} is the firm-specific productivity term. This productivity term depends on

[5] There is no population growth and we use the normalization $\log(L_{sk}) + \log(L_{bk}) = 1$.

the particular city and industry where the firm is located, with

$$A_{fck} = \underbrace{A_c A_k}_{A_{ck}} \epsilon^A_{fck} \tag{2}$$

and

$$A_c = A \left(\sum_{k=1}^{K} L_{ck} \right)^{\delta} \epsilon^A_c, \tag{3}$$

where $L_{ck} = \sum_{f=1}^{N_{ck}} l_f$ is total employment in the local labor market and N_{ck} is the local number of firms. Parameter $\delta > 0$ in equation (3) is the *agglomeration elasticity*, which ensures that the big city, where total employment is larger, has a productivity advantage. This captures, in a reduced form way, agglomeration economies such as knowledge spillovers or labor pooling (Duranton and Puga, 2004). Firms do not internalize the marginal effect of their employment decisions on the city-level productivity, i.e.

$$\frac{\partial A_c}{\partial l_f} = 0.$$

Firms are perfectly mobile across cities, but entry takes one period.[6] The number of firms in each city, endogenously determined in equilibrium, is denoted by N_c, with $N_c = \sum_{k=1}^{K} N_{ck}$.

Given labor supply $W(L_c) = \beta_{c(c')k} L_{ck}^{\eta^{-1}}$, where c' denotes the other city, firms choose employment l_f to maximize profits

$$\pi_{fck} = \max_{l_f} A_{fck} l_f - W(L_{ck}) l_f,$$

The first-order condition gives

$$W_{ck} = \underbrace{(1 + \eta^{-1} \text{HHI}_{ck})^{-1}}_{\text{Markdown}} \text{AMRPL}_{ck}, \tag{4}$$

where

$$\text{HHI}_{ck} = \sum_{f=1}^{N_{ck}} s_f^2 = \sum_{f=1}^{N_{ck}} \left(\frac{l_f}{L_{ck}} \right)^2, \tag{5}$$

is the Herfindahl-Hirschman Index (HHI) and

$$\text{AMRPL}_{ck} = \theta A_c A_k \sum_{f=1}^{N_{ck}} s_f \epsilon^A_{fck} l_f^{\theta - 1}$$

is the average marginal revenue product of labor. If the labor market is perfectly com-

[6] The period subscript t is, for ease of exposition, suppressed for now.

petitive, i.e., if firms' employment shares are infinitesimal ($s_f \to 0$), HHI is zero and workers are paid the marginal revenue product of labor AMRPL_{ck}. If the number of firms is finite, on the other hand, firms exert labor market power and pay them a fraction $(1+\eta^{-1}\text{HHI}_{ck})^{-1}$ of AMRPL_{ck}. This fraction, which we refer to as the *markdown*, shrinks as local labor markets become more concentrated (i.e., as HHI_{ck} increases). Thus, markdowns provide a sufficient measure of labor market power in the model. The employment dynamics of large firms in the market are particularly relevant for the evolution of HHI_{ck} over time because employment shares enter equation (5) with a square. This captures the idea that dominant firms are the main actors behind market power in local labor markets.

Markdowns depend on $\eta^{-1}\text{HHI}_{mt}$, i.e., the extent of labor market competition between firms and workers' ability to "escape" from it by out-migrating. Indeed, there is no labor market power if either:

i. $\text{HHI}_{ck} \to 0$, i.e., there is perfect competition in the labor market;

ii. $\eta^{-1} = (z+\kappa) = 0$, i.e., there are no idiosyncratic preferences (perfect mobility) *and* the elasticity of housing is perfectly elastic (no congestion).

Because η is assumed to be fixed, variations in labor market power come only from changes in HHI_{ck}.

To obtain a simple closed-form solution for firms' profits, we assume that the production function has constant returns to scale ($\theta = 1$) and that labor supply is linear ($\eta = 1$). Moreover, we assume that firms are symmetric within local labor markets ($\epsilon^A_{fck} = 1$), so that $l_f = \frac{L_{ck}}{N_{ck}}$, $\text{HHI}_{ck} = \frac{1}{N_{ck}}$, and $\text{AMRPL}_{ck} = A_{ck}$.

Given equation (4), profits are then given by

$$\pi_{ck} = \frac{1}{(1+N_{ck})^2}\frac{A^2_{ck}}{\beta_{c(c')k}}.$$

Firms pay a market-specific fixed cost F_{ck} of production, which captures, for example, the cost of maintaining a human resource department or the bureaucratic burden of operating in the market. Free entry commands $\pi_{ck} = F_{ck}$. Thus, there is perfect arbitrage between the *small* and *big* cities:

$$\frac{1}{(1+N_{sk})^2}\frac{A^2_{sk}}{\beta_{s(b)k}} - F_{sk} = \frac{1}{(1+N_{bk})^2}\frac{A^2_{bk}}{\beta_{b(s)k}} - F_{bk} = 0, \tag{6}$$

where $\beta_{b(s)k} = \beta_{s(b)k}\frac{b_s}{b_b}$.

Consider the case in which $F_{sk} = F_{bk}$. Then, equation (6) can be rewritten as

$$\frac{A^2_{sk}b_s}{(1+N_{sk})^2} = \frac{A^2_{bk}b_k}{(1+N_{bk})^2}$$

If amenities are similar between the small and the big city ($b_s \simeq b_b$), then $A_{bk} > A_{sk}$ implies $N_{bk} > N_{sk}$.[7] In other words, if the *big* city is more productive, we should expect its local labor markets to be more competitive ($\text{HHI}_{bk} < \text{HHI}_{sk}$), which is indeed confirmed by the data (see Figure 1). With $A_{bk} > A_{sk}$, city b is more attractive to firms and higher firm entry translates to lower labor market power. Thus, the two channels contribute to the city-size wage premium ($W_{bk} > W_{sk}$) through the market equilibrium condition (4): higher productivity ($A_{bk} > A_{sk}$) and lower labor market power ($\text{HHI}_{bk} < \text{HHI}_{sk}$) in city b.

Finally, by the free entry condition (6),

$$\text{HHI}_{ck} = \frac{\sqrt{\beta_{c(c')k} F_{ck}}}{A_{ck} - \sqrt{\beta_{c(c')k} F_{ck}}}. \tag{7}$$

City c is more attractive to firms if it has low fixed costs because gross profits are higher in that case. It is also more attractive if it has relatively higher amenities (low $\beta_{c(c')k}$) because workers accept lower wages. Consequently, HHI_{ck} increases with F_{ck} and $\beta_{c(c')k}$. In the next section, I introduce another source of variation in HHI_{ck}, which I exploit for the construction of the IV in the empirical strategy: changes in the employment of local *public* firms.

2.3 Private and Public Sector

Assume that firms belong to either the *private* or *public* sector, $N_{ck} = N_{ck}^{\text{priv}} + N_{ck}^{\text{pub}}$. *Private* firms maximize profits and set wages according to the first-order condition (4). As a result, a higher A_{ck} or a lower $\beta_{c(c')k}$ will, all else equal, induce entry of private firms in the city (increase in N_{ck}^{priv}) and affect the employment decisions of incumbent private firms. Conversely, *public* firms are not profit maximizers and their entry or exit decisions are not related to market conditions. In particular, the employment shares of each public firm f evolves over time according to $s_{ft+1}^{\text{pub}} = (1 + p_{fckt}) s_{ft}^{\text{pub}}$, where p_{fckt} is unrelated to AMRPL_{ck}, $\beta_{c(c')k}$ or F_{ck}. For example, p_{fckt} may be the outcome of regional governmental policies that change after an election and are assumed to evolve over time unrelated to *local* economic conditions.

Public firms employ workers. For simplicity, assume that $W_{ck} = W_{ck}^{\text{priv}} = W_{ck}^{\text{pub}}$ and that there are no frictions in hiring, such that workers are indifferent between being employed in the public or private sector.[8] Thus, changes in the employment shares

[7] The evidence discussed in Section 4.3, indicating that exogenous amenities do not account for a significant portion of the city-size wage premium, is in line with the assumption that $b_s \simeq b_b$.

[8] In a more realistic model in which private jobs tend to be more remunerative and volatile, i.e. $\mathbb{E}(W_{ck}^{\text{priv}}) > \mathbb{E}(W_{ck}^{\text{pub}})$ and $\text{Var}(W_{ck}^{\text{priv}}) > \text{Var}(W_{ck}^{\text{pub}})$, risk-averse workers may still be indifferent between public and private firms operating in the same market. The Spanish public sector's hiring process typically involves applicants passing public exams, which challenges the assumption of frictionless hiring

of public firms have a direct impact on the degree of local labor market power; that is, $\text{HHI}_{ck} = \sum_{f=1}^{N_{ck}^{\text{pub}}} s_f^2 + \sum_{f'=1}^{N_{ck}^{\text{priv}}} s_{f'}^2$. For instance, a decrease in concentration driven by a lower $\sum_{f=1}^{N_{ck}^{\text{pub}}} s_f^2$ means that workers in the private sector have more outside options and, as a consequence of the increased competition among their potential recruiters, can expect their wages to raise.[9] Such changes in $\sum_{f=1}^{N_{ck}^{\text{pub}}} s_f^2$ affect earnings only through their impact on HHI_{ck} (see first-order condition (4)), given the assumption that p_{fckt} is unrelated to AMRPL_{ck}, $\beta_{c(c')k}$ or F_{ck}[10]. Therefore, changes in the local size of public firms provide exogenous variation in labor market power, which can be used to identify the impact of concentration on private wages.

2.4 Summary

The following equations and the directed acyclic graph (DAG) in Figure 3 summarize the equilibrium of the model. The nodes in the diagram represent the relevant variables in the data generating process, that can either be unobservable, in which case they are enclosed within dashed lines, or observable. Causal relationships in the model are represented by arrows from the cause to the caused variable. To simplify notation, I index local labor markets (defined by their city c and industry k) with m.

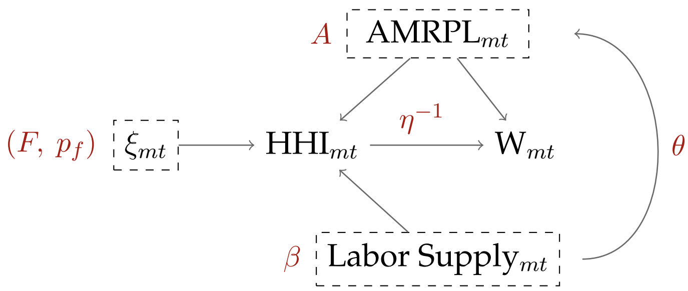
The Herfindahl-Hirschman Index and wages are directly related through the first-order condition (4). As firms take the number of competitors as given when setting the wage level, employment concentration has a causal impact on workers' earnings, and the effect is stronger if η^{-1} is higher. On the other hand, HHI_{mt} and W_{mt} are endogenous objects that are indirectly related because they are both influenced by A_{mt}, which is exogenous, and β_{mt}, which depends on exogenous amenities, b_{ct}, as well as on wages and employment in other local labor markets (taken as given). Therefore, a positive shock to productivity ($\uparrow A_{mt}$) induces firm entry ($\downarrow \text{HHI}_{mt}$) and increases wages through the FOC ($\uparrow W_{mt}$). Similarly, a positive amenity shock ($\downarrow \beta_{mt}$) induces entry ($\downarrow \text{HHI}_{mt}$) and, if there are decreasing returns to scale, negatively affects wages by lowering AMRPL_{mt} ($\downarrow W_{mt}$), as firms are induced to hire more workers who now

decisions. Nonetheless, worries are alleviated by the significant worker flows between private and public firms observed in the IV sample, which averages 10-20% of the total (see Section 4.5). Moreover, a robustness check was conducted focusing only on markets with the highest rate of turnover between public and private firms (see Section 4.5.1).

[9]Note that equation (7) only holds if $N_{ck}^{\text{pub}} = 0$. When $N_{ck}^{\text{pub}} > 0$, firms in the private sector take the number of public firms and their employment decisions as given when maximizing profits. From the assumption that $W_{ck}^{\text{priv}} = W_{ck}^{\text{pub}}$ and of symmetric firms, it follows that firms' employment is identical in the public and private sector.

[10]A potential threat to exogeneity arises from the fact that the employment shares of public firms include changes in the total employment of private firms in the denominator, which could be affected by shifts in productivity. To address this issue, the instrumental variable used in the analysis is designed to exclude the influence of private firms in the denominator. For further details, refer to Sections 4.5 and Appendix C.2.

Figure 3: A DAG summary of the model



$$W_{mt} = (1 + \eta^{-1}\text{HHI}_{mt})^{-1}\text{AMRPL}_{mt}, \qquad \text{(FOC)}$$
$$\text{HHI}_{mt} = h(A_{mt}, \beta_{mt}, F_{mt}, p_{fmt}), \qquad \text{(Firms)}$$
$$W_{mt} = \beta_{mt} L_{mt}^{\eta^{-1}}. \qquad \text{(Workers)}$$

Note: This figure draws a directed acyclic graph (DAG) summarizing the equilibrium of the model. The variables are enclosed within dashed lines if they are unobservable. Arrows represent causal relationship between the variables in the data generating process.

accept lower wages, and these additional workers are marginally less productive.[11]

Is there, then, a source of *exogenous* variation ξ_{mt} in HHI_{mt} that allows us to identify the effect of a change in labor market competition on wages? In the model, changes in the employment shares of public firms (p_{fmt}) and shocks to fixed costs F_{mt} can serve this purpose since they only affect W_{mt} through changes in HHI_{mt}. In Section 3, I propose an estimation strategy that uses this source of exogenous variation in the Herfindahl-Hirschman Index to quantify the extent to which the city-size wage premium can be attributed to differences in labor market imperfections in cities of different sizes. Finally, note that all endogeneity concerns are ultimately due to AMRPL_{mt}, which is typically unobserved in the data.

3 Estimation

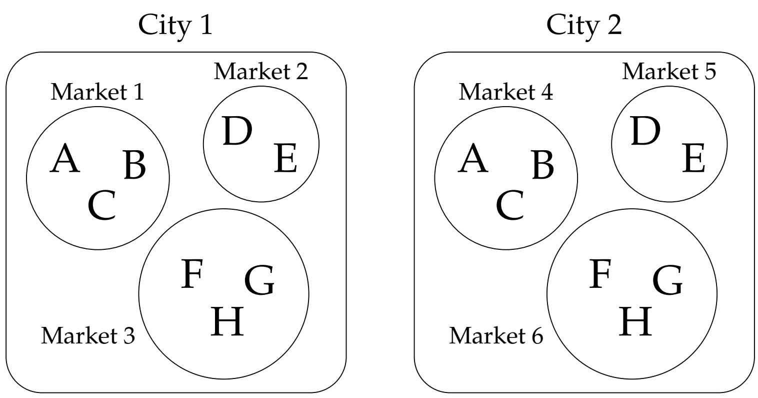
3.1 Local Labor Markets

In the model, local labor markets are islands within cities. If labor market power increases in market m of city c, then workers employed in that market will see a decrease

[11]See Appendix D.1.2 for the derivations of AMRPL_{mt} in the case of decreasing returns to scale technologies. Note that, since firm entry takes one period in the model while changes in AMRPL_{mt} impact wages immediately, shocks to A_{mt} and β_{mt} create an endogenous connection between HHI_{mt} and W_{mt} only if they are serially correlated over time. Furthermore, the presence of asymmetric productivity shocks raises additional endogeneity concerns, as shown in Appendix D.2. As with the symmetric firms model, these concerns are ultimately due to the dynamics of market level productivity AMRPL_{mt}.

in their wages, whereas workers in market m' of the same city will not be affected. This is because workers cannot move across markets within the same city and firms cannot employ workers outside their own labor market. Consistently, for the empirical analysis, I adopt a flexible definition of local labor markets based on worker flow data.

Figure 4: Local labor markets



Note: This figure draws an example of six markets in two cities, grouping subindustries (indexed by letters A-H) that are linked by worker flows into different clusters.

In particular, I use the algorithm proposed by Nimczik (2020) to identify clusters of 3-digit subindustries connected by worker flows. This algorithm views subindustries as nodes in a network, connecting them with job-to-job transitions observed in the data. Two subindustries are then deemed to be part of the same cluster if they share similar probabilities of being linked to the rest of the network.[12] My definition of labor market is given by the combination of these clusters and cities, which can be thought of as commuting zones. Figure 4 shows an example. The group of subindustries A, B, and C forms a distinct market in each city, since in the data we see workers mainly moving between these three subindustries, while flows to and from other subindustries are comparatively limited.

The worker flows used to estimate local labor markets are computed at the national level, so that clusters do not vary across cities.[13] In Section 4.4.2, I present robustness

[12]Worker flows reveal directed links across any two subindustries, which are weighted by the count of job-to-job transitions across them. Given this structure, I estimate a Stochastic Block Model, an algorithm for the detection of latent communities that is used extensively in network analysis, to identify the clusters of subindustries that are consistent with the observed worker flows. The algorithm is micro-founded in Nimczik (2020) with a simple firm-choice model, where two subindustries belong to the same cluster if, for workers employed in the two subindustries, the utility cost of moving to other subindustries in the economy is identical (e.g., skill transferability costs).

[13]To focus exclusively on relatively stable relationships, I adopt the following criteria: First, a change of employer is categorized as a job-to-job transition if there is no more than six months of non-employment between two consecutive employment spells. Second, the sample is limited to transitions in which workers had a minimum tenure of one year in both their previous and current jobs. Third, I restrict the sample to transitions involving firms with two or more employees.

exercises using local labor market definitions that are more standard in the literature, such as city-industry or city-occupation combinations (e.g., Azar et al., 2020, Benmelech et al., 2022), along with other aggregations that exploit worker flows. There are 80 2-digit industries in the data and I can define 80 proxies for occupations (16 1-digit industries $\times$ 5 skill groups). Therefore, for comparability, I estimate 80 flow-based clusters out of 232 3-digit subindustries for my baseline definition of local labor markets.[14] In the remainder of this paper, I will refer to these flow-based clusters as industries.

The flexibility of this strategy means that the resulting clusters sometimes consist of a single 3-digit sector, while at other times they encompass multiple sectors. This distinguishes my approach from a more rigid local labor market definition based on 2-digit sectors. For instance, among the largest flow-based markets, there are some that only include a single 3-digit sector, such as “cleaning activities” in one market and “health care activities” in another. These are fairly broad categories: the former covers exterior and specialized building cleaning, industrial machinery cleaning, and water supply pipe cleaning; the latter includes hospital services for both short- and long-term care, as well as medical diagnosis and treatment in general hospitals.

In contrast, a 2-digit sector-based definition would group “cleaning activities” with unrelated activities like gardening services and other facility management functions, such as mail handling and reception, likely making the market overly broad. Similarly, “health care activities” would be grouped with more specialized healthcare services, like dentistry or veterinary practices, creating an arguably less accurate reflection of the relevant local labor market.

On the other hand, some large flow-based markets encompass multiple 3-digit sectors. For example, one market includes four 3-digit sectors related to food manufacturing (e.g., meat, oil, and bread production) and retail activities mainly associated with farming and agriculture. A 2-digit sector-based approach would not have merged retail and manufacturing firms into a single market, despite the substantial worker flows between these 3-digit sectors. Conversely, it would have grouped them with six other specialized food manufacturing sectors in the first case, and with six specialized retail sectors unrelated to food production or farming in the second case. Thus, the flow-based approach provides a more nuanced and accurate aggregation of subsectors, reflective of actual labor market dynamics.

[14] Local labor markets derived in the original analysis by Nimczik (2020) cover multiple cities. In my empirical exercise, instead, local labor markets are defined at the city level and thus do not encompass multiple commuting zones. Given this assumption, productivity in the local labor market is a function of a *single* city’s common productivity term (see Appendix D.3). This highly simplifies the estimation strategy of Section 3.2.

3.2 Estimation Strategy

In this section, I present an empirical strategy to estimate the part of the city-size wage premium that can be attributed to systematic differences in labor market power between small and large urban areas (Section 3.2.1). This strategy is derived from the model's equilibrium equation (4), which establishes a causal link between employment concentration and wages. The joint influence of unobserved market level productivity on monopsony power and workers' earnings introduces endogeneity concerns. To address these challenges, Section 3.2.2 presents an alternative specification that controls for local revenue productivity, whereas Section 3.2.3 describes the IV strategy.

3.2.1 Estimating Agglomeration Economies

The log version of the equilibrium equation (4) with $\tau = -\eta^{-1}$ is given by

$$\log \mathrm{W}_{mt} = \log(\mathrm{AMRPL}_{mt}) - \log\left(1 - \tau\mathrm{HHI}_{mt}\right).$$

If productivity $\log(\mathrm{AMRPL}_{mt})$ is unobserved, then we can estimate

$$\log \mathrm{W}_{mt} = \alpha_m + \alpha_t + \tau\mathrm{HHI}_{mt} + \alpha\mathrm{X}_{mt} + \varepsilon_{mt}, \tag{8}$$

where $\log\left(1 - \tau\mathrm{HHI}_{mt}\right) \simeq -\tau\mathrm{HHI}_{mt}$ because $\hat{\tau}\mathrm{HHI}_{mt}$ is estimated to be small, and X_{mt} is a vector of market observables that are relevant for wage determination but are not modelled explicitly in Section 2.

The two-way fixed effects structure $\alpha_m + \alpha_t$ controls for the part of $\log(\mathrm{AMRPL}_{mt})$ that systematically affects wages W_{mt}. As $\log(\mathrm{AMRPL}_{mt})$ also influences HHI_{mt} (e.g., shocks to market level productivity and/or to amenities that induce firm entry), these fixed effects need to be included to consistently estimate the effect of labor market power on workers' earnings.

In particular, the market fixed effect α_m in equation (8) measures the part of market level productivity that translates into higher wages. Due to the presence of *agglomeration economies* (see equation (3)), larger cities benefit from a productivity advantage. As a result, α_m should be *positively* correlated with the population of the city in which the local labor market is located. This can be tested using the following *two-step procedure*:

$$\begin{aligned} &\textbf{Step 1:} \quad \log \mathrm{W}_{mt} = \alpha_m + \alpha_t + \tau\mathrm{HHI}_{mt} + \alpha\mathrm{X}_{mt} + \varepsilon_{mt}, \\ &\textbf{Step 2:} \quad \hat{\alpha}_m = \alpha_k + \delta\mathrm{logCitySize}_c + \upsilon_m. \end{aligned}$$

Step one corresponds to equation (8). *Step two* takes estimates of α_m coming from *step one* and regresses them on city size controlling for industry fixed effects α_k. Notice that δ is the *agglomeration elasticity* introduced in equation (3) and that population size is used to proxy for city employment $\sum_{k=1}^{K} L_{ck}$. In Appendix D.3, I list the assumptions that $\log(\text{AMRPL}_{mt})$ needs to obey so that δ can be identified with this strategy.

De la Roca and Puga (2017) estimate this agglomeration elasticity with a similar procedure and using the same Spanish administrative data. They emphasize that the estimation of coefficient δ in *step two* is subject to endogeneity concerns because, for example, highly productive cities encourage workers' migration (reverse causality bias). As in their analysis, in Section 4.3 I deal with this endogeneity issue with an IV for city size based on historical determinants of population, plausibly unrelated to current productivity.

The crucial difference from De la Roca and Puga (2017) is that in their analysis, labor markets are assumed to be perfectly competitive ($\text{HHI}_{mt} = 0$). If labor market power is relevant and systematically related to city size, ignoring HHI_{mt} in *step one* leads to estimate a biased *agglomeration elasticity*. In this case, lower wages in smaller cities will be entirely attributed to lower productivity levels in those markets and not to possibly higher levels of labor market power.

Let $\hat{\delta}^{pc}$ be the estimate of the potentially biased agglomeration elasticity. The formula for the relative extent of the bias,

$$\frac{\hat{\delta} - \hat{\delta}^{pc}}{\hat{\delta}^{pc}}, \tag{9}$$

is provided in Appendix D.3. In Section D.3.1, I show that the bias disappears if $\tau = 0$ (i.e., if employment concentration has no effect on wages) and/or HHI_{mt} is uncorrelated with logCitySize_c. Equation (9) can be interpreted as the fraction of the city-size wage premium explained by labor market power.

Finally, Appendix D.3.2 shows that city amenities can further bias the agglomeration elasticity δ. This occurs if the level of amenities is correlated with city size. Suppose, for instance, that amenities are, on average, lower in large urban areas (e.g., because of lower air quality). Part of the urban earnings premium may then act as compensation for individuals to live and work in larger cities despite the higher disamenity levels, while being totally unrelated to agglomeration economies. As shown in Appendix D.3.2, the bias disappears if an additional control for city amenities is introduced in *step two*.[15]

[15] As explained in Appendix D.3.2, the bias increases with the degree of decreasing returns to scale in the economy. With a higher degree of decreasing returns to scale, indeed, a positive supply (amenity) shock leads to lower average productivity in the market. This happens because firms can now hire more workers for the same wage, and these workers are marginally less productive. Since lower productivity in the market directly translates into lower wages, amenity differences between small and big cities have a higher potential to explain the heterogeneity in earnings observed in the data, and failing to account for urban amenities will lead to a higher bias in the estimated agglomeration elasticity. In the opposite extreme case of constant returns to scale technology, amenities have no effect on wages and so

3.2.2 Estimation With Market Revenue Productivity Control and Heterogeneous Linear Trends

We have emphasized that endogeneity in the relationship between HHI_{mt} and W_{mt} is ultimately due to the market level productivity AMRPL_{mt}, which has been treated until now as an unobserved variable. To mitigate concerns about endogeneity, we can control for a proxy of AMRPL_{mt} in *step one* of the estimation procedure described in Section 3.2.1, alongside the two-way fixed effects structure. Saturating the regression with additional city-year and industry-year fixed effects, as well as market-level linear trends, further alleviates endogeneity concerns.

As it is shown in Appendix D.1.1, given the Cobb-Douglas production function assumption, AMRPL_{mt} can be rewritten as the employment share weighted average of each firm's revenues per worker, i.e.

$$\widetilde{\text{AMRPL}}_{mt} = \theta \sum_{f=1}^{N_{mt}} s_{ft} \frac{P_{ft} Q_{ft}}{l_{ft}},$$

a quantity observed in the data.[16] $\widetilde{\text{AMRPL}}_{mt}$ approximates AMRPL_{mt}, but may differ from it because of measurement error and/or because the production function is not Cobb-Douglas. Adding the $\widetilde{\text{AMRPL}}_{mt}$ control to equation (8) allows us to account for variations in market productivity that may not be fully captured by fixed effects α_m and α_t. To further partial out any endogenous variation in HHI_{mt}, we add city-year α_{ct} and industry-year α_{kt} fixed effects, along with market-specific linear trends t with coefficients γ_m.

To compute the agglomeration elasticity δ in this context, the two-step procedure in Section 3.2.1 is slightly modified as follows:

Step 0: $\log(W_{mt}) = \alpha_m + \alpha_t + \alpha_{ct} + \alpha_{kt} + \gamma_m t + \alpha_1 \log(\widetilde{\text{AMRPL}}_{mt}) + \alpha \text{X}_{mt} + \tau \text{HHI}_{mt} + \epsilon_{mt},$

Step 1: $\log(W_{mt}) - \hat{\tau}\text{HHI}_{mt} = \alpha_m + \alpha_t + \alpha \text{X}_{mt} + \varepsilon_{mt},$

Step 2: $\hat{\alpha}_m = \alpha_k + \delta \text{logCitySize}_c + \upsilon_m.$

First, the coefficient τ is estimated in a preliminary step that augments equation (8) by inserting the productivity proxy $\widetilde{\text{AMRPL}}_{mt}$, fixed effects α_{ct} and α_{kt}, and heterogenous linear trends $\gamma_m t$ as additional controls. The τ estimate obtained from this regression can be used to partial out the effect of labor market power, $\hat{\tau}\text{HHI}_{mt}$, from wages W_{mt}. The partialled out wages are then used as dependent variable in *step one* of the procedure to obtain estimates of productivity, captured by market fixed effects

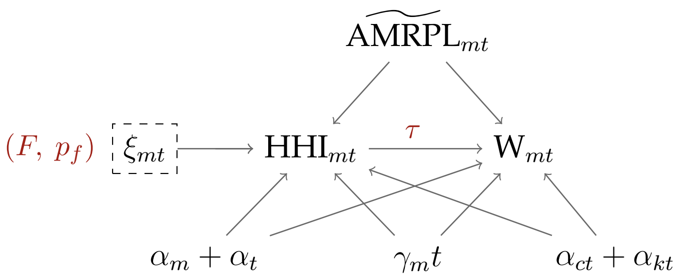
cannot bias the agglomeration elasticity.

[16] I do not observe the degree of decreasing returns to scale θ but, to the extent that the parameter is constant across local labor markets, this is irrelevant for estimation.

α_m, which are not biased by the influence of labor market power on workers' earnings. Finally, *step two* identifies the agglomeration elasticity by regressing the market fixed effects of *step one* on log city size.

The DAG depicted in Figure 5 highlights the source of variation in HHI_{mt} which identifies the coefficient τ in the preliminary step (*step zero*) of the strategy. The model accounts for variation in HHI_{mt} and W_{mt} originating from the flexibile set of fixed effects, heterogenous linear trends, and the observed productivity proxy $\widetilde{\text{AMRPL}}_{mt}$, so that identification is plausibly driven only by exogenous sources ξ_{mt}.

Figure 5: A DAG summary of the model with fixed effects, heterogenous linear trends, and controlling for market revenue productivity ($\widetilde{\text{AMRPL}}_{mt}$)



Note: This figure draws a directed acyclic graph (DAG) of the model with fixed effects, heterogenous linear trends, and controlling for market revenue productivity. The variables are enclosed within dashed lines if they are unobservable. Arrows represent causal relationship between the variables in the data generating process.

3.2.3 Identification with IV

The estimation procedure described in Section 3.2.1 can be alternatively carried out using an instrumental variable strategy. Quasi-experimental variations in labor market power, unrelated to the productivity process causing endogeneity concerns, can indeed be used to identify the effect of monopsony power on earnings, and hence, to estimate the unbiased agglomeration elasticity.

In the model, changes in the local size of the public sector constitute a valid IV for HHI_{mt} because they are unrelated to shocks to AMRPL_{mt} and have an impact on labor market power. If exogeneity of the instrument holds, then the IV coefficient for HHI_{mt}, τ_{IV}, is well-identified. The δ elasticity can then be estimated as in Section 3.2.2, by first subtracting $\hat{\tau}_{\text{IV}}\text{HHI}_{mt}$ from W_{mt} and then using the partialled out wages as dependent variable for the two-step procedure. The DAG depicted in Appendix Figure A1 shows how an instrument Z_{mt} based on the local size of the public sector can identify coefficient τ_{IV}, sidestepping the endogeneity concerns introduced by unobserved productivity AMRPL_{mt}.

4 Results

I now present the results for the *two-step* empirical strategy outlined in Section 3, which I use to estimate the productivity advantage of big cities in the presence of labor market power. After describing the data (Section 4.1) and the labor market controls (Section 4.2), the results are shown in Sections 4.3 and 4.4. To address endogeneity concerns, I first use a two-way fixed effects structure and then further control for a proxy of market level productivity and heterogeneous linear trends. As a complementary and independent procedure, I also use an instrumental variable estimation strategy (Section 4.5). Plausibly exogenous variation in labor market concentration arises from changes in the size of local public firms in health- and education-related markets. In line with the exogeneity assumption, I show that the instrument is unrelated to local revenue productivity.

4.1 Data

The main dataset used in the analysis is Spain's Continuous Sample of Employment Histories (*Muestra Continua de Vidas Laborales* or MCVL). This is a matched employer-employee panel for a 4% non-stratified random sample of individuals affiliated to the Spanish Social Security in 2005-2019, obtained by combining administrative data, income tax, and census records. The MCVL allows us to track workers across space based on their work location. Using this information together with each employer's 3-digit industry, individuals can be assigned to their corresponding local labor market.

Data on employees' daily working hours is also provided in the sample. With this information, we can compute market level mean annual wages W_{mt} expressed as euros per day of full-time equivalent work. Earnings in the MCVL come from tax return data and are not subject to censoring. Information on wages and other workers' observables are provided for the entire working life of the sampled individuals, when available. We focus on 2005–2019, the period in which job spells are matched with tax record data that provide uncensored earnings. Only workers employed in the private sector are considered when computing W_{mt}, as public wages tend to be more regulated and are less likely to respond to labor market concentration.

Employment concentration in the market, HHI_{mt}, is also computed using this dataset. The time series of the Herfindahl-Hirschman Index computed with the MCVL closely follows the analogous time series computed independently with data on the universe of Spanish firms from the Spanish Statistical Office (INE), as shown in Appendix Figure A2. This is checked at the region-sector(2-digit) level, the most granular unit of analysis for local labor markets in the INE data. Information from the MCVL can be used to accurately capture the evolution of HHI over time because (i) labor market concentration is mainly affected by the employment dynamics of big firms, and (ii) employed individuals in the panel are much more likely to be sampled from large establishments, as the sample is random across workers. In particular, approximately 90% of all Spanish establishments with more than ten workers are covered in the MCVL, which ensures a high representativity level.[17] Even if public wages are excluded from the earnings variable W_{mt}, vacancies in public firms still constitute relevant outside options for private employees in many Spanish local labor markets. Therefore, HHI_{mt} is computed taking into account both private *and* public firms operating in each market-year.

The MCVL is also used to measure workers' experience, years of tenure, education (binary indicators for below-secondary, secondary, and tertiary education), and contract type (temporary or permanent), in addition to their gender and nationality. Additionally, each worker is assigned to one of ten occupation categories listed in the social security system, which are meant to capture specific skills required by the job. Following De la Roca and Puga (2017), these categories have been grouped into five skill levels, from low-skiled to very high-skilled.[18] Furthermore, aggregate job-to-job transitions of sampled individuals across 3-digit NACE sectors are used to estimate industry clusters that are linked by worker flows, which form the basis for the definition of local labor markets. Finally, information on the local unemployment rate at the market level is recovered from the sample. Further information on the dataset is provided in Appendix C.1.

One key piece of information missing in the MCVL is firm level production data. To compute the market level productivity proxy, $\widetilde{\text{AMRPL}}_{mt}$, and to control for concentration in the product market, Sales HHI_{mt}, data on firms' revenues need to be used. Therefore, I exploit balance sheet information for the quasi-universe (82%) of Spanish firms during the years 2005-2019 obtained from the *Banco de España Data Laboratory* (BELab). Crucially for our purposes, this data source provides information on firms' headquarters location and their NACE sector code, which can be mapped to my local labor market definition. Using yearly information on sales and employment provided in the sample, I can compute the market level Sales HHI_{mt} and $\widetilde{\text{AMRPL}}_{mt}$ variables for the entire period of analysis.

[17]To compute the Herfindahl-Hirschman Index, I use the number of individuals sampled in the MCVL as a proxy for each establishment's employment level. The number of employees in the sample is computed for each month and then averaged at the yearly level to compute the HHI time series for each local labor market. The MCVL does contain information on the real number of employees in each establishment of the dataset, although the number refers to April of the following year. We can also use this information, lagged by one year, to compute the HHI. Although the two methods yield similar results, the Herfindahl-Hirschman Index computed using sampled workers as a proxy for employment follows more closely the time series of the HHI measured with independent INE data (see Appendix Figure A2). Therefore, this is the preferred method of choice for the analysis.

[18]For example, the upper contribution group, which includes very high-skilled occupations, is reserved for jobs that require an engineering or bachelor's degree and for top managerial positions.

Furthermore, I obtain the coverage of collective agreements, which proxies for the influence of unions, from the Spanish Ministry of Labor and Social Economy. Additionally, the share of production exported is computed using data from the Spanish Ministry of Industry, Trade, and Tourism (*DataComex*).

Finally, local labor markets are defined as combinations of 76 *urban areas* and of clusters of 3-digits NACE subindustries which are based on worker flows (Section 3.1). I use official definitions of urban areas constructed by Spain's Ministry of Housing in 2008. Urban areas group municipalities linked by commuting and employment patterns. They cover 68% of Spain's population and 10% of its surface area. As in De la Roca and Puga (2017), the population size of urban areas is given by the number of people within 10 km of the average person in the city, which they compute on the basis of a 1-km population grid for the year 2006 created by Goerlich and Cantarino (2013). The advantage of this measure over plain population density, a popular choice in the related literature measuring the productivity advantages of large cities (e.g., Combes et al., 2010), is that it is less subject to the noise introduced by the fact that municipality boundaries may be arbitrarily drawn and may enclose large uninhabited areas.

4.2 Market-Level Controls

Before presenting the results, this section briefly describes the X_{mt} vector of variables that is used as control in *step one* of the estimation procedure, equation (8).[19]

Sorting of higher skilled workers into bigger cities could explain part of the city-size wage premium. Because we do not want sorting to bias the agglomeration elasticity, we control for workers' observables in equation (8). Market-year level mean experience and tenure as well as education, skill level (as described in Section 4.1), contract type (permanent or temporary), gender, and native shares are included as controls.

Product market power may also bias the estimation, since it could affect wages (Nickell et al., 1994), and it is likely correlated with labor market power and city size. We control for it with the Sales Herfindahl-Hirschman Index (Sales HHI_{mt}), which is defined as the HHI_{mt} of equation (5), with the difference that revenues shares are used instead of employment shares. Controlling for market *revenue* productivity also helps to account for oligopoly power in the goods market.

Unions may play an important role in determining wages when labor markets are imperfectly competitive, by limiting the monopsony power of employers (Azkarate-Askasua and Zerecero, 2022). The region-sector(1-digit)-year level coverage of collective agreements is used as a proxy for the importance of unions in the market.[20]

[19]Despite the X_{mt} subscript, not all controls vary at the market-year level – as it is made clear below.

[20]In a study on the Spanish economy, Arellano et al. (2002) argue that union affiliation is relatively low in Spain and that the coverage of collective agreements is a better proxy for the impact of unions on wage determination.

The *local unemployment rate* may also be related to HHI; e.g., a concentrated local labor market tends to have a higher local unemployment rate, which puts additional downward pressure on wages. Using the matched employer-employee data, the unemployment rate can be first computed at the city-year level, and then further attributed to the local labor market level by using information on the last industry where unemployed workers used to work before losing their job.

Finally, the *exporter status* of firms may matter, as exporting firms' rents could differ from those of non-exporting firms due to selection and product market competition effects (Bernard and Jensen, 1999). Additionally, the product market of firms in non-traded sectors is geographically limited to the urban areas in which they operate, which could affect their revenue productivity and, hence, the wage they offer. As a proxy for both the exporter status of firms and the tradability of their final products, the sector (2-digit) year level share of production devoted to exports is inserted as control in equation (8).

4.3 OLS Results

Results for the *two-step* procedure

$$\textbf{Step 1:} \quad \log \mathrm{W}_{mt} = \alpha_m + \alpha_t + \alpha_{m1} \times \alpha_{t1} + \tau \mathrm{HHI}_{mt} + \alpha \mathrm{X}_{mt} + \varepsilon_{mt},$$
$$\textbf{Step 2:} \quad \hat{\alpha}_m = \alpha_k + \delta^{\mathrm{HHI}} \mathrm{logCitySize}_c + \upsilon_m,$$

with and without controlling for HHI_{mt} in *step one*, are presented in Table 1.

Columns (1), (2), (3), and (4) report estimates from the *step one* regression, progressively adding market and year fixed effects, city-year and industry-year fixed effects, market-level linear trends, and finally including all of these controls. The coefficient of HHI is comparable across specifications, and shows that labor market concentration is associated with lower earnings. In the baseline specification with two-way fixed effects, column (1), moving from $\mathrm{HHI}_{mt} = 0$ (fully unconcentrated markets) to $\mathrm{HHI}_{mt} = 1$ (single monopsonist) is associated with a decrease in mean wages of approximately 7.4%.

When labor market concentration is not controlled for, regressing the market fixed effects of *step one* against log city size yields an agglomeration elasticity estimate of 0.089 (column (5)). This elasticity is reduced to 0.075 when we control for HHI_{mt} in *step one* (column (6)) using the two-way fixed effects model as the baseline. Because labor market power is negatively correlated with city size and has a sizeable impact on wages, failing to account for differences in monopsony power between small and large urban areas biases the agglomeration elasticity upward. This occurs because lower wages in smaller cities are entirely attributed to lower productivity levels in those markets and not to higher levels of labor market power. Therefore, not controlling for HHI in *step one* leads to lower estimates of the market fixed effects in small urban areas and, hence, to a higher estimated agglomeration elasticity (see Appendix Figure A3). By computing the relative extent of the bias using formula (9), we conclude that labor market power accounts for approximately 16% of the city-size wage premium.

The agglomeration elasticity estimates in columns (5) and (6) of Table 1 suffer from endogeneity concerns. On the one hand, reverse causality issues arise as high wages offered in productive cities attract migrants, which increases city size. On the other hand, omitted variable bias may originate from unobserved city characteristics that jointly increase wages and attract worker migration.

Table 1: OLS estimates

	Step 1: log W				**Step 2:** $\hat{\alpha}_m$	
	(1)	(2)	(3)	(4)	(5)	(6)
Log City Size					0.0888***	0.0745***
					(0.0058)	(0.0054)
HHI	-0.0738***	-0.0613***	-0.0594***	-0.0569***		
	(0.0141)	(0.0200)	(0.0146)	(0.0144)		
Sales HHI	-0.0101	-0.0050	-0.0062	-0.0054		
	(0.0072)	(0.0074)	(0.0069)	(0.0071)		
Labor Market Controls	✓	✓	✓	✓		
Market-Level Linear Trends			✓	✓		
City-Year, Industry-Year FE		✓		✓		
Year FE	✓	✓	✓	✓		
Market FE	✓	✓	✓	✓		
Industry FE	Absorbed	Absorbed	Absorbed	Absorbed	✓	✓
R^2	0.85	0.86	0.90	0.91	0.38	0.38
Observations	64,246	64,246	64,246	64,246	5,027	5,027
Step 1 with HHI	✓	✓	✓	✓		✓

Note: This table reports estimates of Step 1 regressions (columns (1)-(4)) and of Step 2 regressions when HHI is included in Step 1 (columns (6)) or not included (column (5)), in line with the procedure presented in Section 3.2.1. Labor market controls include average worker experience and tenure years, share of workers with high school and university education levels, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract type shares (temporary or permanent), share of Spanish native citizens, share of male workers, and share of exported revenue. The market fixed effects used as dependent variable in column (6) are estimated in column (1). Standard errors are clustered at the market level in columns (1)-(4), and at the industry level in columns (5) and (6). *p<0.1, **p<0.05, ***p<0.01.

To address these concerns, I use an IV based on the historical determinants of population for the city size variable. As in De la Roca and Puga (2017), the variables used to instrument log city size are historical population figures for 1900, historical transportation networks (number of roman roads within 25 km from the city center), and

geographical variables that likely influenced early settlement patterns but are arguably uncorrelated with current productivity levels (i.e. land fertility, water availability, terrain slope, and elevation).[21] As Appendix Table B1 shows, the instruments are jointly and individually significant. The agglomeration elasticity estimates are slightly larger when we use this IV (see Appendix Table B2). Nonetheless, the relative extent of the bias coming from the omission of labor market power controls in Step 1 is very similar at 14%. This is in line with the fact that endogeneity of city size does not constitute a significant threat to identification in this type of analysis (Combes et al., 2010).
Finally, in Appendix Tables B3 and B4 we compare the previously estimated agglomeration elasticities with those obtained by controlling for a set of city amenities in *step two*. Natural amenities include precipitation, distance from the coast, mean temperature, and the percentage of land with water within 25 km of the city center (Table B3). Along with these plausibly exogenous amenities, other relevant urban amenities that are endogenous to city size are also included in Table B4: pollution (NO2 concentration levels), mean commuting time, crimes per person and cinemas per person. The estimate of the productivity advantage of large cities remains virtually unchanged when exogenous amenities are considered. However, the estimated agglomeration elasticity decreases when we also account for endogenous city amenities (column (2) of Table B4). This is because this set of amenities tends to be negatively correlated with city size (e.g., larger cities are more polluted and have more crimes), and workers want to be compensated more to live in cities with low amenities.[22] The wage premium offered in larger urban areas can then be partly explained as compensation for the disamenities arising from living in dense cities and is not entirely attributed to higher productivity levels.

However, the extent of the agglomeration elasticity bias due to the omission of labor market power, computed using formula (9), is estimated to be similar to the one previously obtained in Table 1, when city amenities were not taken into account. Differences in the degree of imperfect competition in the labor market between cities of varying sizes account for approximately 16-18% of the city-size wage premium, depending on whether only exogenous or both endogenous and exogenous amenities are considered.

[21] Details on measurement and on historical and geographical data sources, which include Goerlich and Azagra (2006) and McCormick et al. (2008), can be found in De la Roca and Puga (2017).

[22] The estimates for the agglomeration elasticity are reduced to 0.0636 and 0.0795, in case the HHI_{mt} control is, respectively, inserted or not inserted in *step one*.

4.4 Robustness

4.4.1 Revenue Productivity, Local Unemployment, Recession Years, and New Hires

In Appendix Table B5, I estimate the usual *two-step* procedure and additionally control for revenue productivity ($\widetilde{\text{AMRPL}}_{mt}$) in *step one*, as described in Section 3.2.2. From columns (1) to (5), I progressively control for an increasingly flexible set of fixed effects: city, industry, and year; market and year; market, city-year, and industry-year; and market, city-year, and industry-year fixed effects with market-specific linear trends. The measured productivity log($\widetilde{\text{AMRPL}}$) positively affects log wages, but its associated coefficient shrinks as we saturate the regression with fixed effects. However, the coefficients of HHI are virtually the same as those estimated in the baseline regressions in Table 1. This finding suggests that, as highlighted in Section 3.2.2, the flexible fixed effects structure captures the variation in log(AMRPL) and allows us to estimate the effect of HHI_{mt} on W_{mt} in an arguably consistent manner.

As additional robustness checks, I control for the local unemployment rate at the market level and exclude recession years – specifically, 2008-2009 and 2011-2013. Comparing columns (2) and (3) of Appendix Table B6 with column (1), the HHI estimates remain largely unchanged across these alternative specifications. Consequently, the resulting agglomeration elasticity is nearly identical to that of the baseline. Labor market power explains approximately 16% of the city-size wage premium.

Finally, I test whether the results hold when restricting the sample to new hires and calculating HHI specifically for this group. To make the robustness check operational, I limit the sample to workers who have recently changed establishments. Specifically, I include only the months where these workers were employed at a different establishment in the previous month. This reduces the sample size to 3.4% of its original size. For consistency, I compute HHI, market controls from administrative data, and market wages using this subset of new hires. However, certain market controls $-$ namely, the sales HHI and the export share of revenue $-$ cannot be restricted to new hires, while union coverage is not limited to new hires due to data constraints. For these variables, I use the baseline values from the full sample.

As shown in column (4) of Table B6, the coefficient for HHI among new hires (-0.0685) is statistically significant and is comparable to the baseline regression in column (1) (-0.0738). When the sample is restricted to new hires, concentration accounts for 27% of the city-size wage premium, exceeding the baseline OLS effect of 16.1%. This result is largely driven by the fact that new hires tend to be more concentrated among fewer firms in smaller cities, causing concentration to vary more with city size and thus contributing more to the explanation of the urban wage premium. We should however interpret these findings cautiously, as the small sample size may introduce noise into the results.

4.4.2 Alternative Definitions of Local Labor Markets

Instead of relying solely on the baseline definition of markets based on worker flows across 3-digit sectors, I explore alternative market definitions. First, I consider more standard approaches, i.e. city-industry and city-occupation combinations (e.g., Azar et al., 2020, Benmelech et al., 2022). I start by using 80 2-digit industries, which I interact with urban areas. This results in a number of markets comparable to the baseline definition, that uses 80 clusters within each urban area. The HHI coefficient under this alternative definition is -0.0521, compared to -0.0738 in the baseline (see column (5) of Table B7). This implies that labor market concentration accounts for 12.2% of the city-size wage premium.

Next, I explore an alternative definition based on occupations, using skill information. In Spain, workers are classified into different social security contribution groups according to the educational requirements of their roles, which is recorded administratively. Following the methodology from De la Roca and Puga (2017), I classify these groups into five skill levels: *very high-skilled*, *high-skilled*, *medium-high-skilled*, *medium-low-skilled*, and *low-skilled* occupations. For instance, the top contribution group, which includes very high-skilled occupations, is reserved for jobs that require an engineering or bachelor's degree and for top managerial positions.

To proxy for occupations, I interact 1-digit sectors with these five skill categories. With 16 1-digit sectors, this interaction creates 80 unique cells, aligning the number of local labor markets with the baseline definition (80 clusters). The findings are reported in column (6) of Table B7: the HHI coefficient is -0.0867, compared to -0.0738 in the baseline. This implies that 18.2% of the city-size wage premium can be attributed to variations in labor market concentration across cities, up from 16.1% in the baseline model. Given the greater flexibility of my baseline definition based on worker flows, I prefer it over these alternative, though widely used, definitions.

Furthermore, I can incorporate information on skills in my flow-based definition of local labor markets. First, I categorize workers into cells based on a combination of 3-digit sectors and skill groups, resulting in $232 \times 5 = 1160$ distinct cells. I then calculate worker flows among these cells and apply the classification algorithm from Nimczik (2020) to group cells into 80 clusters based on the worker flows. This method potentially provides a more precise representation of occupations by integrating skill information, relative to the baseline definition that only considers flows across 3-digit sectors. The results with this new labor market definition are shown in column (2). Here, the HHI coefficient is -0.0755, very similar to the baseline model. This implies that 11.9% of the city-size wage premium can be attributed to labor market power.

It should be noted that some market controls – namely the sales HHI, the share of revenue exported, and union coverage – cannot be computed at the 3-digit sector $\times$ skill level. For these variables, I have used baseline values computed at the broader sectoral level and merged them into the main dataset. Similarly, market-level revenue productivity (AMRPL), which is crucial to motivate the exogeneity assumption of the IV (Section 4.5), is only available at the 3-digit sector level and cannot be perfectly mapped to this alternative local labor market definition. Due to these limitations, the baseline definition using flows across 3-digit sectors remains preferred, though the alternative definition incorporating skill data produces comparable results.

The baseline definition relies on labor flows which are endogenous to wages and influenced by labor market power, making the classification of local labor markets endogenous as well. As a further robustness check, I then define local labor markets using flows from 1970 to 2004 $-$ before the analysis period (2005-2019) $-$ which are not affected by the treatment. This is feasible due to the structure of the administrative data (MCVL), which tracks the complete employment histories of sampled workers. The results, presented in column (3) of Table B7, are very close to the baseline: the HHI coefficient is -0.0892, with labor market concentration explaining 18.4% of the city-size wage premium, compared to 16.1% in the baseline. Since the endogeneity of labor market definitions does not appear to significantly affect the results, I prefer to retain the baseline definition (2005-2019 worker flows). This definition is more up-to-date and better reflects the current dynamics of local labor markets.

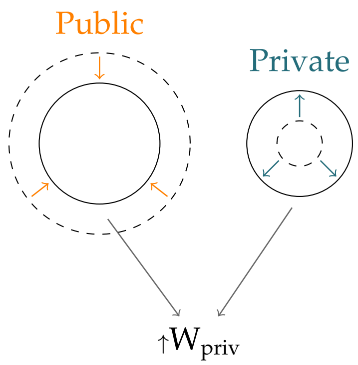
Finally, I consider the possibility of different local labor markets between cities of different size. Specifically, I apply the classification algorithm from Nimczik (2020) to worker flows between 3-digit sectors, treating cities of different sizes separately. I divide cities into quintiles based on size and implement the classification strategy within each of these five groups, ensuring that 80 clusters are estimated within each city group to maintain comparability with the baseline. The results for this alternative local labor market definition are shown in column (4) of Table B7. Here, the HHI coefficient is -0.0669, and 10.1% of the city-size wage premium can be explained, compared to -0.0738 and 16.1% in the baseline regression. While this method offers greater flexibility in defining local labor markets, it comes at the cost of reduced precision, as fewer worker flows within each city quintile make the estimation more noisy.

4.5 IV with Changes in Size of Public Sector

Next, I use an IV strategy to estimate the agglomeration elasticity. In particular, I instrument the Herfindahl-Hirschman Index with changes in the local size of the public sector, following the procedure described in Appendix Section 3.2.3. While Guillouzouic et al. (2022) is a recent paper that highlights that public firms tend to be relatively large, and so are likely to exert substantial monopsony power on workers, I am aware of no study in the literature that directly uses changes in the size of local public firms as an instrument for HHI.[23]

To fix ideas, imagine a labor market for nurses in a small city, where the only employers operating are a large public hospital and a small private hospital. After a year, the public hospital shrinks for reasons unrelated to local economic conditions (e.g., a regional election leading to a change in governmental policies at a higher administrative level), whereas the private hospital increases in size for reasons that are potentially endogenous to the business cycle. This stylized example is illustrated in Figure 6.

Figure 6: A local labor market with two firms



Note: This figure draws a stylized example of a local labor market with a big public firm and a small private firm. The public firm shrinks while the private firm increases in size, which reduces labor market power and puts upward pressure on private wages.

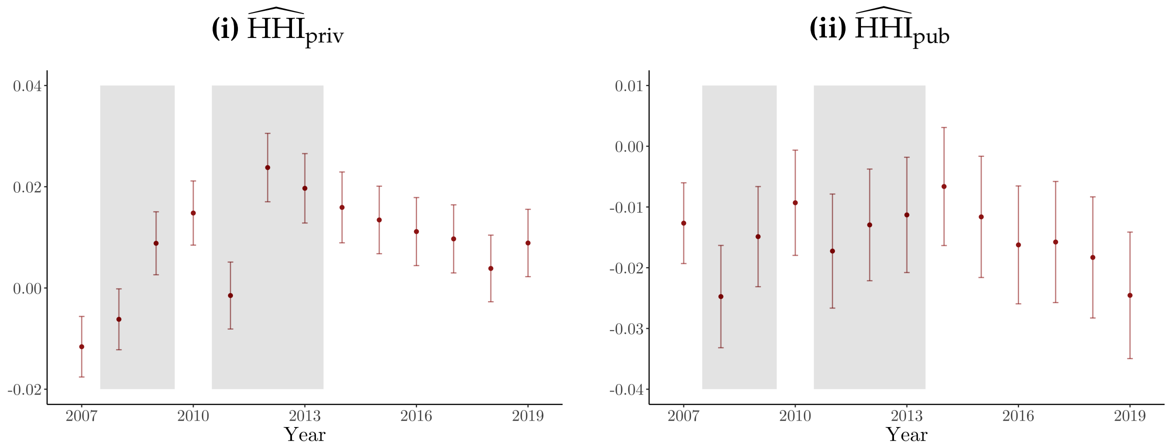
Prior to changes in the size of the two firms in the local labor market, nurses seeking employment were likely to find open positions only in the much larger public hospital, which could act as a de facto monopsonist and pay workers less than their marginal product. The private hospital could take this market wage as given, having no reason to pay nurses more than the public hospital. After the employment changes have substantially reduced the firms' size gap, however, the private hospital has become a relevant competitor employer. The resulting increase in labor market competition should, all else equal, push nurses' wages up – as employees with more outside options have some bargaining power to turn down bad offers. The proposed IV strategy only exploits the (plausibly exogenous) variation in HHI due to changes in

[23] Arnold (2022), Benmelech et al. (2022), and Prager and Schmitt (2021) exploit mergers and acquisitions events as a source of exogenous variation in labor market concentration. Yet, these events are likely to be partly driven by local economic conditions that may contemporaneously affect earnings. To alleviate concerns about the validity of the IV, the authors either control for labor productivity or perform a series of robustness checks. Schubert et al. (2024) pursue an alternative strategy by using shift-share shocks based on national firms' hiring growth as IV.

the size of the public hospital while focusing on the wages offered by the *private* firm as the relevant outcome.[24]

I construct the IV by computing the mechanical impact that the employment changes of local *public* firms have on HHI. This instrument estimates the evolution of the Herfindahl-Hirschman Index in each period had only public firms altered their employment levels as recorded in the data, disregarding the impact of private firms. The construction of the IV, denoted by $\widehat{\text{HHI}}_{\text{pub}}$, is detailed in Appendix C.2. Similarly, the impact on HHI caused by changes in the employment of local *private* firms only is denoted by $\widehat{\text{HHI}}_{\text{priv}}$ (see Section C.2). Figure 7 shows how $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$ evolve over time.

Figure 7



Note: This figure plots the evolution of $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$ over time. $\widehat{\text{HHI}}_{\text{priv}}$ denotes changes in HHI coming from *private* firms, whereas $\widehat{\text{HHI}}_{\text{pub}}$ denotes changes in HHI coming from *public* firms. Point estimates and standard errors are year fixed effects of two separate regressions where market fixed effects and all sectors are included. The year 2005 is excluded because the variables $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$ are defined in changes with respect to the previous year, whereas the year 2006 is the excluded fixed effect. Recession years are highlighted in grey. The quarterly periods of recession in Spain were 2008Q2-2009Q4 and 2010Q4-2013Q2 (Source: *Spanish Business Cycle Dating Committee*, *Spanish Economic Association*).

It can be seen that the evolution of $\widehat{\text{HHI}}_{\text{priv}}$ has a clear business cycle component: the negative productivity shocks that come with recessions lead to firms' exit and hurt small establishments more than larger ones, which leads, with some lag, to an increase in employment concentration. The opposite occurs during periods of expansion, and $\widehat{\text{HHI}}_{\text{priv}}$ tends to decrease as a consequence. The evolution of $\widehat{\text{HHI}}_{\text{pub}}$ is much less related to the business cycle, which lends credibility to the instrument's exogeneity

[24]In both the OLS and IV regressions, public wages are not considered because they are likely more rigid in the short-run and should respond less to variations in HHI. In addition, compensation in the public sector may be directly affected by the IV (i.e., a change in policy at the regional level affecting public employment *and* wages), which would violate exogeneity of the instrument. In the rest of the section, I estimate a battery of related regressions that lend credibility to the exogeneity assumption. Among these robustness exercises, I show that the effect of the instrument on public wages is not significant.

assumption.

As described in the DAG of Appendix Figure A1, exogeneity of the IV hinges on two assumptions. First, changes in the size of local public firms in some market-year must not be related to the productivity AMRPL_{mt} of establishments operating in the same market, conditional on market and year fixed effects. Second, the only relevant (short-run) consequence of a change in $\widehat{\text{HHI}}_{\text{pub}}$ for the wage-setting behavior of establishments in the market, conditional on market and year FE, is that it affects HHI_{mt} in the local labor market.

To assess whether changes in public employment are correlated with shocks to local economic conditions that contemporaneously affect earnings W_{mt}, I check whether revenue productivity ($\widetilde{\text{AMRPL}}_{mt}$) is a statistically significant predictor of the instrument $\widehat{\text{HHI}}_{\text{pub}}$. Results are reported in Appendix Table B8.[25] Productivity is negatively related to $\widehat{\text{HHI}}_{\text{priv}}$, which is in line with the plot of Figure 7 panel (i). Although the effect has a lower statistical significance level, productivity is also positively related to $\widehat{\text{HHI}}_{\text{pub}}$, which raises concerns about the instrument's validity. Therefore, I restrict the attention to *health*- and *education*-related markets to isolate a set of industries in which movements in local public employment are less likely to be related to the business cycle. Indeed, as Columns (2) and (4) of Table B8 show, revenue productivity in these markets is not significantly related to either $\widehat{\text{HHI}}_{\text{priv}}$ or, importantly, $\widehat{\text{HHI}}_{\text{pub}}$. In the rest of the IV analysis, I restrict my attention to this set of health and education related industries and claim that the instrument is exogenous conditional on market and year fixed effects.[26] As a further check, I find that changes in log market productivity do not predict log changes in the IV, both across all markets and within the health and education sectors. This is shown in columns (3) and (4) of Table B9, respectively. Notably, this relationship holds even without accounting for market observables and fixed effects. In contrast, as expected, positive productivity shocks reduce the contribution of private firms to the HHI (see column (1)), possibly due to increased entry in response to these shocks, which leads to lower market concentration.

Approximately 60% of the total public employment in my sample are in the health and education sectors, and around 55% of workers in these markets are employed by public firms. Hence, these sectors cover a significant share of public employment. In

[25]Because the dependent variables $\widehat{\text{HHI}}_{\text{pub}}$ and $\widehat{\text{HHI}}_{\text{priv}}$ are vectors of numbers between 0 and 1, I estimate a set of logit regressions.

[26]The markets include the following subindustries: "Medical and dental activities", "Hospital activities", "Social service activities for the elderly and the disabled", "Assistance in residential facilities for the elderly and the disabled", "Assistance in residential care facilities with health care", "Residential care activities for persons with intellectual disabilities, mental illness and drug addiction", "Other residential care activities", "Other social work activities", "Other health-related activities", "Pre-primary education", "Primary education", "Secondary education", "Postsecundary education", "Research activities" ,"Research and development in Social Sciences and Humanities", "Auxiliary activities to education", "Other educational activities".

Spain, the public recruitment process involves a national exam, held separately for the education and healthcare sectors. Candidates compete to work in one of the country's 17 autonomous regions, each of which is responsible for managing its own vacancies through the *Oferta de Empleo Público* (OEP). Retirement is the most significant factor influencing employment reductions in the public sector. When employees retire, their positions may be refilled, but the nature of public exams may lead this process to be neither immediate nor one-to-one.

Vacancies tend to be announced in advance, but a lag typically exists between their publication and when they are filled. For example, vacancies for teaching positions are often published a year before the competitive public exams are held. Additionally, recruitment for primary and secondary school teachers occurs in alternating years. Therefore, even if the availability of public vacancies were related to local economic conditions, the timing of this recruitment cycle introduces a degree of exogeneity, which may explain why the instrumental variable appears uncorrelated with productivity levels or their log changes (as shown in columns (4) of Tables B8 and B9).

The IV strategy hinges on the assumption that firms in the private and public sectors belonging to the same industry and city are part of the same local labor market. Our sample shows that worker flows between private and public firms in health and education related markets are indeed sizeable: among workers that change jobs within markets, approximately 10% switch from the private to the public sector or vice-versa. If job-to-job flows are not restricted to be within markets, the fraction of private-public or public-private switches out of the total increases to around 20%. Additionally, as can be seen in Appendix Table B10, private and public wages in the same local labor market are similarly affected by changes in HHI, which suggests that public and private firms in the same industry and city belong to the same local labor market.

Because HHI is a number bounded between 0 and 1, I estimate a nonlinear first stage for the IV.[27] In my preferred specification, the prediction exercise is carried out with a random forest algorithm to allow for a high degree of nonlinear interactions between regressors.[28] In Section 4.5.1 I also present the results obtained using a logit first stage for comparison. The baseline results are presented in Table 2. Columns (1) and (2) report the OLS estimates for the overall and IV samples (health and education markets), whereas columns (3) and (4) report the IV estimates and the new agglomeration elasticity.

[27]Because the endogenous regression is bounded, the nonlinear first stage prediction must be used as *instrument* (Kelejian, 1971). In practice, I use a three-step procedure, where I first estimate a nonlinear "stage zero" with HHI as dependent variable and $\widehat{\mathrm{HHI}}_{\mathrm{pub}}$ as regressor (along with controls and fixed effects), using the random forest algorithm. Then, I take the predicted values from the previous step and, together with the controls and fixed effects, but without $\widehat{\mathrm{HHI}}_{\mathrm{pub}}$, I use them as regressors in a linear first-stage regression. Finally, I estimate the second stage as usual.

[28]To avoid overfitting, the model is trained with two thirds of the sample, whereas the remaining third is used for prediction.

The IV strategy confirms that higher levels of labor market concentration lead to lower earnings; moving from fully unconcentrated labor markets to the single monopsonist case is associated with a decrease in wages of approximately 14.5%. The estimated IV effect is slightly larger than the OLS coefficient for HHI in the same set of health and education markets. Given the IV result, the new estimated agglomeration elasticity is 0.063 and labor market power is estimated to account for approximately 30% of the city-size wage premium.

Similar to the findings in Arnold (2022) and Benmelech et al. (2022), OLS estimates appear to underestimate the causal effect of HHI on wages identified by the instrument. The IV coefficient I find is similar to the ones that Benmelech et al. (2022) and Prager and Schmitt (2021) estimate for the U.S. context, using merger-induced variation in employment concentration for identification.[29] By comparing columns (2) and (1), it also appears that the LATE effect is stronger than the treatment effect in the full sample, which partially explains the difference between the OLS and IV estimates. The first stage F-statistic is well above the conventional thresholds associated with strong instruments. The first stage and the reduced form are reported in columns (1) and (2) of Appendix Table B11.

4.5.1 Robustness

As a further check of the instrument's validity, I estimate the impact of IV on public wages. The concern is that earnings in the *public* sector may change with $\widehat{\text{HHI}}_{\text{pub}}$ for reasons unrelated to the overall change in employment concentration (e.g., a regional government that decides to invest more in the public sector and increases public employment and wages simultaneously). Since public wages are the relevant outside options for workers employed in private firms operating in the same local labor market, this channel would create a direct link between $\widehat{\text{HHI}}_{\text{pub}}$ and earnings that is not mediated by HHI, which violates exogeneity. However, the effect of the instrument on public wages is not significant (see column (3) of Table B11).

The IV results obtained using a logit model instead of the random forest algorithm to construct the instrument are reported in Appendix Table B12. Further alternative

[29] In Table 1 of their 2019 working paper, Prager and Schmitt (2021) linearly relate log wages to changes in HHI predicted by the merger and acquisition IV. They report estimates of -0.128 and -0.198 for nursing and pharmacy employees and for skilled workers, respectively (the coefficient for unskilled workers is not statistically significant). The -0.145 coefficient that I find in Table 2 falls within this window. Benmelech et al. (2022) find an IV estimate of -0.041 for log HHI. I find that the OLS coefficients of regressions that estimate the effect of log HHI, instead of HHI, on wages tend to be approximately 3.35 times lower. This back-of-the-envelope adjustment gives an estimated coefficient of -0.135, which is very close to the IV estimate in Table 2. The estimates in Arnold (2022) are instead not directly comparable, because they are presented as wage elasticities to top quartile log changes in HHI in above-median concentration markets. Similarly, the IV coefficient in Schubert et al. (2024), which leverages a shift-share instrument based on national firms' hiring growth, is difficult to compare since it relates to log HHI rather than levels. Nevertheless, it is similar to Benmelech et al. (2022), ranging from -0.028 to -0.053 depending on the specification, and therefore still comparable to my estimate.

Table 2: IV estimates

	Step 1: log W			**Step 2:** $\hat{\alpha}_m$
	(1)	(2)	(3)	(4)
Log City Size				0.0628***
				(0.0052)
HHI	-0.0734***	-0.1044***	-0.1449**	
	(0.0132)	(0.0279)	(0.0623)	
Labor Market Controls	✓	✓	✓	
Year FE	✓	✓	✓	
Market FE	✓	✓	✓	
Industry FE	Absorbed	Absorbed	Absorbed	✓
R^2	0.85	0.82	0.33	0.38
Observations	70,569	13,572	13,572	5,027
Estimation Method	OLS	OLS	IV	OLS
F-test (First Stage)	–	–	2,561	–
All Markets	✓			✓
Education and Health Markets		✓	✓	

Note: This table reports estimates of Steps 1 and 2 regressions with OLS and IV. Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of Spanish native citizens, share of male workers, share of exported revenue. The market fixed effects used as the dependent variable in Column (4) are recovered using the HHI coefficient estimate in Column (3). Standard errors are clustered at the market level in columns (1), (2), and (3), and at the industry level in column (4). *p<0.1, **p<0.05, ***p<0.01.

IV specifications are presented in Table B13. Column (1) reports the baseline IV regression with an additional market revenue productivity control. In column (2), the IV sample is restricted to the health and education industries with the highest worker flows between the public and private firms (i.e., industries with higher than median worker flows). Finally, column (3) reports IV estimates for the full sample, that is, the sample not restricted to markets related to the health and education sectors. In all these alternative specifications, the IV coefficients for HHI are slightly larger, but comparable to the baseline result.[30]

4.6 Discussion

The analysis suggests that labor market power explains a significant portion of the urban wage premium. However, since the strategy only accounts for changes in market concentration (through HHI) and frictions across local labor markets (captured by the coefficient τ), it should not be considered as accounting for all potential sources of monopsony power. Other factors, such as search costs, workplace differentiation, and additional frictions within local labor markets, may also contribute to monopsony power (Berger et al., 2022, Card et al., 2018, Lamadon et al., 2022, Manning, 2011).

[30] Estimates range from -0.167 to -0.216.

Specifically, since the frictions modeled here are between, rather than within, local labor markets, firms only internalize the market-level labor supply function. This assumption is crucial for identifying the effect of labor market power on wages for two main reasons.

First, the current assumption establishes the local labor market as the relevant unit of analysis. To identify the effect of labor market power on wages, one must find an instrument for HHI that is uncorrelated with unobservables. If the unit of analysis were the individual firm, finding such an instrument at the firm level would be quite challenging (see e.g. Datta, 2022). My strategy, which leverages changes in the size of public firms, allows for plausibly quasi-exogenous variation at the market level that would not be valid under an individual firm-level labor supply elasticity assumption. In that case, shifts that affect public firms would not necessarily influence the wages of private firms operating in the same market.

Second, even with a valid instrument at the firm level, the reduced-form estimate of the elasticity may not reflect the true structural elasticity, due to potential biases originating from strategic interactions between firms in a Cournot setting (Berger et al., 2022). This is particularly problematic when the size of the firm is substantial, or when the instrument exploits shocks that are either common to multiple firms or are purely idiosyncratic, which are scenarios that commonly arise in practice. In order to recover the true structural elasticity from the reduced-form estimate, one would need to simulate a structural model. In contrast, a market-level shock that affects all firms equally allows for an unbiased estimation of the market-level labor supply elasticity (Berger et al., 2022), ensuring the validity of my identification strategy.

To justify this approach, I have placed significant emphasis on defining local labor markets that are as homogeneous as possible, by using worker flows instead of arbitrary administrative categories (Nimczik, 2020). Additionally, I provide extensive robustness checks (see Section 4.4.2). However, this strategy, though necessary for identification, has limitations if within-market frictions are systematically related to city size. In such cases, these frictions could explain part of the city-size wage premium, on top of what is captured in my model. Some of these frictions, such as commuting time, increase with city size in Spain.[31] Other frictions, like search costs, may be lower in larger cities with thicker labor markets. Furthermore, unobserved amenities at the firm level could vary substantially between small and large urban areas, though it is difficult to predict this variation a priori. Whether these unaccounted-for frictions lead to upward or downward bias is uncertain. This is an important avenue for future research, particularly with models that can capture these additional sources

[31]The correlation between log city size and log commuting time (in minutes) is substantial, at 0.61. The median commuting times (in minutes) by city size quintile are as follows, from smallest to largest cities: 13.3, 14.2, 14.5, 15.7, and 17.7.

of monopsony power.

Another channel my model does not account for is labor market frictions that lead to higher assortative matching between firms and workers, and consequently, higher productivity in larger cities (Dauth et al., 2022). While this effect is captured in a reduced-form manner through the exogenous agglomeration elasticity (Duranton and Puga, 2004), further disentangling the impact of labor market frictions on productivity from their effect on wages through monopsony is a promising area of research. Nevertheless, to the extent that my identifying variation coming from changes in the local size of public firms does not significantly influence matching in the labor market, the effect of labor market concentration on wages that I estimate is not biased. While this cannot be tested in the current setting, the fact that the IV is not correlated with measured productivity (which also reflects matching) in either levels or changes is reassuring.

Finally, a limitation of my IV strategy is that the identification stems from a subset of markets (the health and education sectors), covering only 19% of all markets. When comparing the OLS estimates for this subset with those for all markets, I find that the effect is stronger (-0.1044 versus -0.0734), which may partly explain why the LATE estimate identified using the IV (-0.1449) is lower than the OLS estimate. A further limitation is that testing for the exogeneity of the IV would require correlating it with AMRPL, which is not directly observed. Instead, I use a proxy for it, measured productivity $\widetilde{\text{AMRPL}}$, which introduces potential noise and requires the assumption of a Cobb-Douglas production function (see Section D.1.1). Nevertheless, as expected, $\widetilde{\text{AMRPL}}$ predicts the contribution of private firms to HHI, as well as wages, but not the IV.

5 Conclusion

Local labor markets in larger cities tend to be more competitive. If firms in small cities have higher labor market power and pay workers less than their marginal products, this could generate part of the city-size wage premium observed in the data. I use administrative data for Spain to quantify this channel.

A Rosen-Roback model with imperfect competition in the labor market rationalizes the correlation between labor market concentration and wages observed in the data as a spatial equilibrium in which neither firms nor workers have an incentive to move. The model also guides the empirical strategy.

I use two complementary approaches for identification. First, I control for latent productivity using a rich set of fixed effects, market-level time trends, and balance sheet revenue data. I then exploit quasi-experimental variation in labor market power stemming from changes in the local size of the public sector. The results from both approaches indicate that differences in labor market power across urban areas are a significant factor driving the city-size wage premium, accounting for $20-30\%$ of the wage gap between small and large cities.

The finding that labor market power contributes to the urban wage premium has a range of implications that warrant further investigation. For instance, it is indicative of another cost associated with restrictive land use regulations in large, productive cities. Moreover, it may suggest the need for a more spatially-oriented approach to antitrust policy. Additionally, it provides insight into discussions surrounding the decentralization of government employment, potentially as a means of promoting competition in smaller cities.

References

Almunia, Miguel, David Lopez Rodriguez and Enrique Moral-Benito. 2018. "Evaluating the Macro-Representativeness of a Firm-Level Database: An Application for the Spanish Economy." Working Paper.

Arellano, Manuel, Olympia Bover, and Samuel Bentolila. 2002. "The Distribution of Earnings in Spain During the 1980s: The Effects of Skill, Unemployment and Union Power." in *The Economics of Rising Inequalities*, 3-53, Oxford University Press and CEPR.

Arnold, David. 2022. "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes." Working Paper.

Azar, José, Steven Berry, and Ioana Marinescu. 2022. "Estimating Labor Market Power." Working Paper.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska. 2020. "Concentration in US Labor Markets: Evidence From Online Vacancy Data." *Labour Economics*, 66, 101886.

Azar, José, and Xavier Vives. 2021. "General Equilibrium Oligopoly and Ownership Structure." *Econometrica*, 89(3), 999-1048.

Azkarate-Askasua, Miren, and Miguel Zerecero. 2022. "The Aggregate Effects of Labor Market Concentration." Working Paper.

Bamford, Iain. 2022. "Monopsony Power, Spatial Equilibrium, and Minimum Wages." Working Paper.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu. 2022. "Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies." *Journal of Human Resources*, 57(S), S50–S86.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy*, 122(3), 507-553.

Benmelech, Efraim, Nittai Bergman, and Hyunseob Kim. 2022. "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?." *Journal of Human Resources*, 57(S), S200-S250.

Berger, David W., Kyle F. Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *American Economic Review*, 112(4), 1147-1193.

Bernard, Andrew B., and J. Bradford Jensen. 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics*, 47(1), 1-25.

Berry, Steven, Martin Gaynor and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives*, 33(3), 44-68.

Caldwell, Sydnee and Oren Danieli. 2024. "Outside Options in the Labour Market." *Review of Economic Studies*.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics*, 36(S1), S13-S70.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. "Estimating Agglomeration Effects with History, Geology, and Worker Fixed Effects." in *Agglomeration Economics*, pp. 15–65, Chicago, IL: Chicago University.

Datta, Nikhil. 2022. "Local Monopsony Power." Working Paper.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum. 2022. "Matching in Cities." *Journal of the European Economic Association*, 20(4), 1478-1521.

De la Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." *Review of Economic Studies*, 84(1), 106–142.

Duranton, Gilles, and Diego Puga. 2004. "Micro-Foundations of Urban Agglomeration Economies." in *Handbook of Regional and Urban Economics, Vol. 4*, pp. 2063-2117, Henderson, V. and Thisse, J.-F.

Glaeser, Edward L. and David C. Maré. 2001. "Cities and Skills." *Journal of Labor Economics*, 19(2), 316–342.

Goerlich, Francisco José, and Joaquín Azagra. 2006. "La localización de la poblacion española sobre el territorio. Un siglo de cambios: un estudio basado en series homogéneas." *Bilbao, Spain: BBVA Foundation*.

Goerlich, Francisco José, and Isidro Cantarino. 2013. "A Population Density Grid for Spain." *International Journal of Geographical Information Science*, 27(12), 2247–2263.

Guillouzouic, Arthur, Emeric Henry, and Joan Monras. 2022. "Local Public Goods and the Spatial Distribution of Economic Activity." Working Paper.

Hirsch, Boris, Elke J. Jahn, Alan Manning, and Michael Oberfichtner. 2022. "The Urban Wage Premium in Imperfect Labour Markets." *Journal of Human Resources*, 57(S), S111-S136.

Jarosch, Gregor and Jan Sebastian Nimczik, and Isaac Sorkin. 2024. "Granular Search, Market Structure, and Wages." *Review of Economic Studies*.

Kelejian, Harry H. 1971. "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables." *Journal of the American Statistical Association*, 66, 373-374.

Moretti, Enrico. 2011. "Local Labor Markets." in *Handbook of Labor Economics*, vol. 4, 1237-1313, Elsevier.

Lamadon, Thibaut, Magne Mogstad and Bradley Setzler. 2022. "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market." *American Economic Review*, 112(1), 169-212.

Langella, Monica, and Alan Manning. 2021. "The Measure of Monopsony." *Journal of the European Economic Association*, 19(6), 2929-2957.

Lipsius, Ben. 2018. "Labor Market Concentration Does Not Explain the Falling Labor Share." Working Paper.

Manning, Alan. 2010. "The Plant Size-Place Effect: Agglomeration and Monopsony in Labour Markets." *Journal of Economic Geography*, 10(5): 717-744.

Manning, Alan. 2011. "Imperfect Competition in the Labor Market." in *Handbook of Labor Economics*, vol. 4, pp. 973-1041, Elsevier.

Manning, Alan and Barbara Petrongolo. 2017. "How Local Are Labor Markets? Evidence From a Spatial Job Search Model." *American Economic Review*, 107(10): 2877-2907.

McCormick, Michael, Huang, Guoping, Zambotti, Giovanni, and Jessica Lavash. 2008. "Roman Road Network." *Cambridge, MA: Digital Atlas of Roman and Medieval Civilizations (DARMC)*, Center for Geographic Analysis, Harvard University.

Nickell, Stephen, Jari T. Vainiomaki, and Sushil Wadhwani. 1994. "Wages and Product Market Power." *Economica*, 61: 457-473.

Nimczik, Jan S. 2020. "Job Mobility Networks and Data-Driven Labor Markets." Working Paper.

Prager, Elena and Matt Schmitt. 2021. "Employer Consolidation and Wages: Evidence from Hospitals." *American Economic Review*, 111(2): 397-427.

Sokolova, Anna and Todd Sorensen. 2021. "Monopsony in Labor Markets: A Meta-Analysis." *ILR Review*, 74(1): 27-55.

Schubert, Gregor, Anna Stansbury, and Bledi Taska. 2024. "Employer Concentration and Outside Options." Working Paper.

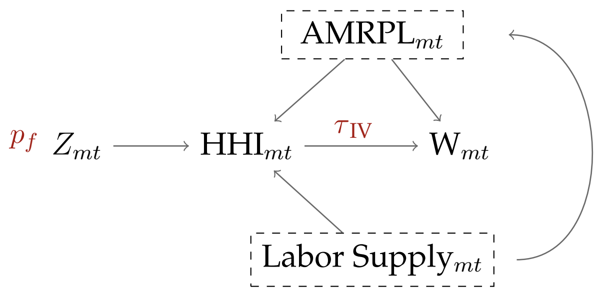
Appendix

This Appendix is organized as follows. Sections A and B contain additional tables and figures referenced in the text. Section C provides details about the data and the construction of the instrument. Section D presents further details and extensions of the model and the estimation strategy.

A Additional Figures

The DAG shown in Figure A1 illustrates how the instrument Z_{mt}, based on shifts in local public firms' size, helps identify the coefficient τ_{IV}, bypassing the endogeneity issues caused by unobserved productivity AMRPL_{mt}.

Figure A1: Identification of τ_{IV} using the size of local public firms as IV for HHI_{mt}



Note: This figure draws a directed acyclic graph (DAG) of the IV model. The variables are enclosed within dashed lines if they are unobservable. Arrows represent causal relationship between the variables in the data generating process.

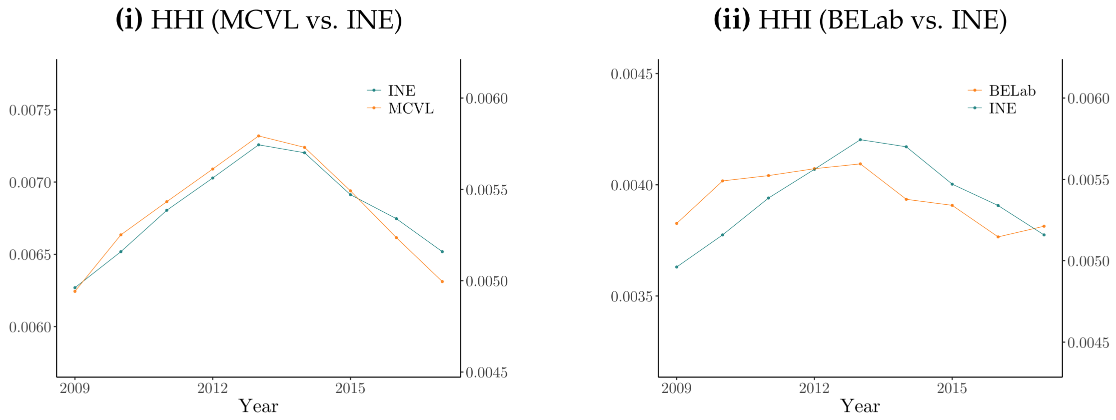
Figure A2 evaluates the accuracy of the HHI measurement. The most reliable benchmark comes from the universe of firms in the INE, which compiles data from the *Demografía Armonizada de Empresas*. However, the granularity of this data is limited to the regional (Comunidad Autónoma) and 2-digit NACE sector levels. Furthermore, establishments are categorized into bins based on employee count: 1-4 employees, 5-9 employees, and over 10 employees, without any further detail available within those categories. Additionally, the data is only accessible for the years 2009-2017. Consequently, while this source serves as a valuable benchmark, it is unsuitable for detailed analysis compared to the two alternatives: MCVL and BELab.

To facilitate comparison, establishments in both the MCVL and BELab are similarly categorized into the bins of 1-4 employees, 5-9 employees, and over 10 employees. The HHI is then computed for all three data sources, based on the average employment levels within bins derived from the MCVL.[32] The local labor market is defined at the

[32]Using data from BELab delivers very similar results.

regional and 2-digit sector level, the most granular unit of analysis available in the INE data. The national level weighted average of HHI is computed using market employment as weight and plotted in Figure A2.

Figure A2: Mean HHI: MCVL and BELab vs. INE



Note: This figure plots the time series of the employment HHI computed with MCVL (yellow line) and INE data (blue line). Markets are defined at the regional (Comunidad Autónoma) and 2-digit NACE sector level. The national level weighted average of HHI is computed using market employment as weight. The scales of the vertical axes are such that equal *percentage* changes over time of HHI are represented equivalently in the two series. INE data comes from the *Demografía Armonizada de Empresas*, which measures the stock of all establishments operating in Spain by dividing them into bins of establishments with 1-4 employees, 5-9 employees, and more than 10 employees. The average number of workers per firm in each category is recovered using the MCVL, so that the approximate employment distribution of firms from INE data can be recovered accordingly. This allows us to compute the employment HHI using the INE data. For comparibility, establishments in the MCVL and BELab are equally categorized in bins of 1-4 employees, of 5-9 employees, and more than 10 employees. The HHI is then computed with MCVL and BELab data using the same procedure.

When evaluating these alternatives against the INE, it becomes clear that the MCVL performs significantly better. This is illustrated by comparing panel (i) of Figure A2 with panel (ii), where we can observe that the MCVL more accurately captures fluctuations over time.[33] While this may seem surprising, given the smaller sample size of the MCVL, several factors contribute to this outcome.

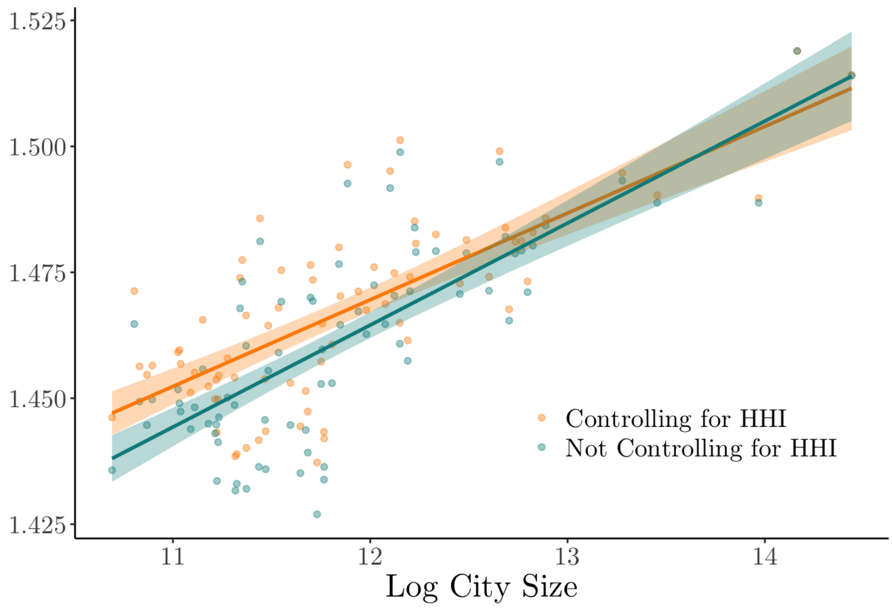
First, it is important to clarify that the BELab has information on the quasi-universe of firms, rather than the universe. Although BELab has a high and stable coverage ratio at 82% (see Almunia et al., 2018), it does not encompass all firms. The MCVL, despite covering a relatively small sample of workers, achieves exceptionally high coverage for larger establishments, with 89.6% of all Spanish firms with more than ten employees included.[34] This high representativity is crucial, since the dynamics of large firms play a disproportionate role in the evolution of the HHI, given that employment shares enter in the HHI formula with a square. Another reason behind the better fit of the MCVL over BELab may be the greater measurement error present in the BELab data, which is susceptible to reporting inaccuracies common in large-scale

[33]Even if the HHI levels are marginally different with respect to INE, the scales of the vertical axes are such that equal percentage changes over time of HHI are represented equivalently in the two series.
[34]The coverage is reduced to 7.4% among firms with 1-4 employees and to 28.2% for firms with 5-9 employees.

microdatasets. Finally, the BELab is a firm-level dataset, whereas the MCVL provides more accurate information at the establishment level.

It should also be noted that the MCVL enables a broader range of robustness checks, offering flexibility through alternative local labor market definitions that incorporate skill information (exclusively available in the MCVL). These include 1-digit sector $\times$ skill markets or flow-based clusters that aggregate 3-digit sectors $\times$ skills cells (see Section 4.4.2). For all these reasons, I conclude that the MCVL is the superior choice for measuring the HHI.

Figure A3: Market fixed effects of regression (8)



Note: This figure plots the market fixed effects of regression (8) as a function of the size of the city where markets are located. The market fixed effects are separately estimated by controlling and not controlling for HHI (yellow and blue dots, respectively). Market fixed effects are averaged at the city level. City size is population within 10km of the average resident (De la Roca and Puga, 2017).

Figure A3 plots the market fixed effects of regression (8) as a function of the size of the city where markets are located. The market fixed effects are separately estimated by controlling and not controlling for HHI (yellow and blue dots, respectively), and then averaged at the city level. Not controlling for HHI in regression (8) leads to lower estimates of the market fixed effects in small urban areas and, hence, to a higher estimated agglomeration elasticity.

B Additional Tables

Table B1: First stage regression of IV for city size

	Log City Size
Log City Size in 1900	0.6538***
	(0.0017)
Fertile Land Within 25km (%)	0.0143***
	(0.0002)
Water Within 25km (%)	0.0058***
	(0.0000)
Steep Terrain Within 25km (%)	-0.0134***
	(0.0001)
Log Mean Elevation Within 25km (%)	0.2800***
	(0.0025)
Roman Road Rays Within 25km	0.0694***
	(0.0009)
Industry FE	✓
R^2	0.66
Observations	5,027
F-test	1,591

Note: This table reports estimates of the first stage regression for the IV strategy of the log city size variable in Step 2. Standard errors are clustered at the industry level. *p<0.1, **p<0.05, ***p<0.01.

Table B2: Step 2 regression with and without IV

	$\hat{\alpha}_m$			
	(1)	(2)	(3)	(4)
Log City Size	0.0888***		0.0745***	
	(0.0058)		(0.0054)	
$\widehat{\text{Log City Size}}$		0. 0966***		0.0835***
		(0.0066)		(0.0063)
Industry FE	✓	✓	✓	✓
R^2	0.38	0.35	0.38	0.37
Observations	5,027	5,027	5,027	5,027
Estimation Method	OLS	IV	OLS	IV
F-test (First Stage)	–	1,591	–	1,591
Step 1 with HHI			✓	✓

Note: This table reports estimates of the Step 2 regression with and without IV (columns (2) and (4) vs. columns (1) and (3)), and with and without the HHI control variable in Step 1 (columns (3) and (4) vs. columns (1) and (2)). Standard errors are clustered at the industry level. *p<0.1, **p<0.05, ***p<0.01.

Table B3: Step 2 regression with and without controlling for natural city amenities

	$\hat{\alpha}_m$			
	(1)	(2)	(3)	(4)
Log City Size	0.0888***	0.0905***	0.0745***	0.0761***
	(0.0058)	(0.0061)	(0.0054)	(0.0058)
Log Precipitations		0.0123*		0.0097
		(0.0067)		(0.0067)
Log Distance from Coast		0.0053		0.0037
		(0.0032)		(0.0032)
Log Mean Temperature		-0.0321		-0.0366
		(0.0228)		(0.0225)
Water Within 25km (%)		0.0001		-0.0000
		(0.0002)		(0.0002)
Industry FE	✓	✓	✓	✓
R^2	0.38	0.39	0.38	0.39
Observations	5,027	5,027	5,027	5,027
Step 1 with HHI			✓	✓

Note: This table reports estimates of the Step 2 regression with and without natural amenities controls (columns (1) and (3) vs. columns (2) and (4)), and with and without the HHI control variable in Step 1 (columns (3) and (4) vs. columns (1) and (2)). Standard errors are clustered at the industry level. *p<0.1, **p<0.05, ***p<0.01.

Table B4: Step 2 regression with and without controlling for natural and endogenous city amenities

	$\hat{\alpha}_m$			
	(1)	(2)	(3)	(4)
Log City Size	0.0888***	0.0747***	0.0745***	0.0611***
	(0.0058)	(0.0078)	(0.0054)	(0.0075)
Log Precipitations		0.0171**		0.0143**
		(0.0067)		(0.0066)
Log Distance from Coast		0.0110**		0.0087**
		(0.0044)		(0.0043)
Log Mean Temperature		-0.0509**		-0.0536**
		(0.0234)		(0.0230)
Water Within 25km (%)		0.0006**		0.0005*
		(0.0002)		(0.0002)
Log Pollution (NO2 Conc.)		0.0349***		0.0308***
		(0.0081)		(0.0080)
Log Mean Commuting Time		-0.0391		-0.0276
		(0.0315)		(0.0310)
Log Crimes per Person		0.0538***		0.0473***
		(0.0073)		(0.0071)
Log Cinemas per Person		0.0319***		0.0340***
		(0.0106)		(0.0105)
Industry FE	✓	✓	✓	✓
R^2	0.38	0.39	0.38	0.39
Observations	5,027	5,027	5,027	5,027
Step 1 with HHI			✓	✓

Note: This table reports estimates of the Step 2 regression with and without natural and endogenous amenities controls (columns (1) and (3) vs. columns (2) and (4)), and with and without the HHI control variable in Step 1 (columns (3) and (4) vs. columns (1) and (2)). Standard errors are clustered at the industry level. *p<0.1, **p<0.05, ***p<0.01.

Table B5: Step 1 regression controlling for revenue productivity

	log W				
	(1)	(2)	(3)	(4)	(5)
HHI	-0.1047***	-0.0982***	-0.0738***	-0.0613***	-0.0569***
	(0.0176)	(0.0171)	(0.0140)	(0.0141)	(0.0144)
Sales HHI	0.0299***	0.0192*	-0.0100	-0.0049	-0.0049
	(0.0109)	(0.0110)	(0.0076)	(0.0079)	(0.0076)
Log Productivity (AMRPL)		0.0251***	0.0000	-0.0001	-0.0006
		(0.0056)	(0.0031)	(0.0031)	(0.0031)
Labor Market Controls	✓	✓	✓	✓	✓
Market-Level Linear Trends					✓
City-Year, Industry-Year FE				✓	✓
Market FE			✓	✓	✓
Year FE	✓	✓	✓	✓	✓
City FE	✓	✓	Absorbed	Absorbed	Absorbed
Industry FE	✓	✓	Absorbed	Absorbed	Absorbed
R^2	0.71	0.71	0.85	0.86	0.91
Observations	64,246	64,246	64,246	64,246	64,246

Note: This table reports estimates of Step 1 regressions controlling for revenue productivity. Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of Spanish native citizens, share of male workers, share of exported revenue. Standard errors are clustered at the city and industry level in (1) and (2) and the market level in (3), (4) and (5). *p<0.1, **p<0.05, ***p<0.01.

Table B6: Robustness checks for Step 1 regression

	log W			
	(1)	(2)	(3)	(4)
HHI	-0.0738***	-0.0756***	-0.0747***	-0.0685***
	(0.0140)	(0.0160)	(0.0149)	(0.0070)
Sales HHI	-0.0101	-0.0056	-0.0111	-0.0080
	(0.0072)	(0.0076)	(0.0088)	(0.0099)
Log Productivity (AMRPL)		0.0016		
		(0.0033)		
Unemployment Rate		0.0224		
		(0.0684)		
Labor Market Controls	✓	✓	✓	✓
Market FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
R^2	0.85	0.86	0.88	0.53
Observations	64,246	59,540	42,621	53,168
% CSWP Explained	16.1%	-	15.9%	27%
Sample	All Years (2005-2019)	All Years (2005-2019)	No Recession Years	Only New Hires

Note: This table reports estimates of Step 1 regressions controlling for local unemployment and productivity in column (2), excluding recession years, i.e. years 2008-2009 and 2011-2013, in column (3) and using a sample that only includes new hires in column (4). Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of Spanish native citizens, share of male workers, share of exported revenue. Standard errors are clustered at the market level. *p<0.1, **p<0.05, ***p<0.01.

Table B7: Alternative local labor market definitions

	log W					
	(1)	(2)	(3)	(4)	(5)	(6)
HHI	-0.0738***	-0.0755***	-0.0892***	-0.0669***	-0.0521***	-0.0867***
	(0.0140)	(0.0150)	(0.0196)	(0.0126)	(0.0115)	(0.0156)
Sales HHI	-0.0101	0.0165	0.0024	0.0104	-0.0049	0.0233**
	(0.0072)	(0.0109)	(0.0065)	(0.0076)	(0.0068)	(0.0097)
Labor Market Controls	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
R^2	0.85	0.88	0.77	0.75	0.82	0.84
Observations	64,246	68,060	68,236	69,472	56,512	67,896
% CSWP Explained	16.1%	11.9%	18.4%	10.1%	12.2%	18.2%
Def. of Local Labor Markets (Within UAs)	Baseline	Flows 3-Digit Sectors - Skills	Flows Years 1970-2004	Flows Varying by City Size	2-Digit Sectors	1-Digit Sectors - Skill

Note: This table reports estimates of Step 1 regressions with alternative local labor market definitions. These include the baseline model in column (1), the data-driven definition based on worker flows between cells of 3-digit sectors interacted with skill groups in column (2), using worker flows of years 1970-2004 in column (3), separately estimating the data-driven markets by city size in column (4), using 2-digit sectors in column (5) and 1-digit sector by skill combinations in column (6). Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of Spanish native citizens, share of male workers, share of exported revenue. Standard errors are clustered at the market level. *$p<0.1$, **$p<0.05$, ***$p<0.01$.

Table B8: Effect of revenue productivity on $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$

	$\widehat{\text{HHI}}_{\text{priv}}$		$\widehat{\text{HHI}}_{\text{pub}}$	
	(1)	(2)	(3)	(4)
Log Productivity (AMRPL)	-0.0403**	-0.0314	0.0829*	0.0409
	(0.0169)	(0.0425)	(0.0451)	(0.0352)
Labor Market Controls	✓	✓	✓	✓
Market FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
R^2	0.68	0.72	0.80	0.83
Observations	59,979	10,292	10,007	6,385
All Markets	✓		✓	
Education and Health Markets		✓		✓

Note: This table reports estimates of the effect of revenue productivity on $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$. All markets are included in columns (1) and (3), whereas only education and health-related labor markets are included in columns (2) and (4). Such markets include the following industries: "Medical and dental activities", "Hospital activities", "Social service activities for the elderly and the disabled", "Assistance in residential facilities for the elderly and the disabled", "Assistance in residential care facilities with health care", "Residential care activities for persons with intellectual disabilities, mental illness and drug addiction", "Other residential care activities", "Other social work activities", "Other health-related activities", "Pre-primary education", "Primary education", "Secondary education", "Postsecundary education", "Research activities" ,"Research and development in Social Sciences and Humanities", "Auxiliary activities to education", "Other educational activities". Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of spanish native citizens, share of male workers, share of exported revenue. Logit model. Standard errors are clustered at the market level. *$p<0.1$, **$p<0.05$, ***$p<0.01$.

Table B9: Effect of AMRPL (log changes) on $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$ (log changes)

	Δ log$\widehat{\text{HHI}}_{\text{priv}}$		Δ log$\widehat{\text{HHI}}_{\text{pub}}$	
	(1)	(2)	(3)	(4)
Δ Log Productivity (AMRPL)	-0.0119**	0.0063	0.0168	-0.0037
	(0.0055)	(0.0140)	(0.0149)	(0.0172)
R^2	0.00	0.00	0.00	0.00
Observations	53,119	8,932	8,428	5,445
All Markets	✓		✓	
Education and Health Markets		✓		✓

Note: This table reports estimates of the effect of AMRPL (log changes) on $\widehat{\text{HHI}}_{\text{priv}}$ and $\widehat{\text{HHI}}_{\text{pub}}$ (log changes). All markets are included in columns (1) and (3), whereas only education and health-related labor markets are included in columns (2) and (4). Such markets include the following industries: "Medical and dental activities", "Hospital activities", "Social service activities for the elderly and the disabled", "Assistance in residential facilities for the elderly and the disabled", "Assistance in residential care facilities with health care", "Residential care activities for persons with intellectual disabilities, mental illness and drug addiction", "Other residential care activities", "Other social work activities", "Other health-related activities", "Pre-primary education", "Primary education", "Secondary education", "Postsecundary education", "Research activities" ,"Research and development in Social Sciences and Humanities", "Auxiliary activities to education", "Other educational activities". *p<0.1, **p<0.05, ***p<0.01.

Table B10: Effect of HHI on private, public, and overall wages

	W_{priv}	W_{pub}	W_{all}
	(1)	(2)	(3)
HHI	-0.0734***	-0.0713*	-0.0733***
	(0.0132)	(0.0415)	(0.0127)
Labor Market Controls	✓	✓	✓
Market FE	✓	✓	✓
Year FE	✓	✓	✓
R^2	0.85	0.87	0.86
Observations	70,569	11,987	71,527
Private Firms	✓		✓
Public Firms		✓	✓

Note: This table reports estimates of the effect of HHI on private, public, and overall wages. Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of spanish native citizens, share of male workers, share of exported revenue. Standard errors are clustered at the market level. *p<0.1, **p<0.05, ***p<0.01.

Table B11: Effect of HHI on wages (first stage and reduced form of IV estimates) and of the IV on public wages

	HHI	log W	log W_{pub}
	(1)	(2)	(3)
$\widehat{\text{HHI}}_{\text{pub, forest}}$	1.290***	-0.1872***	-0.0260
	(0.0253)	(0.0553)	(0.1326)
Labor Market Controls	✓	✓	✓
Market FE	✓	✓	✓
Year FE	✓	✓	✓
R^2	0.48	0.13	0.31
Observations	13,572	13,572	8,246
Regression	First Stage	Reduced Form	OLS
F-test	2,561	–	–
Education and Health Markets	✓	✓	✓

Note: This table reports estimates of the effect of HHI on wages (first stage and reduced form of IV estimates) and of the IV on public wages. Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of spanish native citizens, share of male workers, share of exported revenue. The nonlinear "stage zero" prediction estimated with the random forest algorithm and used as instrument is denoted by $\widehat{\text{HHI}}_{\text{pub, forest}}$. The random forest model is trained with two thirds of the sample, whereas the remaining third is used for prediction. Standard errors are clustered at the market level. *p<0.1, **p<0.05, ***p<0.01.

Table B12: Effect of HHI on wages (first stage and reduced form of logit IV estimates)

	HHI	log W	
	(1)	(2)	(3)
$\widehat{\text{HHI}}_{\text{pub, logit}}$	1.066***	-0.2092*	
	(0.0556)	(0.1135)	
HHI			-0.1964*
			(0.1065)
Labor Market Controls	✓	✓	✓
Market FE	✓	✓	✓
Year FE	✓	✓	✓
R^2	0.82	0.82	0.82
Observations	13,900	13,166	13,166
Regression	First Stage	Reduced Form	IV
F-test (First Stage)	621.5	–	621.5
Education and Health Markets	✓	✓	✓

Note: This table reports estimates of the effect of HHI on wages (first stage and reduced form of logit IV estimates). Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of spanish native citizens, share of male workers, share of exported revenue. The nonlinear "stage zero" prediction estimated with the logistic regression and used as instrument is denoted by $\widehat{\text{HHI}}_{\text{pub, logit}}$. Standard errors are clustered at the market level. *p<0.1, **p<0.05, ***p<0.01.

Table B13: Effect of HHI on wages (IV estimates)

	log W		
	(1)	(2)	(3)
HHI	-0.1670**	-0.2161**	-0.2076***
	(0.0656)	(0.0736)	(0.0305)
Productivity (AMRPL) Control	✓		
Labor Market Controls	✓	✓	✓
Market FE	✓	✓	✓
Year FE	✓	✓	✓
R^2	0.31	0.38	0.31
Observations	12,658	7,084	56,747
F-test (First Stage)	2,080	1,118	14,034
Sample	Health and Education (H&E)	Highest Pub-Priv Flows H&E	All Markets

Note: This table reports estimates of the effect of HHI on wages (IV estimates). Column (1) reports the baseline IV regression with an additional market revenue productivity control. In column (2), the IV sample is restricted to the health and education industries with the highest worker flows between the public and private sector (higher than median worker flows). Column (3) reports IV estimates for the full sample, i.e. not restricted to markets related to the health and education sectors. Labor market controls include average worker experience and tenure years, share of workers with high school and university education level, share of jobs by task content (five skill levels), share of workers covered by collective agreements (unions), contract types shares (temporary or permanent), share of spanish native citizens, share of male workers, share of exported revenue. Standard errors are clustered at the market level. *p<0.1, **p<0.05, ***p<0.01.

C Empirical Analysis

C.1 Data Appendix: MCVL

The MCVL (*Muestra Continua de Vidas Laborales*, or Continuous Sample of Employment Histories) is a 4% non-stratified sample of individuals affiliated to the Spanish social security. The panel records any change in individuals' labor market status (working, receiving unemployment benefits, or receiving a pension). Job changes and contractual modifications within the same firm are also recorded. Information on wages is provided for the entire working life of the sampled individuals when available. We focus on 2005–2019, the period in which job spells are matched with tax record data that provide uncensored earnings, and compute daily full-time equivalent wages using the available information on working hours. For a small number of cases, the computed wages are much higher than workers' contributions to social security. To prevent these outliers from affecting the results, I remove the observations corresponding to the top 1% of the wage distribution. Workers' tenure and experience are measured by counting the number of employment days in the current establishment and during the entire working life, respectively. Furthermore, the MCVL provides information on workers' gender and age, which are contained in social security records. The sample is also matched with Spain's Continuous Census of Population (*Padrón Continuo*), so that individual characteristics such as country of birth, nationality, and educational attainment can be recovered.

Employers assign workers to different social security contribution groups that are highly related to the level of education required to perform the job. Following De la Roca and Puga (2017), I organize these groups into five skill categories: *very high-skilled*, *high-skilled*, *medium-high-skilled*, *medium-low-skilled*, and *low-skilled* occupations. For example, the upper contribution group, which includes *very high-skilled* occupations, is reserved for jobs that require an engineering or bachelor's degree and for top managerial positions. The MCVL further reports detailed information on employers, such as their firms' employment levels or ownership status (private or public). Finally, the NACE 3-digit sector of the establishment and workplace location are reported so that each employer can be assigned to a single local labor market.

The panel used for the analysis covers working individuals aged 18 or older. Data collected in the Basque Country and Navarre are excluded from the analysis because they do not provide information on uncensored earnings. Furthermore, local labor markets in three small urban areas are not considered because workplace locations are not reported for municipalities with less than 40,000 inhabitants.

C.2 IV Construction

In this section, I describe the construction of the instrumental variable presented in Section 4.5. Let the IV be denoted by Z_{mt}. This vector measures the predicted impact of changes in the size of local public firms on HHI_{mt} for all markets m and years t in which some public firms operate. For m and t where there are no public firms, Z_{mt} equals zero.

The following example illustrates how the instrument is computed. Consider a market with four competing establishments denoted by *a*, *b*, *c*, and *d*. Firms *a* and *b* are public, whereas firms *c* and *d* are private. Total employment E and the employment HHI in the market at time t are given by

$$E_t = \underbrace{e_{a,t}^{\text{pub}} + e_{b,t}^{\text{pub}}}_{E_t^{\text{pub}}} + \underbrace{e_{c,t}^{\text{priv}} + e_{d,t}^{\text{priv}}}_{E_t^{\text{priv}}}$$

$$\text{HHI}_t = \frac{\left(e_{a,t}^{\text{pub}}\right)^2 + \left(e_{b,t}^{\text{pub}}\right)^2 + \left(e_{c,t}^{\text{priv}}\right)^2 + \left(e_{d,t}^{\text{priv}}\right)^2}{(E_t)^2}$$

In the next period, establishments change their employment levels. The new level of labor market concentration is then given by

$$\text{HHI}_{t+1} = \frac{\left(e_{a,t}^{\text{pub}} + \Delta e_{a,t}^{\text{pub}}\right)^2 + \left(e_{b,t}^{\text{pub}} + \Delta e_{b,t}^{\text{pub}}\right)^2 + \left(e_{c,t}^{\text{priv}} + \Delta e_{c,t}^{\text{priv}}\right)^2 + \left(e_{d,t}^{\text{priv}} + \Delta e_{d,t}^{\text{priv}}\right)^2}{\left(E_t + \Delta E_t^{\text{pub}} + \Delta E_t^{\text{priv}}\right)^2}$$

Suppose changes in public employment are "exogenous" to local productivity shocks. Then

$$Z_{mt} = \widehat{\text{HHI}}_{t+1}^{\text{pub}} = \frac{\left(e_{a,t}^{\text{pub}} + \Delta e_{a,t}^{\text{pub}}\right)^2 + \left(e_{b,t}^{\text{pub}} + \Delta e_{b,t}^{\text{pub}}\right)^2}{\left(E_t + \Delta E_t^{\text{pub}}\right)^2}$$

is our candidate instrument for HHI_{t+1}. Similarly, we can define

$$\widehat{\text{HHI}}_{t+1}^{\text{priv}} = \frac{\left(e_{c,t}^{\text{priv}} + \Delta e_{d,t}^{\text{priv}}\right)^2 + \left(e_{b,t}^{\text{priv}} + \Delta e_{b,t}^{\text{priv}}\right)^2}{\left(E_t + \Delta E_t^{\text{priv}}\right)^2}.$$

This quantity measures the variation in HHI driven by changes in the employment of private firms and that is then likely endogenous to local productivity shocks.

Finally, changes in HHI that are driven by the entry or exit of public firms to and from a local labor market are not included the instrument $\widehat{\text{HHI}}_{t+1}^{\text{pub}}$, since these events are less likely to be exogenous. For example, people may anticipate the construction of an hospital and migrate to the city, causing a supply shock. Additionally, an hospital shutting down may be indicative of an unobserved demographic shock in the area.

D Model Appendix

D.1 Market Equilibrium with Decreasing Returns to Scale

D.1.1 Asymmetric Firms

Suppose N_{mt} firms compete *à la Cournot* for workers in the local labor market m and at time t, with the market being defined as a cluster of subindustries indexed by k within a city c (i.e., local labor markets are city-cluster combinations). Each firm f has a Cobb-Douglas production function with labor l_{ft} as the sole input,

$$Q_{ft} = A_{fmt} l_{ft}^{\theta}, \quad \theta \leq 1,$$

and sells its product at the competitive price $P_{ft} = 1$. Firms are heterogeneous in productivity A_{fmt}, and A_{fmt} has a market-time component that is common across all firms that compete in market m at time t. For example, this captures the productivity advantage of markets m located in large cities. In particular, we assume that $A_{fmt} = A_{mt}\epsilon_{fmt}^{A} \geq 0$. Firms internalize that they face the upward sloping labor market supply curve

$$W_{mt} = \beta_{mt} L_{mt}^{\tau},$$

where $\tau = \eta^{-1}$ is the inverse labor supply elasticity, assumed to be constant across markets. $L_{mt} = \sum_{f=1}^{N_{mt}} l_{ft}$ denotes total market employment, and the intercept β_{mt} is indexed by m to reflect market (and time) varying factors that influece local supply (e.g. consumption amenities that affect migration across cities).

Firms choose l_{ft} to maximize profits

$$\pi_{ft} = A_{fmt} l_{ft}^{\theta} - W_{mt}(L_{mt}) l_{ft}.$$

Denoting with $s_{ft} = \frac{l_{ft}}{L_{mt}}$ the employment share of firms, the first-order condition of each firm is given by

$$W_{mt} \left(1 + \tau s_{ft}\right) = \theta A_{fmt} l_{ft}^{\theta - 1}. \tag{10}$$

This indicates that more productive firms are larger in size. Multiplying both sides of the equation by s_{ft} and summing across all firms in the market, we obtain the wage setting formula

$$W_{mt} \left(1 + \tau \text{HHI}_{mt}\right) = \text{AMRPL}_{mt}, \tag{11}$$

where we have defined the Herfindahl-Hirschman Index

$$\text{HHI}_{mt} = \sum_{f=1}^{N_{mt}} s_{ft}^2,$$

and the average marginal revenue productivity of labor

$$\text{AMRPL}_{mt} = \sum_{f=1}^{N_{mt}} s_{ft} \theta A_{fmt} l_{ft}^{\theta-1} = \theta A_{mt} \sum_{f=1}^{N_{mt}} s_{ft} \epsilon_{fmt}^{A} l_{ft}^{\theta-1}. \tag{12}$$

If the labor market is perfectly competitive, then firms are atomistic ($s_{ft} \to 0$), HHI_{mt} goes to zero, and productivity is fully passed through to wages. On the other hand, with imperfect competition, we have $\text{HHI}_{mt} > 0$, and firms force a markdown upon workers unless their supply is perfectly elastic ($\tau = 0$). Finally, it is easy to see that

$$\text{AMRPL}_{mt} = \theta \sum_{f=1}^{N_{mt}} s_{ft} \frac{\overbrace{P_{ft}}^{=1} Q_{ft}}{l_{ft}} \tag{13}$$

D.1.2 Symmetric Firms

In the symmetric Cournot model, $s_{ft} = \frac{l_{ft}}{N_{mt} l_{ft}} = \frac{1}{N_{mt}}$. Therefore, the Herfindahl-Hirschman Index corresponds to the inverse number of firms in the market,

$$\text{HHI}_{mt} = \frac{1}{N_{mt}},$$

while productivity is given by

$$\text{AMRPL}_{mt} = \theta A_{mt} l_{mt}^{\theta-1},$$

where $l_{mt} = \frac{L_{mt}}{N_{mt}}$ denotes the number of workers employed by the representative firm. Therefore, the market equilibrium (11) can be rewritten as

$$\beta_{mt} L_{mt}^{\tau} \left(1 + \tau \frac{1}{N_{mt}}\right) = \theta A_{mt} \left(\frac{L_{mt}}{N_{mt}}\right)^{\theta-1},$$

so that total employment is given by

$$L_{mt} = \left[\left(\frac{1}{N_{mt}}\right)^{\theta-1} \left(1 + \tau \frac{1}{N_{mt}}\right)^{-1} \frac{\theta A_{mt}}{\beta_{mt}}\right]^{\frac{1}{\tau+1-\theta}}$$

and firms' employment is

$$l_{mt} = \left[\left(\frac{1}{N_{mt}} \right)^{\tau} \left(1 + \tau \frac{1}{N_{mt}} \right)^{-1} \frac{\theta A_{mt}}{\beta_{mt}} \right]^{\frac{1}{\tau+1-\theta}}.$$

Because $\theta \leq 1$, L_{mt} is decreasing with respect to $\text{HHI}_{mt} = \frac{1}{N_{mt}}$. In particular,

$$\frac{\partial \log L_{mt}}{\partial \frac{1}{N_{mt}}} = -\left(\frac{\tau(2-\theta)\frac{1}{N_{mt}} + (1-\theta)}{(\tau+1-\theta)(\frac{1}{N_{mt}}(1+\tau\frac{1}{N_{mt}}))} \right) \leq \chi \simeq -1, \tag{14}$$

where χ is a constant that is approximately -1 if $\tau \frac{1}{N_{mt}} \simeq 0$ (as assumed in Section 4, since τ is estimated to be small). However, firms' individual employment increases with HHI_{mt}, since

$$\frac{\partial \log l_{mt}}{\partial \frac{1}{N_{mt}}} = \frac{\tau}{\tau+1-\theta} \left(\frac{1+\frac{1}{N_{mt}}(\tau-1)}{\frac{1}{N_{mt}}(1+\tau\frac{1}{N_{mt}})} \right) \geq 0, \tag{15}$$

which follows from $\frac{1}{N_{mt}} \leq 1$. In this model, an increase in labor market concentration is associated with a decrease in market employment and the number of firms, and the latter decreases faster than the former, so that each firm's number of workers increases as a consequence.

Now,

$$\log \text{AMRPL}_{mt} = \log \theta A_{mt} - (1-\theta) \log l_{mt}$$

or

$$\log \text{AMRPL}_{mt} = \omega \log \theta A_{mt} + (1-\omega) \log \beta_{mt} - (1-\omega) f\left(\frac{1}{N_{mt}}\right), \tag{16}$$

where

$$\omega = \frac{\tau}{\tau+1-\theta}$$

and

$$f\left(\frac{1}{N_{mt}}\right) = \left[\tau \log \left(\frac{1}{N_{mt}} \right) - \log \left(1 + \tau \frac{1}{N_{mt}} \right) \right],$$

a function that is increasing in $\text{HHI}_{mt} = \frac{1}{N_{mt}}$. The first-order Taylor expansion for this function is

$$f(\text{HHI}_{mt}) \simeq \tau \log \overline{\text{HHI}} + \tau \frac{\text{HHI}_{mt} - \overline{\text{HHI}}}{\overline{\text{HHI}}} - \log(1+\tau\overline{\text{HHI}}) - \tau \frac{\text{HHI}_{mt} - \overline{\text{HHI}}}{1+\tau\overline{\text{HHI}}},$$

where $0 < \overline{\text{HHI}} < 1$ is a small constant around which we expand and, since τ is estimated to be small in Section 4, we assume that $\tau\overline{\text{HHI}} \simeq 0$. Therefore,

$$
\begin{aligned}
f\left(\text{HHI}_{mt}\right) &\simeq \tau \log \overline{\text{HHI}} + \tau \left(\frac{1}{\overline{\overline{\text{HHI}}}} - 1\right) \text{HHI}_{mt} \\
&= \tau(\psi_1 + \psi_2 \text{HHI}_{mt}),
\end{aligned}
$$

$$
\begin{aligned}
\psi_1 &= \log \overline{\text{HHI}} \geq 0, \\
\psi_2 &= \left(\frac{1}{\overline{\overline{\text{HHI}}}} - 1\right) \geq 0
\end{aligned}
$$

and

$$
\log \text{AMRPL}_{mt} \simeq \omega \log \theta A_{mt} + (1-\omega) \log \beta_{mt} - (1-\omega)\tau(\psi_1 + \psi_2 \text{HHI}_{mt}). \tag{17}
$$

With decreasing returns to scale ($\theta < 1$, $\omega < 1$), a positive supply (amenity) shock – that is, a reduction in the intercept β_{mt} – leads to lower average productivity in the market: firms can now hire more workers for the same wage, but these workers are marginally less productive. Similarly, markets with high HHI_{mt} have firms which are larger in size (see equation (15)) and, hence, with decreasing returns to scale, that are less productive on average.

D.2 Endogeneity of HHI_{mt} in the Asymmetric Firms Model

In this section, I show that the asymmetric firms model presents an additional source of endogeneity in HHI_{mt} with respect to those highlighted in Section 2.4. Indeed, if a positive productivity shock hits market m, then workers are paid higher wages ($\uparrow W_{mt}$), and if large firms benefit relatively more from the shock, these firms grow in size and the market becomes more concentrated, that is, $\uparrow \text{HHI}_{mt}$. If, instead, small firms benefit relatively more from the productivity shock, then labor market power is reduced ($\downarrow \text{HHI}_{mt}$), as larger firms lose part of their dominant position. In both cases, labor market concentration and wages correlate for reasons other than the causal relationship between the two variables.

Without loss of generality, assume that market m has only two firms, f and j, and that f is the dominant firm, that is, $s_f > s_j$, with $s_f + s_j = 1$. By dividing the individual first-order conditions (10) of the two firms, and by assuming constant returns to scale without loss of generality, we get

$$
\frac{\cancel{W_{mt}}\,(1 + \tau s_{ft})}{\cancel{W_{mt}}\,(1 + \tau s_{jt})} = \frac{\cancel{A_{mt}}\,\epsilon^A_{fmt}}{\cancel{A_{mt}}\,\epsilon^A_{jmt}}
$$

Then, any asymmetric productivity shock that increases ϵ^A_{fmt} more than ϵ^A_{jmt} leads to an increase in s_f and a decrease in s_j. Since s_f was greater than s_j to start with, and since

$$\text{HHI}_{mt} = s_f^2 + s_j^2,$$

we have that labor market concentration increases as a consequence of the productivity shock.[35] It is also easy to see that average productivity in the market

$$\text{AMRPL}_{mt} = A_{mt}(s_f \epsilon_{fmt}^A + s_j \epsilon_{jmt}^A)$$

increases following the shock. This, by equation (11), puts upward pressure on W_{mt}. In other words, the asymmetric productivity shock induces a positive correlation between HHI_{mt} and W_{mt}, which goes in the opposite direction to the causal effect between the two variables highlighted in equation (11). Although this section focuses on the example of a positive productivity shock, the same endogeneity concerns arise in the case of *negative* changes in market productivity that asymmetrically impact firms.

D.3 Estimation Strategy

In this section, I present a strategy to estimate the *agglomeration elasticity* by examining the linear relationship between firms' log productivity and the log population density of the city in which they operate. I also emphasize the potential for estimation bias when failing to account for variables that influence wages and systematically vary with city size – most notably, labor market concentration.

D.3.1 Constant Returns to Scale

D.3.1.1 Controlling for Labor Market Concentration

We assume that firms are symmetric and use a constant returns to scale technology. Then, $\theta = \omega = 1$ and, from equation (16),

$$\log \text{AMRPL}_{mt} = \log A_{mt}, \tag{18}$$

i.e., revenue productivity in the market fully reflects the linear term in firms' productivity.[36] We know that $\log A_{mt}$ has the following functional form:

[35] The opposite would happen ($\downarrow \text{HHI}_{mt}$) if, following the productivity shock, ϵ_{jmt}^A were to increase more than ϵ_{fmt}^A.

[36] For simplicity, I focus on the symmetric firms case. As shown in equation (12), with asymmetric firms and constant returns to scale ($\theta = 1$), I would obtain an analogous expression to (18) if I assume that

$$\sum_{f=1}^{N_{mt}} s_{ft} \epsilon_{fmt}^A = 1.$$

Remembering that $A_{fmt} = A_{mt} \epsilon_{fmt}^A$, this assumption is intuitively guaranteeing that the employment

$$\log(A_{mt}) = \log(A_m) + \log(A_t) + \epsilon^A_{mt}, \tag{19}$$

which is the time-varying version of equation (2) in the symmetric firms case. In equation (19), A_m denotes productivity in the market, A_t is the overall productivity time trend, and ϵ^A_{mt} is the variation in productivity that is left once these components are partialled out. Similarly, and remembering that markets are city-cluster combinations (indexed by c and k, respectively), we assume that

$$\log(A_m) = \log(A_c) + \log(A_k) + \epsilon^A_m.$$

Finally, from equation (3), we know that the log productivity of city c, $\log(A_c)$, is a linear function of the city's log employment level. For example, this may be the case because the proximity of workers and firms facilitates the generation of new ideas:

$$\log(A_c) = \log(A) + \delta \text{logCitySize}_c + \epsilon^A_c,$$

where δ denotes the *agglomeration elasticity*, and CitySize_c proxies for total employment – which delivers a functional form that is in line with the literature studying agglomeration effects (De la Roca and Puga, 2017). Even if $\log \text{AMRPL}_{mt}$ is unobservable (or partially unobservable), the structure of the problem is sufficiently simple to allow us to estimate δ from data on logCitySize_c, W_{mt}, HHI_{mt} and some market controls X_{mt} that capture potentially important features of the market that are not modeled explicitly (e.g., the degree of unionization of workers in the labor market or the extent of product market power).

Indeed, using (18), we can rewrite equilibrium equation (11) in logs as

$$\log \text{W}_{mt} = \log A_{mt} - \tau \log \text{HHI}_{mt} + \upsilon_{mt},$$

where υ_{mt} is the sampling error and $\log(1 + \tau \log \text{HHI}_{mt}) \simeq \tau \log \text{HHI}_{mt}$ as $\tau \overline{\text{HHI}} \simeq 0$. We can thus estimate

$$\textbf{Step 1:} \quad \log \text{W}_{mt} = \alpha_m + \alpha_t + \tau \log \text{HHI}_{mt} + \alpha \text{X}_{mt} + \varepsilon_{mt}, \tag{20}$$

where the market fixed effect α_m captures $\log(A_m)$ and the time fixed effect α_t cap-

weighted average of firms productivity in the market is equal to the market component of productivity, i.e., that

$$\sum_{f=1}^{N_{mt}} s_{ft} A_{fmt} = A_{mt}.$$

tures $\log(A_t)$. Also, $\varepsilon_{mt} = \upsilon_{mt} + \epsilon^A_{mt}$, and we assume that

$$\mathbb{E}[\varepsilon_{mt}|\alpha_m, \alpha_t, \mathrm{X}_{mt}, \mathrm{HHI}_{mt}] = 0. \tag{21}$$

This implies that the $\tau < 0$ coefficient is identified by the part of HHI_{mt} variation that is not determined by the market and overal time trend components of firms' productivity.[37] The market fixed effect estimated in (20) can be rewritten as

$$\textbf{Step 2:} \quad \alpha_m = \alpha_k + \delta \mathrm{logCitySize}_c + \upsilon_m, \tag{22}$$

where $\alpha_k = \log(A) + \log(A_k)$ and $\upsilon_m = \varepsilon^A_c + \varepsilon^A_m$. Thus, the parameter of interest δ could be readily estimated in *step two* (regression (22)), where we substitute the dependent variable α_m for the $\hat{\alpha}_m$ estimated in *step one* (regression (20)), were it not for the fact that

$$\mathbb{E}[\upsilon_m|\alpha_k, \mathrm{logCitySize}_c] \neq 0.$$

The reason why strict exogeneity fails in equation (22) is that $\log(A_c)$, contained in the dependent variable α_m, likely causes $\mathrm{logCitySize}_c$ – as workers are attracted to migrate to high-productivity, high-paying cities, and this creates a reverse causality problem. In Section 4.3, we deal with this identification problem with an IV based on the historical determinants of population density, which are plausibly unrelated to time t productivity.

D.3.1.2 Not Controlling for Labor Market Concentration

If we estimate the agglomeration elasticity in the same way as highlighted in the previous section, with the only difference that we do not control for HHI_{mt} in *step one* (equation (20)), then the agglomeration elasticity estimate is likely to be biased. The bias will be present as long as labor market concentration is relevant ($\tau \neq 0$) and is correlated with city size. In this section, I quantify the extent of the bias and show that, given the correlation between HHI_{mt} and city population density observed in the data, failing to account for systematic differences in labor market concentration across markets leads to an *overestimation* of the agglomeration elasticity.

First, we decompose HHI_{mt} into its market component h_m, general time-trend component h_t, and residual variation ϵ^h_{mt}:

$$\mathrm{HHI}_{mt} = h_m + h_t + \epsilon^h_{mt}. \tag{23}$$

This is similar to the decomposition of the market level productivity A_{mt} in equa-

[37] An example of such exogenous variation in HHI_{mt} are shocks to fixed costs of production unrelated to market productivity and affecting firm entry.

tion (19). The h_m and h_t components correspond to market and time fixed effects in a regression that has the form of equation (23). These fixed effects capture both the variation in HHI_{mt} that is endogenous to the productivity terms $\log(A_m)$ and $\log(A_t)$, and the variation that is unrelated to productivity and is shared across markets and/or time within the same market. On the other hand, ϵ^h_{mt} corresponds to the variation in HHI_{mt} that, given assumption (21), is fully idiosyncratic and unrelated to productivity. It should be noted that this variation identifies the τ parameter in equation (20) and that condition (21) implies that $\mathbb{E}(\epsilon^A_{mt}\epsilon^h_{mt}) = 0$.

We posit the following functional form for h_m:

$$h_m = h + h_k + \lambda \text{logCitySize}_c + \epsilon^h_m. \tag{24}$$

In the data, λ is estimated to be negative, since markets in larger cities attract more firms and are thus systematically less concentrated (see Figure 1). In this context, if we apply the same two-step procedure of equations (20) and (22) without controlling for HHI_{mt} in *step one*, we obtain an upward biased estimate of the agglomeration elasticity:

$$\log(W_{mt}) = \underbrace{\alpha_m}_{\log(A_m)+\tau h_m} + \alpha_t + \alpha \text{X}_{mt} + \underbrace{\varepsilon_{mt}}_{\epsilon^A_{mt}+\tau\epsilon^h_{mt}+\upsilon_{mt}}$$

$$\hat{\alpha}_m = \alpha_k + (\delta + \underbrace{\tau\lambda}_{>0})\text{logCitySize}_c + \underbrace{\upsilon_m}_{\epsilon^A_m+\tau\epsilon^h_m+\epsilon^A_c}$$

Calling $\hat{\delta}$ the agglomeration elasticity that we estimate when we control for HHI_{mt}, and $\hat{\delta}^{pc}$ the elasticity estimated when we do not control for it, the extent of the bias can be estimated as

$$\frac{\hat{\delta}^{pc} - \hat{\delta}}{\hat{\delta}^{pc}} \longrightarrow \frac{\tau\lambda}{\delta + \tau\lambda}.$$

This can also be interpreted as the percentage of the city-size wage premium that can be explained by labor market concentration differences across cities, and not by agglomeration economies. Note that the bias is substantial if $\tau\lambda$ is large with respect to δ, and disappears if either $\tau = 0$ (i.e., there is no labor market power) or $\lambda = 0$ (i.e., labor market power is not systematically related to city size).

D.3.2 Decreasing Returns to Scale

With decreasing returns to scale ($\omega < 1, \theta < 1$), $\log(\text{AMRPL}_{mt})$ is a function of $\log(A_{mt})$, $\log(\beta_{mt})$ and HHI_{mt} (see equation (17)). To the extent that we control for HHI_{mt} in *step one* (regression (20)), the fact that labor market concentration affects wages not only directly but also through $\log(\text{AMRPL}_{mt})$ changes our interpretation of some estimates but does not introduce any additional source of bias. However, the fact that $\log(\beta_{mt})$

may systematically vary across cities of different sizes is a source of concern. To observe this, let us assume the usual decomposition for $\log(\beta_{mt})$:

$$\begin{aligned} \log(\beta_{mt}) &= \log(\beta_m) + \log(\beta_t) + \epsilon^{\beta}_{mt}, \\ \log(\beta_m) &= \log(\beta_c) + \log(\beta_k) + \epsilon^{\beta}_{m}, \\ \log(\beta_c) &= \log(\beta) + \rho \text{logCitySize}_c + \epsilon^{\beta}_{c}. \end{aligned}$$

In principle, we do not know if $\rho > 0$, $\rho < 0$ or $\rho = 0$; that is, if amenities are, on average, lower in bigger cities, higher in bigger cities, or unrelated to city size. Substituting in the log version of equation (11) and using the usual Taylor approximation, we now get

$$\begin{aligned} \log \mathrm{W}_{mt} = & \underbrace{(\omega-1)\tau\psi_1 + \omega \log\theta + \omega\log A_m + (1-\omega)\log\beta_c}_{\alpha_m} + \underbrace{\omega\log A_t + (1-\omega)\log\beta_t}_{\alpha_t} + \\ & -\tau(1+\underbrace{(1-\omega)\psi_2}_{>0})\mathrm{HHI}_{mt} + \alpha \mathrm{X}_{mt} + \underbrace{\omega\epsilon^{A}_{mt} + (1-\omega)\epsilon^{\beta}_{ct} + \upsilon_{mt}}_{\varepsilon_{mt}}. \end{aligned}$$

Two remarks are in order. First, fixed effects are now a combination of constants and of weighted averages of the productivity and amenity terms. Second, decreasing returns to scale magnify the effect of labor market concentration on wages.

If amenities are not controlled for and we proceed with *step two* of the estimation procedure, we get

$$\hat{\alpha}_m = \alpha_k + (\omega\delta + (1-\omega)\rho)\text{logCitySize}_c + \upsilon_m.$$

As long as $\rho \neq 0$, the agglomeration elasticity is biased, and the direction of the bias depends on the sign of ρ. Because $\log(\beta_c) = b - b_c$, where b_c are city amenities, we can avoid this bias by controlling for city amenities in *step two*:[38]

$$\hat{\alpha}_m = \alpha_k - (1-\omega)b_c + \omega\delta\text{logCitySize}_c + \upsilon_m,$$

and the agglomeration elasticity δ is identified up to the constant ω.

Finally, let $\hat{\delta}^{b,pc}$ and $\hat{\delta}^{b}$ be the agglomeration elasticities estimated by controlling for amenities and for both HHI_{mt} and amenities, respectively. Then, the extent of the different kind of biases described can be estimated as

$$\frac{\hat{\delta}^{pc} - \hat{\delta}^{b}}{\hat{\delta}^{pc}} \longrightarrow \frac{(1-\omega)\rho + \tau\lambda}{\omega\delta + (1-\omega)\rho + \tau\lambda} \tag{25}$$

[38]In the general equilibrium model of Section 2, we have $\log(\beta_{c(c')k}) = g_{c'k} - b_c$ (see equation (1)). Here, we are assuming that workers take the attractiveness $g_{c'k}$ of all other cities in the economy as given when evaluating city c amenities in period t, i.e. $g_{c'k} = b$.

and

$$\frac{\hat{\delta}^{b,pc} - \hat{\delta}^{b}}{\hat{\delta}^{b,pc}} \longrightarrow \frac{\tau\lambda}{\omega\delta + \tau\lambda}.$$

Equation (25) shows how the correlation between city amenities and population influences the direction of bias in estimating the agglomeration elasticity. If amenities are, on average, lower in large urban areas (e.g., because of lower air quality), then $\rho < 0$ and the bias increases. In this case, part of the urban earnings premium acts as compensation for individuals to live and work in larger cities despite the higher disamenity levels, and is not related to agglomeration economies. Conversely, if amenities are higher in larger urban areas, then $\rho > 0$ and the bias decreases. If amenities are not related to city size ($\rho = 0$), instead, they do not bias the agglomeration elasticity.

Additionally, the bias grows with the degree of decreasing returns to scale in the economy, which affects ω. When returns to scale are lower, a positive supply (amenity) shock reduces average productivity in the market, as firms hire more workers at the same wage, but these workers are marginally less productive. Since lower productivity translates into lower wages, differences in amenities between smaller and larger cities have a greater potential to explain the wage gap observed in the data. Ignoring urban amenities in such cases would result in a larger bias in the estimated agglomeration elasticity. In contrast, under constant returns to scale ($\omega = 1$), amenities have no effect on wages and therefore cannot bias the agglomeration elasticity.