Informality and the City Size Wage Premium: Evidence from Brazil[*]

Sophie Nottmeyer [†]
CEMFI

July 2022

Abstract

Cities in middle and low income countries are rapidly urbanizing, generating important economic opportunities and challenges. This paper analyzes how pervasive informality shapes the distribution of productivity gains typically associated with larger cities in Brazil. Using Census data, it shows that informality attenuates the urban wage premium by generating differential returns to density across worker types. Decomposing employment density into formal and informal components, the analysis reveals that formal workers experience strong wage gains from formal density, which are partially offset by a negative effect from informal density, while informal workers only benefit from formal density. These findings suggest that productive advantages of cities are primarily driven by the presence of formal workers, and that informality alters the composition of productivity spillovers in urban labor markets of emerging economies.

[*]I gratefully acknowledge financial support from Spain's State Research Agency through its María de Maeztu Units of Excellence program (MDM-2016-0684). All remaining errors are mine.

[†]CEMFI, Calle Casado del Alisal, 5, 28014 Madrid. E-Mail: sophie.nottmeyer@cemfi.edu.es

1 Introduction

Cities in middle and low income countries are rapidly urbanizing. By 2050, the majority of the world's urban population will live in developing countries, generating considerable economic opportunities and challenges (Bryan et al., 2019; Glaeser and Henderson, 2017). Among potential benefits, a core result in the urban economics literature states that individual earnings are higher in bigger cities. However, recent empirical efforts to revisit the so-called city size wage premium in less developed countries do not take into account important features of the local context. Informality is among the most prominent overlooked characteristics, despite its pervasiveness. In Brazil, where I draw my data from, over 30 percent of the entire workforce is employed informally, which is likely to have deep implications for a range of economic phenomena.

In particular, segmentation of labor markets in terms of informality may hinder urban development. For example, the lack of formal contracts may limit incentives for workers to improve their skills, thus limiting the scope of agglomeration benefits through knowledge spillovers in cities. It may also introduce important search and matching frictions in the allocation of workers to employers that interact with city size. Regardless of the mechanism, it is not clear that formal and informal workers would benefit and contribute to productive advantages of bigger cities in the same way.

How does informality affect the relationship between wages and city size? In this paper, I study this question by analyzing the variation in formal and informal wages across cities in Brazil, a highly urbanized context with a history of widespread informality. I find that higher levels of informality are associated with a lower city size wage premium. Figure 1 illustrates this result using raw data. Specifically, it plots the distribution of wages in cities above and below median density and informality. We can see that wage distributions for both formal and informal workers shift to the right when going from a high informality context to lower levels of informality, both comparing low and high density cities, respectively. Importantly, this shift is larger for big cities.

This paper extends a standard two-step estimation procedure to measure the city size wage premium in the presence of informality. The main data source is the 2000 Population Census in Brazil. The data allows me to observe wages and the registration status of the universe of wage earners. The latter is defined in terms of access to social security and measured by the (reported) existence of a 'signed work card' required to claim the benefits of contributory contracts. The census also provides information on commuting flows between municipalities, which I use to construct my spatial units of analysis based on a local labor market approach. For this, I am using an aggregation algorithm based on

Duranton (2015) that joins basic spatial units connected by a commuting flow of at least 10 percent of origin working population. The analysis works with cities of more than 50,000 residents that consist of at least one municipality with a density above 100 inhabitants per km^2.

First, I estimate an individual wage equation with city fixed effects in order to isolate spatial variation in wages from the influence of worker characteristics. Second, I relate the estimated city fixed effects to city size as measured by total employment density, which gives me a simple OLS estimate of the static advantages of bigger cities. I repeat these steps for both sub-samples of formal and informal workers to explore a potential heterogeneity in the effect of city size by group. My main results are derived from the following two additional specifications: I separately introduce formal and informal employment density to investigate the contribution of each by worker group, and then I repeat the analysis interacting total employment density with the local informality share to directly test the effect of informality on the city size wage premium.

I find that location matters for individual wage disparities in Brazil. In particular, city size is a strong predictor of spatial variation in earnings, even for observationally equivalent workers. The elasticity of wages with respect to total employment density is estimated at 0.0539 for the pooled sample, consistent with estimates in the literature. Even though the effect of total employment density appears to be the same in magnitude for both formal and informal workers, I show that density means different things for each group, as different types of workers experience the concentration of formal and informal workers differently.

In particular, the net effect of density for formal workers is composed of a strong positive impact of formal density that is partially offset by a negative effect stemming from an increased concentration of informal workers. The effects for informal workers are not symmetric: they also benefit from an increased density in terms of formal workers (although to a lesser extent), but their earnings are not sensitive to density from their peers in the informal sector. Hence, productivity advantages of larger cities seem to be mainly driven by the presence of formal workers.

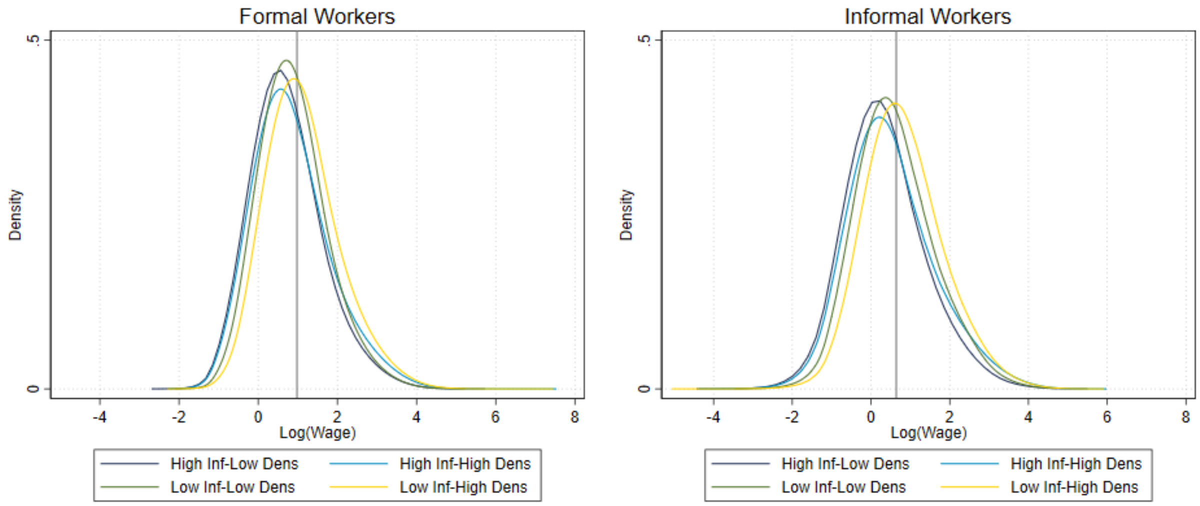
This suggests that a higher share of informal workers in the local labor market reduces the city size wage premium, both because there are relatively less formal workers around that productively contribute to agglomeration economies and because the presence of informal workers depresses formal wages. Indeed, the negative and significant coefficient on the interaction term indicates that the higher the extent of informality in a given city, the lower the benefits from city size.

This paper mainly relates to the empirical literature on the productivity benefits of bigger cities, going back to at least Ciccone and Hall (1996). Many papers have since estimated the city size wage premium in different contexts (see Ahlfeldt and Pietrostefani 2019 for a recent survey), with the following key advances in terms of its understanding. Glaeser and Mare (2001) and, more recently, Combes et al. (2008) introduce worker fixed effects to show that standard approaches overestimate the effect of city size due to worker sorting on unobserved ability into bigger cities. More recent contributions demonstrate that sorting on unobservables may contribute relatively little to the observed city size wage premium and instead highlight the role of additional effects, e.g Baum-Snow and Pavan (2012) using a structural model. De La Roca and Puga (2017) demonstrate that in addition to static advantages from working in bigger cities, there are equally important dynamic benefits resulting from the accumulation of experience. This paper contributes to this line of research by showing how the city size wage premium depends on the specific composition of the workforce that populates a city, especially when there are different types of contracts introducing frictions in how workers benefit from and contribute to productive advantages in cities.

Within this broader literature, this paper also contributes to a smaller set of papers revisiting the estimation of productivity benefits for the case of developing countries, such as Duranton (2016) for Colombia and Chauvin et al. (2017) for cities in Brazil, China and India. Specifically, I provide additional evidence for Brazil, which is a country that has been extensively studied in the development literature researching causes and consequences of informality. The papers most closely related to this analysis are Matano et al. (2020), who estimate the city size wage premium in Ecuador, and Bernedo Del Carpio and Patrick (2021), who use establishment-level data to estimate productivity advantages of bigger cities in Peru. With respect to these papers, I go beyond simply differentiating the estimated density elasticity for each group of workers by exploring the heterogeneous effect of their contribution to density and explicitly taking into account local informality levels. To the best of my knowledge, this is the first paper to document spatial patterns of wages in relation to informality in this way.

The remainder of this article is organized as follows. Section 2 discusses the institutional setting to clarify what informality means in this context. Section 3 provides some thoughts on how we can think about the different ways in which informality may affect the city size wage premium. In section 4, I present the data and describe how the sample is selected, including a discussion of how I delineate urban areas. Section 5 outlines the empirical framework and section 6 presents the empirical results, followed by a discussion of robustness, potential mechanisms and limitations of the analysis. Section 7 concludes with policy implications and suggestions for future research.

Figure 1. Distribution of Wages Above and Below Median Density and Informality



Note: Distribution of the logarithm of individual nominal hourly wages for formal and informal workers in cities above and below the median density and median share of informal workers. Survey sampling weights used. Data drawn from the 2000 Brazilian Population Census, see Section 4.

2 Labor Market Regulations and Informality in Brazil

In order to investigate the role of informality in shaping benefits from working in bigger cities, we first need to understand what informality means in the present context. Brazil has one of the least flexible labor markets in the world, resulting in high levels of labor informality as defined by the lack of access to social insurance. Despite efforts to enforce regulations and reduce informality starting in the 1990s, around 35 percent of the entire workforce were employed informally, out of which around 40 percent by formal firms, in the beginning of the 2000s (Ulyssea, 2018).

As part of the labour regulations in Brazil, formal workers are required to hold a booklet (*carteira de trabalho*) issued by the Ministry of Labour that must be signed by the employer and that contains workers' entire formal employment histories. Having the labour contract registered in this booklet means that contributions to social insurance are being made on behalf of the worker and guarantees access to all social insurance benefits (such as pension system, unemployment insurance, severance payments, sickness and maternity leave, and paid vacation). In addition to social security contributions, a formal contract entails additional labor taxes (such that the total tax rate is around 42.1 percent as a share of commercial profits), hiring and firing costs and the obligation to follow the federal minimum wage regulation.

In addition to labour informality, firms and entrepreneurs may choose to evade tax obligations altogether. In Brazil, informal firms can be defined as firms that are not registered with tax authorities, which means that they do not possess the tax identification number required for Brazilian firms (*Cadastro Nacional de Pessoa Juridica*, CNPJ).[1]

Importantly, there is significant overlap in the data: informal workers may be found in both formal or informal firms. Although the formal sector (workers and firms) tends to be more productive, you can find workers with the same skill level in both formal or informal jobs, in the same way as firms of the same productivity level coexist within narrowly defined industries with a different registration status (Meghir et al., 2015; Ulyssea, 2018).

3 How Informality May Affect Theoretical Mechanisms

Having discussed what informality means in this context, this section outlines different ways in which informality may inhibit or enhance productivity advantages of bigger cities through the lens of the main theoretical mechanisms: sharing, matching and learning. The main idea is that the segmentation between sectors in terms of the different (legal) constraints that economic actors face in each sector has implications for how they can benefit from and contribute to productivity benefits of larger cities.

Duranton and Puga (2004) distinguish three main categories of mechanisms that are thought to give rise to productive advantages of cities. First, a larger market allows for a more efficient *sharing* of a pool of workers. Second, bigger cities also allow for better *matching* between employers and employees, or buyers and suppliers. Finally, a larger market can facilitate *learning* through the transmission and accumulation of skills or by promoting the adoption of new technologies and business practices.

Although hiring informally in principle may allow firms to share a large labour pool more flexibly, the practice may have adverse implications for gains from specialization and other mechanisms because of an under-investment in on-the-job training of informal workers. For example, Ponczek and Ulyssea (2022) show that greater flexibility introduced by informality allows both formal firms and low-skill workers to cope better with adverse labor market shocks in Brazil. However, informal workers can be fired just as ‘flexibly’, discouraging task specialization and investment in on-the-job training of informal workers on the firm side.

Similarly, informality may re-introduce hold-up problems mitigated by larger markets in terms of human capital investment between employer and employees, but also in the

[1] Even though firm informality is an important margin, the analysis mainly focuses on labour informality as productivity advantages of cities are measured using individual-level data.

more traditional sense between buyers and suppliers. In particular, informal workers may be discouraged to invest in their human capital ex ante because they have lower bargaining power. On the other hand, the inability of informal firms to sign business contracts discourages vertical linkages between formal and informal firms, thereby introducing frictions in the matching of buyers and suppliers.

In addition to affecting the chances of matching, informality may also impact the quality of matches. Although competition for limited formal sector jobs may have a positive effect, search frictions likely decrease match quality. One way to think about this is that informal job opportunities are likely not officially advertised but allocated through social networks.[2] In particular, Meghir et al. (2015) show that informality leads to an increased allocation of workers to low productivity jobs. Hence, a higher extent of informality could off-set the benefits from better matches in larger cities.

Finally, advantages of cities in terms of learning externalities may be the most affected by informality given the previous discussion. If informal workers do not (have an incentive to) accumulate human capital, in general or on-the-job, they do not provide learning spillovers to formal workers nor may they learn from others around them. Further, limited overlap between formal and informal workers in the workplace or at residence may imply limited face-to-face interactions that are thought to produce social learning. At the firm level, lack of access to credit may result in a large gap in technology and management practices across sectors, such that informal firms have limited capabilities to absorb spillovers from formal firms, putting them at a productivity disadvantage despite being located in larger markets.

In line with this discussion, Matano et al. (2020) find suggestive evidence that matching and learning channels play a role in generating different spatial wage premia, exploiting job transitions in a panel data set from Ecuador. Both are more relevant for formal than informal sector wages, and matching quality seems to be more important but there is also evidence of learning effects.

4 Data and Descriptive Statistics

The main data source is the 2000 Brazilian Population Census conducted by IBGE (*Instituto Brasileiro de Geografia e Estatística*).[3] In addition to individual characteristics, the

[2] In a recent paper Bobba et al. (2022) develop an equilibrium model with search and matching frictions that is able to replicate the main features of labor markets with high informality, supporting the presence of different types of contracts in equilibrium.

[3] Microdata is publicly available here. In addition, I used the `datazoom` Stata programs developed by the Department of Economics at PUC-Rio to read in the microdata.

survey allows me to compute wages and determine the formality status of all wage earners in the 10% census sample (20.6 million individuals). The Census further provides information on commuting flows between municipalities, which I use to define the cities in my sample, as described below. To illustrate the kind of variation I am using in my empirical analysis, I document that informality is not equally distributed in space.

4.1 Delineating Cities

I use a local labor market approach based on commuting flows between municipalities and an iterative aggregation algorithm to construct my spatial units of analysis. Cities are defined as urban areas that reach a conventional population and density threshold. This process yields a total of 202 cities, containing 697 out of 5,507 municipalities, that account for 56 percent of Brazil's total population and roughly 3 percent of its surface.

The types of interactions I am interested in in my analysis do not stop at administrative boundaries, but require an appropriate definition of economically integrated units. Building on the insights from a recent discussion in the literature on the delineation of urban areas, I define cities as local labor markets using commuting flows (Duranton, 2021).[4] In practice, I use an iterative algorithm based on Duranton (2015) to aggregate *municipios* into endogenously defined urban areas subject to a minimum commuting threshold of 10 percent of the working population at origin.[5]

I initially compute commuting flows for all municipality pairs in the sample. At each iteration, the algorithm picks the pair with the strongest commuting tie, aggregates them into a new unit and recomputes commuting flows. In that way, it recursively joins units to the local labor market with which they share the strongest commuting tie that exceeds the minimum threshold until there are not more units that satisfy this criterion.[6]

Doing this results in 4,905 urban areas considered local labor markets, the majority of which are individual municipalities. I restrict my sample of spatial units to those local labor markets that have a total population above 50,000 and consist of at least one municipality with a density of at least 100 inhabitants per km^2, based on a universally accepted definition of cities (Dijkstra et al., 2021).

[4] In particular, I exploit a question in the Census survey that asks each respondent in which municipality, state or foreign country they work or study (although I only consider those that work).

[5] While threshold choice is inherently arbitrary, I follow Dingel et al. (2021) in applying the same 10 percent threshold to the case of Brazil that Duranton (2015) initially proposed for Colombia.

[6] The algorithm does not impose contiguity, which is not a problem for the analysis. However, there are 18 cases where the algorithm would join municipalities that are further than 150 km apart (around 1000km, on average). These are mainly very small origin municipalities that are connected to a large metropolitan area (Sao Paolo or Rio de Janeiro) in the data. I treat these exceptions as outliers by preventing these aggregations to happen within the loop.

Table A1 lists the 20 largest cities in my sample. Figure A1 shows all sample cities in a map, shaded differently according to their total population. Importantly, one can see that many of the units I identify as economically integrated local labor markets combine multiple municipalities.[7]

Information on the surface area at the municipality level used to compute density is provided by the Institute for Applied Economic Research (IPEA) Brazil.[8]

4.2 Worker Sample Selection

The sample based on which I estimate wage levels across locations results from standard restrictions according to location of residence, demographics, employment status and sector of economic activity.

The starting sample includes 20,274,411 observations. I subset the data to respondents located in the cities identified above using municipality of residence. In order to ensure that I only include those who are most likely working in the urban parts of the municipality, I restrict the sample to respondents reported to live in *'zonas urbanas'*, since municipalities in Brazil can be quite big with stretches of very sparsely populated areas. This yields an initial sample of 9,302,402 observations.

Next, I focus on the working age population (15 to 65 years) that reports non-zero earnings for their main job and worked at least 20 hours in the reference week. Crucially, I focus on wage earners only, excluding the self-employed (or own-account workers) and micro-entrepreneurs because I cannot disentangle wages from profits.[9] This yields a sample of 2,568,409 observations.

Finally, I drop workers in agriculture and mining sectors because they are typically found in rural areas. I further exclude public officials because their wages tend to be determined by different forces than the ones I am analyzing here. Then, the final sample consists of 2,421,291 observations.[10]

As discussed in Section 2, a worker is defined as formal if they report having a worker's card signed by the firm.[11] Table A2 shows the individual characteristics of all workers

[7] Interestingly, defining cities based on commuting flows in this way does not correspond with the boundaries of 'micro-regions' provided by IBGE although they are supposed to reproduce the idea of local economies similar to commuting zones in the US.

[8] Accessed here.

[9] My definition of wage earner further excludes unpaid family helpers, interns and subsistence workers.

[10] In the estimation below I include dummies for ethnicity, which is missing in some cases, which is why the number of observations reported there is 2,408,145.

[11] The exact question I use to determine registration status is number 4.47. Accordingly, I define workers in the following two categories as informal: 2 - Domestic worker without signed work card (*Trabalhador domestico sem carteira de trabalho assinada*) and 4 - Employee without signed work card (*Empregado sem carteira de trabalho assinada*).

in my sample by registration status. Formal and informal workers are different in many aspects. For example, informal workers are younger, more likely to be female, less educated, working in service and sales, and private household employees.

Summing over the number of formal and informal workers in a city then gives what I consider total employment. Note that survey sampling weights are used whenever computing population counts. In my analysis below, I measure city size in terms of employment density, which is the number of all workers over the surface area of a city.[12] Finally, the informality share of a city is the number of informal workers over total employment. The average informality share in my sample is 36 percent.

Finally, I compute nominal hourly wages from monthly earnings and weekly hours for the reference week in the main job.

4.3 Distribution of Informality Across Cities in Brazil

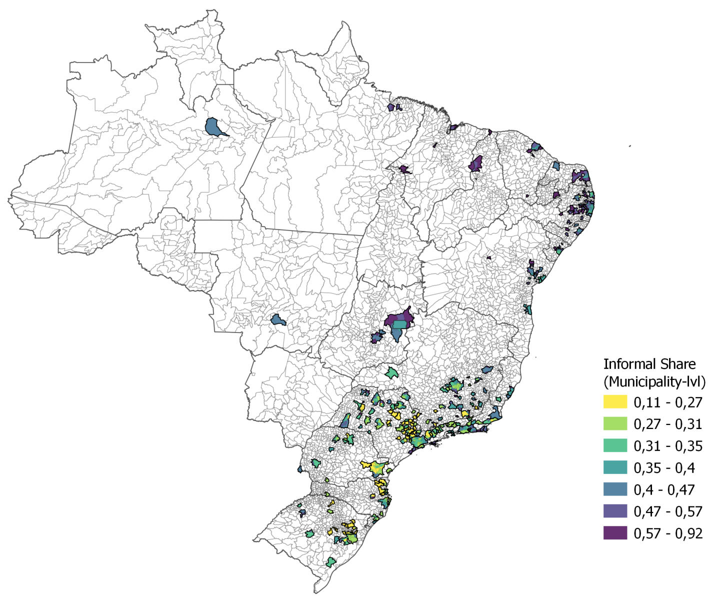
Having defined urban areas and selected a sample of individual wage earners, we can take a closer look at the data. Local informality is unevenly distributed across cities in Brazil and falling with city size (for the most part).

Figure 2 illustrates the distribution of informality at the municipality level across cities in my sample. The informal share ranges from 11 to 92 percent in some places with considerable variation across regions. In particular, cities in the South-East of Brazil tend to be more formal than cities in the Northern parts. However, for those areas there is also a lot of variation within regions and even within some of the larger cities.

Figure 3 further illustrates the relationship between informality share and city size as measured by log employment density. It plots a local mean smoother for informal share (on the left vertical axis) against city size together with a histogram depicting the number of cities for each size bin (right vertical axis). Generally, larger cities tend to be more formal. However, the informal share increases again for very large metropolitan areas such as Sao Paulo, which is in line with a popular belief that the urban informal sector is large in big cities.

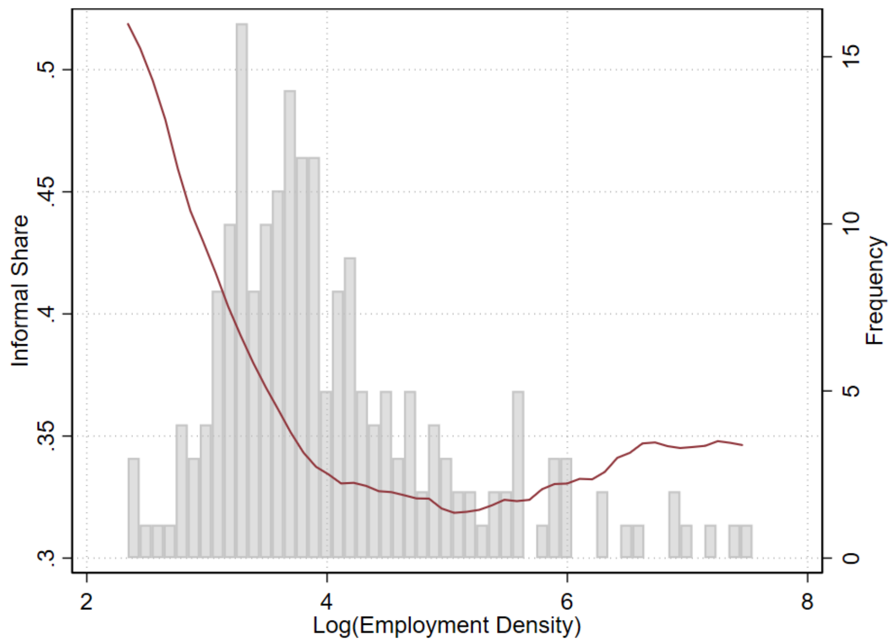
[12] Similarly, formal and informal density are computed as the number of formal and informal workers, respectively, over the area.

Figure 2. Distribution of Informality Across Sample Cities



Note: Own elaboration based on Census data using QGIS, including municipality and state (*unidade federativa*) boundaries.

Figure 3. Informality and City Size in Brazil



Note: Plots a local mean smoother using an Epanechnikov kernel.

5 Empirical Framework: A Two-Step Approach

Following a standard in the literature, the estimation proceeds in two steps (Combes et al., 2008). First, I introduce city fixed effects in an individual wage equation in order to isolate spatial variation in wages from the influence of worker characteristics. Second, I relate the estimated city fixed effects separately to city size, as measured by employment density, formal and informal density, and workforce composition in terms of informality. I repeat these steps for both sub-samples of formal and informal workers, respectively. First, I take the wage of worker i in city c, w_{ic}, to be given by

$$\log w_{ic} = \sum_{c} \gamma_c^k D_{c(i)} + \theta^k X_i + \varepsilon_i \tag{1}$$

where γ_c^k is a city fixed effect, X_i is a vector of individual and job-related characteristics (gender, age, age squared, marital status, indicators for years of education, race, 2-digit occupations, and 2-digit economic sectors) and ε_i is an individual error term.[13] Parameters are indexed by k indicating the sample they were estimated on (pooled, formal or

[13]Unfortunately, the census data does not provide any firm-level information. In particular, I do not observe the registration status of the firm nor its size, which would be interesting to include at this stage.

informal). Equation 1 is estimated by ordinary least squares using individual survey sampling weights provided in the raw data.

Second, I use the city fixed effects estimated in the fist step, that can be interpreted as local wage indices after controlling for observed and unobserved worker characteristics, and regress them on the city characteristics of interest. The advantage of using this two-step procedure, as opposed to directly including city-level variables in the wage equation above, is that I can distinguish local shocks from purely idiosyncratic shocks at the worker level and correctly compute standard errors for the estimated coefficients of aggregate explanatory variables.[14] Since the dependent variable in the second step is estimated in the first, I am weighting the second step estimation by the number of observations used to estimate the respective city fixed effects in order to address potential measurement error. As a consequence, second step estimates can be interpreted with the individual worker as the underlying unit of analysis in mind.

In particular, the main analysis focuses on three different specifications at the second step. I start with estimating the static benefits of working in a bigger city in the standard way, relating city-level wages to log city size measured in terms of total employment density:

$$\gamma_c^k = \alpha_0^k + \beta_T^k \log Den_c^T + \mu_s^k + \xi_c \tag{2}$$

where ξ_c is a city-level error term. I am also including state fixed effects μ_s in order to account for structural differences across states (*unidades federativas*), since many relevant policies and resources are defined at this level. The coefficient of interest, β_T^k, is the static elasticity of wages with respect to density estimated in the literature. It identifies the net effect of larger cities on labor productivity, i.e. benefits minus costs of density that may arise through various channels.

Next, I split total employment density into formal and informal density and estimate the following equation:

$$\gamma_c^k = \alpha_1^k + \beta_F^k \log Den_c^F + \beta_I^k \log Den_c^I + \mu_s^k + \xi_c \tag{3}$$

where β_F^k and β_I^k should approximately add up to β_T^k. This allows me to investigate the contribution of the presence of each 'type' of worker to the overall effect estimated in 2.

Finally, I introduce the share of informal employment and its interaction with total employment density in a separate regression in order to directly test the idea that the extent

[14] See Combes et al. (2008) for an in-depth discussion.

of informality in the local economy moderates the productive advantages of bigger cities:

$$\gamma_c^k = \alpha_2^k + \beta^k \log Den_c^T + \delta^k InfSh_c + \eta^k Den_c^T \times InfSh_c + \mu_s^k + \xi_c \tag{4}$$

All predictors in the second step are standardized in order to aid interpretation, such that $\beta^k + \eta^k$ gives the average marginal effect of total density that workers experience in a city with average informality levels.

Ordinary least squares estimates of this type of specifications may be affected by various sources of endogeneity, both at the individual and the local economy level, as highlighted by Combes and Gobillon (2015). At the worker level, endogeneity occurs when city dummies $D_{c(i)}$ are correlated with the individual error term ε_i, i.e. when workers sort across locations according to their unobserved characteristics such as ability.[15] If we expect workers with higher individual ability to sort into bigger cities, my estimates in the second step are therefore going to overstate the city size wage premium.[16] A second type of endogeneity bias on the individual level arises when workers' location choices depend on an exact job offer with known wages. This is mostly taken care of by the explanatory variables included in the first step.

There are endogeneity concerns at the local economy level, i.e. when one of the regressors in the second step is correlated with the local random component ξ_c, due to both unobserved heterogeneity and reverse causality. The former is an issue because there may be some missing local variable that *directly* affects density, the allocation of workers across formal and informal sectors and wages, for example unobserved productive amenities.[17] Reverse causality is a concern because higher local average wages may attract more workers, increasing city size.

Both of these sources of endogeneity at the aggregate level imply a positive bias in the estimated coefficients and are usually addressed together by instrumentation using local historical or geological variables. However, the literature has found the bias to be of small practical importance (Combes et al., 2010). For the moment, I refrain from developing a more sophisticated identification strategy and instead focus on pure correlations in the

[15] In order to address this, it would be ideal to use panel data that allows to follow workers over time and include individual fixed effects. However, this data is not available at the national level in Brazil.

[16] Although De La Roca and Puga (2017) have shown that this is not only driven by sorting on unobservables, but also the existence of dynamic advantages of bigger cities, which again I cannot account for in the cross-section. Similarly, Baum-Snow and Pavan (2012) indicate that sorting on unobserved ability may be of limited importance.

[17] Even though I am not addressing this concern in my main analysis, I show a version of my results where I try to account for local endowments as much as possible by including a number of city-level controls in section 6.2.

data for my main analysis.[18] My results are therefore not intended to be interpreted in a causal sense.

6 Results and Discussion

I find that location matters for individual wage disparities and that city size is a strong predictor of spatial variation in wages in Brazil. I estimate the elasticity of wages with respect to total employment density to be around 5 percent, for both formal and informal workers. However, that does not imply that they benefit from city size in the same way, because density can mean different things to either group, depending on which type of worker is 'generating' it. In particular, productivity advantages of cities are mainly driven by the presence of formal workers. The elasticity of wages with respect to formal density is much larger for formal compared to informal workers. On the other hand, density in terms of informal workers is associated with lower formal wages and does not matter for informal wages. Hence, density is more productive in places where a smaller share of the workforce is employed informally.

6.1 Main Findings

This section presents my main findings, focusing on the results from the specifications in the second step.

Location matters for individual wage disparities

Table A3 reports the estimation results for the first step (equation 1) for each sub-sample. Earnings are concave in age, a proxy for experience, and increase in the level of education. Most coefficients are similar for formal and informal workers, with a few exceptions. Returns are higher in the formal sector except for technicians, clerks, and service and sales workers. Interestingly, for the sectors of public administration, education, health and social work, workers earn more when they are employed informally.[19] Overall, individual characteristics and city fixed effects explain around 58 percent (R^2) of the variation in wages.

Figure 4 plots the estimated city fixed effects against log city size as measured by total

[18] I had previously explored the possibility to use historical population and a natural experiment involving labor enforcement (also used by Almeida and Carneiro (2012) and Ponczek and Ulyssea (2022)) to instrument density and informality levels, but did not pursue this further because simultaneous instrumentation seemed to introduce other, less straightforward biases.

[19] Possibly related to the institutional setting.

employment density. I find notable geographic differences in wages, even for observationally equivalent workers. For instance, a worker in Salvador earns around 40 percent more than a worker with the same observable characteristics in the smallest place in my sample (Bezerros). City size by itself can explain about 12 percent of the spatial variation in wages that is left after controlling for observable worker characteristics.

Figure 4. Wages and City Size in Brazil

Note: Vertical axis shows city fixed effects from first step estimation, centered around the sample mean. N=202 cities.

City size explains spatial variation in wages

Columns (1) to (3) in Table 1 regress the estimated city fixed effects on log city size and a set of state fixed effects. This yields an elasticity of the wage premium with respect to total employment density of 0.0539 for the pooled sample. Thus, working in a city that is one standard deviation bigger (or more dense) is associated with around 5 percent higher earnings, beyond any variation attributable to differences in individual characteristics.

My pooled OLS estimate of the wage elasticity with respect to employment density is close to estimates in the literature. For example, Chauvin et al. (2017) find an elasticity of wages with respect to urban population of 0.052, and an elasticity with respect to density of 0.026 for Brazil.[20] More broadly, in their meta-analysis of density elasticity estimates, Ahlfeldt and Pietrostefani (2019) find a mean of the density elasticity of wages of 0.04

[20]Note that the sample is restricted to urban prime-age males in microregions with urban population of 100,000 or more.

across studies. Note that my estimate is closer to this than what is usually found for non-high-income countries (0.08).

One concern in the literature is that employment density may not appropriately reflect the density actually faced by individuals (Duranton and Puga, 2020). In order to better capture how close the typical individual is to other people when population is unevenly distributed, De La Roca and Puga (2017) have proposed an "experienced density" measure that counts population within a given radius around each individual and computes an average across individuals in each city. Repeating the analysis using this alternative measure of density yields an elasticity of 0.0605 for the pooled sample (not shown). I continue using the 'naive' employment measure for this analysis because I cannot know how the alternative measure is composed in terms of formality status of workers.

Table 1. OLS Results for Effect of City Size and Informality on Wages By Formality Status

	Basic			Split Density		Interaction	
	(1) Pooled	(2) Formal	(3) Informal	(4) Formal	(5) Informal	(6) Formal	(7) Informal
Log Employment Density	0.0539^{***}	0.0527^{***}	0.0540^{***}			0.0357^{***}	0.0430^{***}
	(0.0040)	(0.0038)	(0.0043)			(0.0048)	(0.0054)
Log Formal Density				0.1173^{***}	0.0468^{*}		
				(0.0271)	(0.0278)		
Log Informal Density				-0.0548^{**}	0.0093		
				(0.0244)	(0.0264)		
Informal Share						-0.0567^{***}	-0.0313^{**}
						(0.0141)	(0.0157)
Log Emp. Density × Inf. Share						-0.0478^{***}	-0.0366^{***}
						(0.0101)	(0.0104)
State FE	✓	✓	✓	✓	✓	✓	✓
Number of Workers	2,408,145	1,586,873	821,272	1,586,873	821,272	1,586,873	821,272
Number of Cities	202	202	202	202	202	202	202
R^2	0.89	0.89	0.88	0.90	0.88	0.91	0.89

Notes: Standard errors in parentheses. * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$. Dependent variable is the respective set of estimated city fixed effects. All predictors are standardized.

(No) Heterogeneity by own formality status

Benefits of working in bigger cities appear to be similar in magnitude for formal and informal workers ($\hat{\beta}_T^F = 0.0527$ and $\hat{\beta}_T^I = 0.0540$). Previous papers that separately estimate the density elasticity of wages for both groups of workers have found a (much) larger elasticity for the *informal* sector (Duranton, 2016; Matano et al., 2020; Bernedo Del Carpio and Patrick, 2021). However, this finding goes against theoretical predictions in the literature, as one would expect formal workers to benefit a lot more than informal ones from city size because they are in a better position to do so (e.g. more educated, better connected). My sense is that some of the difference in estimated elasticities that these other papers find may be driven by including rural areas in the analysis. However, it may also relate to what I am showing next, which is that each group of workers contributes to productive benefits differently, which is something we are missing if we consider the effect of *total* employment density only.

Heterogeneity by others' formality status

The net effect of total employment density 'hides' an important heterogeneity of the effect of density with respect to *others'* formality status. In particular, productivity benefits of larger cities seem to be mainly driven by the presence of formal workers. Columns (4) and (5) of Table 1 report the results for the effects of formal and informal density for each group of workers.[21]

Formal workers benefit very strongly from other formal workers contributing to density ($\hat{\beta}_F^F = 0.1173$). A one standard deviation increase in formal density is associated with around 11 percent higher wages for observationally similar workers in the formal sector. This is more than double of what I found for the elasticity of wages with respect to total density.

On the other hand, formal wages are negatively associated with density in terms of informal workers ($\hat{\beta}_I^F = -0.0548$), offsetting half of the positive effect of formal density. So, comparing cities with the same formal density, a higher concentration of informal workers on average depresses formal sector wages. One possible explanation for this is that formal sector jobs are desirable but may be limited such that competition from informal workers in bigger cities puts downward pressure on formal wages (compensating differentials). However, this cannot explain the other effects I document in this section.

The effects for informal workers are not symmetric and not (very) statistically significant. Similarly to the case of formal workers, informal workers generally seem to benefit from an increase of density in terms of formal workers ($\hat{\beta}_F^I = 0.0468$). This effect is a lot smaller than what I found for formal workers, which is in line with theoretical prediction (as discussed in the previous sub-section).

Informal wages appear to be unaffected by an increase in informal density for a given concentration of formal workers. The coefficient of informal density is slightly positive ($\hat{\beta}_F^I = 0.0093$), so informal workers may also benefit from the presence of their peers at least a little bit, but insignificant.

In sum, it seems that the coefficient on formal density mainly captures the conventional

[21] One concern may be that including city area twice (as the denominator of both density terms) may introduce a multicollinearity bias. However, repeating the analysis with log formal, log informal population and log area separately yields qualitatively similar results.

urban agglomeration economies that the literature seeking to quantify effects of concentration of economic activity on productivity is focusing on. Thus, this may be evidence that density of human capital may be a more relevant measure than employment density (Duranton and Puga, 2020).

Local informality as a moderating factor

The results discussed in the previous subsection suggest important composition effects: a higher concentration of informal workers in the local labor market reduces all workers' ability to benefit from working in bigger cities, both because there are less formal workers whose presence has a (heterogeneous) positive impact and because there are more informal workers that have no impact on an increasing part of the workforce and a negative one the a decreasing group. Column (6) and (7) of Table 1 test this idea directly by interacting city size with the local share of workers that are employed informally.

When introducing the informal share (and its interaction with employment density), estimates of the density elasticity of wages fall. This implies that part of productive advantages estimated before may actually be attributed to a more formal composition of the workforce in larger cities. The coefficient of the informal share itself is negative for both formal and informal workers because positive spillovers from formal workers are reduced and formal workers in addition suffer stronger negative spillovers from informal workers.[22]

The interaction term is negative and significant as well, indicating that we cannot increase employment density without changing informal share as well. The higher the extent of informality in a given city, the lower the benefits from city size will be. On the other hand, relative to the density that the average Brazilian in my sample is experiencing, formalizing the local labor market by one standard deviation (decreasing the informal share by 12 percentage points) is associated with a 10 percent increase in formal wages ($\hat{\delta}^F \Delta InfSh_c + \hat{\eta}^F \Delta InfSh_c = 0.0567 + 0.0478 = 0.1045$).

The city size wage premium at different levels of informality

Figure 5 aims to illustrate how predicted benefits from working in a bigger city vary with the extent of informality in the local labor market, according to my analysis. It plots the average marginal effect of density ($\hat{\beta}^k + \hat{\eta}^k \times InfSh_c$) at varying informality levels, for both the formal and informal worker sample. The idea is that changes in the informality share

[22] Note that I call 'spillovers' the effects discussed in the previous subsection, but they may not actually be externalities.

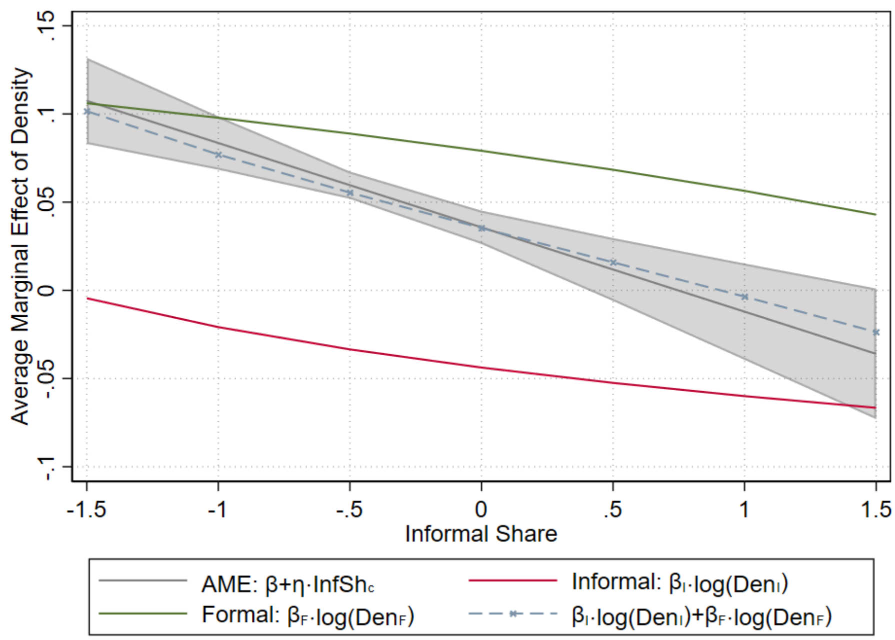
imply different formal and informal densities for a given city, so I can further decompose the average effect including the interaction term using the estimates in Columns (4)-(5) of Table 1.

Focusing on the case of formal workers (coefficients plotted in the upper panel), we can see that higher informality is related to a lower city size wage premium, which results from a combination of a decreasing positive impact of formal density and an increasingly negative impact of informal density. Reducing the informality share by only one standard deviation is associated with a much higher city size wage premium that is close to the one implied by formal density.

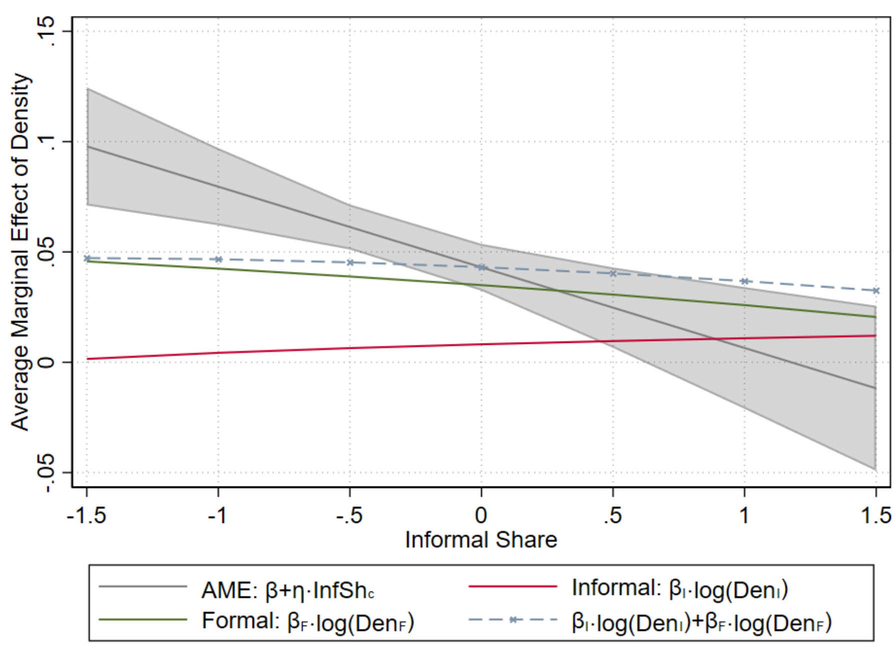
The same graph for informal workers is less conclusive (lower panel). From the negative and significant interaction term from Column (7) of Table 1 we still have that the city size wage premium for informal workers is decreasing with informality share, but this seems to be more negative than implied by the effects for formal and informal density from Column (5). However, the latter two were not statistically significant so this might be related to a lack of statistical power rather than reflecting a deeper reason.

Figure 5. Average Marginal Effects of Density

a. Formal Workers



b. Informal Workers



Note: Each graph plots marginal effects using estimated coefficients from Table 1 at different values of the standardized informal share and their 95% confidence intervals. In order to map these values into formal and informal density, I calibrate total employment density such that the marginal effects of formal and informal density sum up to the average marginal effect at the average share of informality in the sample (which yields a density of 120 and 130 workers per km^2, respectively) and hold it constant at this level when varying informal share.

6.2 Robustness

For the purposes of this first draft, I only provide one robustness check for the analysis presented here (where I include city-level controls). In the future, I plan to replicate my analysis based on newer data for Brazil.

So far, my analysis has looked at pure correlations in the data at the aggregate level, with the exception of controlling for state fixed effects. One concern may be that there are omitted variables related to both labour productivity and the allocation of workers across formal and informal sectors in a city that bias my results. In a first attempt, I try to address this potential endogeneity problem by including geographic controls as well as proxies for market access (distance to state capital) at the level of the city. Table A4 shows the results for the same specifications as above while controlling for these city characteristics. Overall the patterns I document above still hold, but coefficients are smaller in magnitude and loose power, especially for the informal sample. The elasticity of wages with respect to total employment density is 0.0220 for the pooled sample. For formal workers, the relative contribution of the effects of formal and informal density has slightly shifted for the formal sample (the positive effect of formal density is weaker and the negative effect of informal density is stronger) and the coefficient on the interaction of density and informal share is not as negative anymore. For informal workers, none of the effects except for the elasticity of total employment density is significant.

6.3 Discussion of potential mechanisms

In this section, I would like to discuss a few thoughts on what may be behind my results, some of which I am planning to further explore in the future development of this project. There are many possible explanations, however some may not fit the whole picture. For example, one could think that informal workers introduce congestion costs, crowding out benefits for formal workers but this leaves open the question why congestion does not affect other informal workers too. In the end, it is most likely going to be a combination that leads to the findings above. Having said that, I would like to focus on the following three pieces that might each account for part of the story.

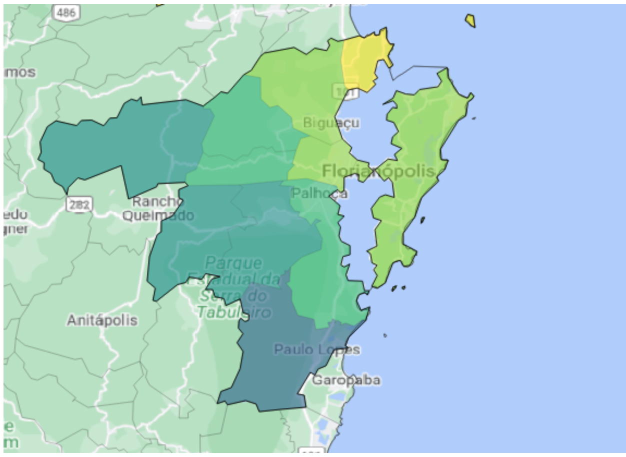
Human capital accumulation and dynamic benefits of bigger cities

De La Roca and Puga (2017) show that there are dynamic benefits of working in bigger cities that workers accumulate over time and that get embedded in workers' human capital. In particular, they show that dynamic benefits may be as important as static ones. Since informal workers are less likely to invest in human capital and are less likely to be trained on-the-job (as discussed in Section 3), and more likely to migrate, it is plausible to assume that they do not accumulate dynamic benefits of working in a big city. Therefore, they are neither capable to learn from their formal counterparts nor do they contribute to positive learning spillovers. My estimates of the effect of density include both static and dynamic benefits (as well as any sorting on unobservables) because I am working with a cross-section. It would be interesting to explore this further with additional data sources that follow workers over time as dynamic benefits seem to be important in this context.

Segregation in terms of worker location in larger cities

One of the major differences between mid to large sized cities in Brazil is where informal workers are located geographically within a city. Although I do not know the exact location of residence for workers in the census, I can exploit the fact that my cities are composed of multiple municipalities to look at informality shares at a more granular level. Figure 6 illustrates the fact that informal workers seem to be pushed towards the outskirts, far away from formal employment opportunities.[23] This would explain why informal workers do not benefit as much as formal workers do from formal density, as this physical separation prevents productive interactions from happening. More granular data would be needed to explore this line of thinking further.

Figure 6. Within-city distribution of informality rates at the municipality level



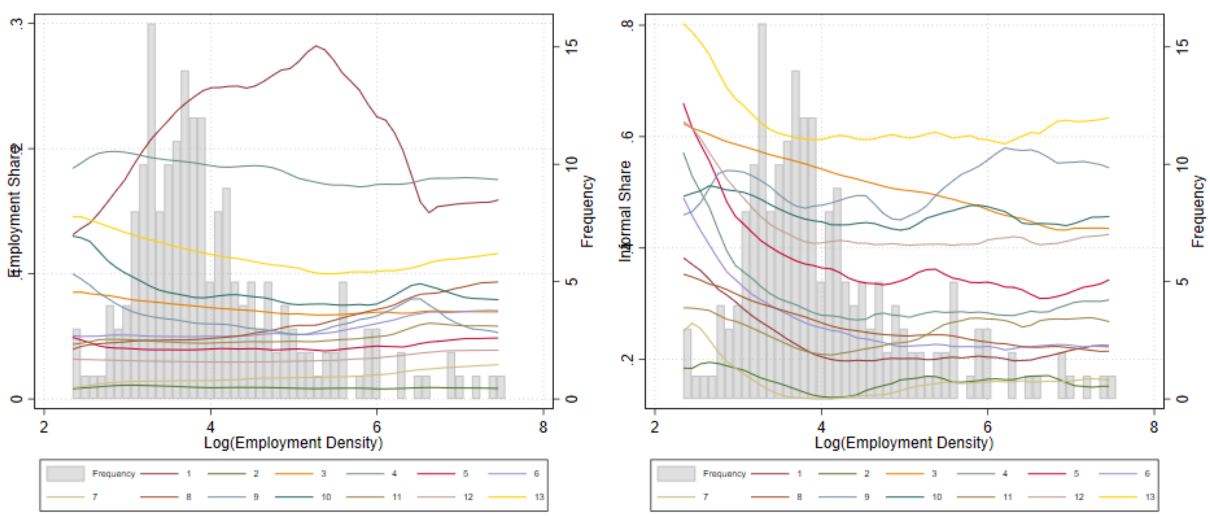
Note: This is a zoomed-in version of Figure 2 above, where darker shades represent a higher informality share. I picked Florianopolis but similar patterns can be found for other multi-municipality cities as well.

[23]Similarly, Zárate (2022) documents for Mexico City that most formal jobs are located in the central areas of the city, while most informal workers reside in the outskirts and that informality rates decline with transit improvements that improve market access of formal employment to informal workers.

Sector specialization at the city level

The extent of informality that I am introducing into the analysis may be capturing something about the city environment, a city's unobserved productivity in particular, that is correlated with the allocation of workers across sectors. The idea is that 'worse' cities (inherently or because of reverse causality) are hosting more informal workers such that wages are low for both groups. Similarly, more productive and bigger cities may attract more productive industries that are more likely to hire formally by the nature of their economic activity. This would explain why formal workers earn more in bigger cities (but not necessarily why informal workers do as well). However, it is also true that there is significant overlap in distributions, such that you can find both formal and informal workers within narrowly defined industries. Still, results might be generated by sectors 'formalizing' differently with city size. Figure 7 illustrates this. For example, a lower share of workers works in manufacturing in very large cities, which is a sector with low levels of informality in most locations. On the other hand, some sectors that tend to be more informal have higher informality shares in larger cities.

Figure 7. Distribution of Sectoral Employment and Informality Shares Across Cities



Note: 1 - Manufacturing, 2 - Electricity, Gas and Water Supply, 3 - Construction, 4 - Wholesale and Retail Trade, 5 - Hotels and Restaurants, 6 - Transport, Storage and Communications, 7 - Financial Intermediation, 8 - Real Estate, Renting and Business Activities, 9 - Public Administration, Defense and Social Security, 10 - Education, 11 - Health and Social Work, 12 - Other Community, Social and Personal Service Activities, 13 - Private Households with Employees

6.4 Limitations

There are several limitations to the results of this analysis. First, internal validity may be limited by different sources of endogeneity, as discussed above. Most importantly, I do not account for sorting of workers on unobserved ability. Unfortunately, there is no panel data at the national level available for Brazil that would allow me to introduce worker fixed effects. I am currently considering to include data from the Brazilian Labor Force Survey (*Pesquisa Mensal de Emprego*) to observe worker transitions across locations and sectors. However, this data is only available for the main six metropolitan regions in Brazil. More generally, note that this analysis is not making any causal statement about the effect of formal and informal density. In addition, there may be better ways to disentangle the effects of size and informality, for example with the help of a structural model. Second, external validity may be limited since informality is necessarily defined by a specific institutional context. What I observe here for the case of Brazil may not hold true in other middle income or lower income countries. Nevertheless, it is the first time a paper is documenting these patterns and regardless of its exact definition, informality is an important enough issue everywhere that we might learn something more general here.

7 Concluding Remarks

In this paper, I show that taking into account informality is important when analyzing the relationship between wages and city size in developing countries. Using individual wage data from Brazil, I find that a higher share of informality is related to a lower city size wage premium. In particular, the net effect of total employment density on wages hides an important heterogeneity in terms of how formal and informal workers contribute and benefit from density. Formal workers benefit from agglomeration of other formal workers but experience a reduction in wages when density results from the concentration of informal workers. Informal workers similarly benefit from formal density but do not seem to be affected by agglomeration of their peers. Hence, productivity advantages of larger cities seem to be mainly driven by the presence of formal workers.

The insights from this analysis have important policy implications for urbanization and formalization strategies in developing countries. Understanding how informality determines the success of cities can help governments leverage the rapid urbanization for economic growth. My results also illustrate indirect consequences of informality, showing that formalizing urban labor markets may lead to wage improvements for formal workers through unlocking agglomeration benefits.

Given the patterns documented in this paper, future research is needed to explore the channels that give rise to these differences. In particular, the empirical evidence presented here may be useful as a basis for a more structural approach incorporating informality into an equilibrium model of cities.

References

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

Almeida, R. and P. Carneiro (2012). Enforcement of labor regulation and informality. *American Economic Journal: Applied Economics 4*(3), 64–89.

Baum-Snow, N. and R. Pavan (2012). Understanding the city size wage gap. *The Review of Economic Studies 79*(1), 88–127.

Bernedo Del Carpio, M. and C. Patrick (2021). Agglomeration and informality: Evidence from peruvian establishments. *Journal of Regional Science 61*(2), 442–471.

Bobba, M., L. Flabbi, and S. Levy (2022). Labor market search, informality, and schooling investments. *International Economic Review 63*(1), 211–259.

Bryan, G., E. Glaeser, and N. Tsivanidis (2019). Cities in the developing world. Technical report, National Bureau of Economic Research.

Chauvin, J. P., E. Glaeser, Y. Ma, and K. Tobio (2017). What is different about urbanization in rich and poor countries? cities in brazil, china, india and the united states. *Journal of Urban Economics 98*, 17–49.

Ciccone, A. and R. E. Hall (1996). Productivity and the density of economic activity. *The American Economic Review 86*(1), 54–70.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*(2), 723–742.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics*, pp. 15–66. University of Chicago Press.

Combes, P.-P. and L. Gobillon (2015). The empirics of agglomeration economies. In *Handbook of Regional and Urban economics*, Volume 5, pp. 247–348. Elsevier.

De La Roca, J. and D. Puga (2017). Learning by working in big cities. *The Review of Economic Studies 84*(1), 106–142.

Dijkstra, L., A. J. Florczyk, S. Freire, T. Kemper, M. Melchiorri, M. Pesaresi, and M. Schiavina (2021). Applying the degree of urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *Journal of Urban Economics 125*, 103312.

Dingel, J. I., A. Miscio, and D. R. Davis (2021). Cities, lights, and skills in developing economies. *Journal of Urban Economics 125*, 103174.

Duranton, G. (2015). Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns. In *The Economics of Interfirm Networks*, pp. 107–133. Springer.

Duranton, G. (2016). Agglomeration effects in colombia. *Journal of Regional Science 56*(2), 210–238.

Duranton, G. (2021). Classifying locations and delineating space: An introduction. *Journal of Urban Economics 125*, 103353.

Duranton, G. and D. Puga (2004). Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics*, Volume 4, pp. 2063–2117. Elsevier.

Duranton, G. and D. Puga (2020). The economics of urban density. *Journal of Economic Perspectives 34*(3), 3–26.

Glaeser, E. and J. V. Henderson (2017). Urban economics for the developing world: An introduction. *Journal of Urban Economics 98*, 1–5.

Glaeser, E. L. and D. C. Mare (2001). Cities and skills. *Journal of Labor Economics 19*(2), 316–342.

Matano, A., M. Obaco, and V. Royuela (2020). What drives the spatial wage premium in formal and informal labor markets? the case of ecuador. *Journal of Regional Science 60*(4), 823–847.

Meghir, C., R. Narita, and J.-M. Robin (2015). Wages and informality in developing countries. *American Economic Review 105*(4), 1509–46.

Ponczek, V. and G. Ulyssea (2022). Enforcement of labour regulation and the labour market effects of trade: Evidence from brazil. *The Economic Journal 132*(641), 361–390.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from brazil. *American Economic Review 108*(8), 2015–47.

Zárate, R. D. (2022). Spatial misallocation, informality, and transit improvements: Evidence from mexico city.

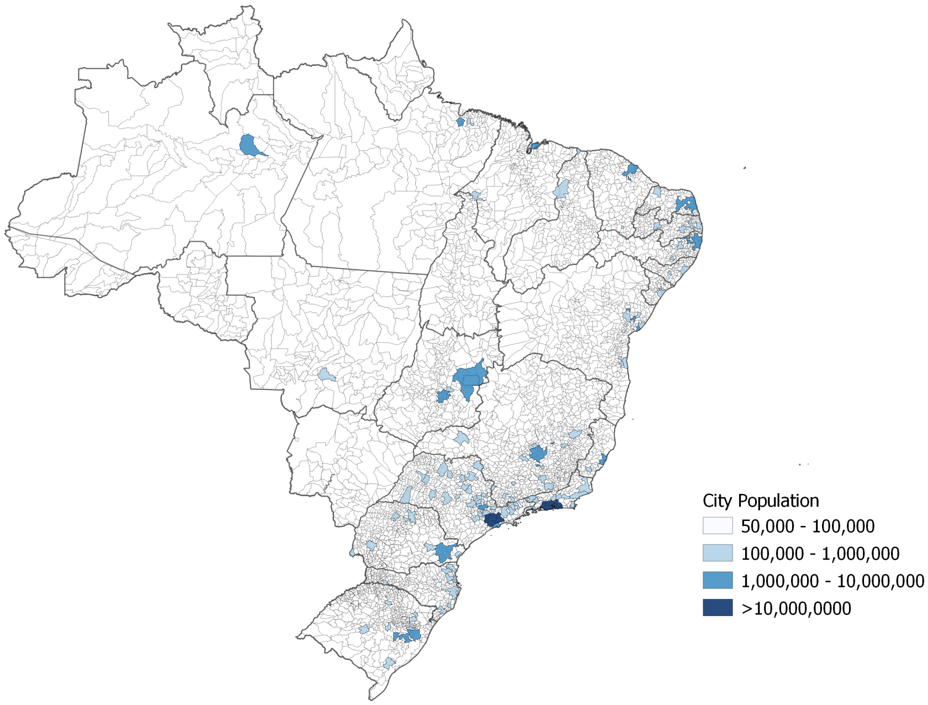
Appendix: Tables and Figures

Appendix Table A1. List of Top 20 Sample Cities in terms of Population

Name of Core Municipality	Number of Constituting Municipalities	Core Population	City Population	City Area (km^2)
Sao Paulo	35	10,434,252	17,803,784	6,539
Rio De Janeiro	18	5,857,904	10,869,255	5,364
Belo Horizonte	30	2,238,526	4,315,011	6,846
Recife	31	1,422,905	3,739,268	5,065
Fortaleza	10	2,141,402	2,823,479	3,867
Porto Alegre	16	1,360,590	2,816,904	6,576
Brasilia	11	2,051,146	2,747,993	24,805
Salvador	8	2,443,107	2,725,077	1,444
Curitiba	19	1,587,315	2,674,102	8,334
Belem	5	1,280,614	1,795,536	1,820
Goiania	15	1,093,007	1,653,574	4,088
Manaus	1	1,405,835	1,405,835	11,408
Campinas	4	969,396	1,355,982	1,252
Vila Velha	6	345,965	1,350,196	1,706
Santos	5	417,983	1,288,237	857
Natal	32	712,317	1,277,900	7,691
Sao Luis	3	870,028	1,053,600	1,382
Maceio	9	797,759	971,452	1,720
Joao Pessoa	17	597,934	959,698	2,540
Teresina	7	715,360	934,885	7,039

Note: The core is defined as the municipality with the largest population for cities that are composed of more than one municipality.

Appendix Figure A1. Distribution of Population Across Sample Cities



Note: Own elaboration based on Census data using QGIS, including municipality and state (*unidade federativa*) boundaries.

Appendix Table A2. Summary Statistics for Individual Workers

	Formal	Informal
Age	33.48	32.85
Male	0.59	0.48
Married	0.44	0.36
Highest Level of Education		
At most literacy	0.04	0.07
Elementary school	0.44	0.50
High school	0.36	0.27
University	0.14	0.14
Post-graduate degree	0.02	0.02
Ethnicity		
Branca (White)	0.61	0.54
Preta (Black)	0.07	0.08
Amarela (Asian)	0.01	0.01
Parda (Mixed White/Black)	0.31	0.37
Indigenous	0.00	0.00
Occupation		
Managers	0.04	0.02
Professionals	0.07	0.09
Technicians	0.11	0.12
Clerks	0.18	0.11
Service and Sales Workers	0.32	0.42
Industrial Production and Services Workers	0.25	0.22
Repair and Maintenance Workers	0.03	0.03
Sector		
Manufacturing	0.23	0.11
Electricity, Gas and Water Supply	0.01	0.00
Construction	0.05	0.09
Wholesale and Retail Trade; Repair of motor vehicles and household goods	0.19	0.15
Hotels and Restaurants	0.05	0.04
Transport, Storage and Communications	0.08	0.04
Financial Intermediation	0.03	0.01
Real Estate, Renting and Business Activities	0.10	0.05
Public Administration, Defense and Social Security	0.04	0.10
Education	0.06	0.11
Health and Social Work	0.06	0.04
Other Community, Social and Personal Service Activities	0.03	0.04
Private Households with Employees	0.06	0.20
Observations	1,586,873	821,272

Note: Highest level of education is defined based on years of schooling. Occupation categories defined according to *Classificação Brasileira de Ocupações Domiciliar*, using ISCO-88 major groups. Economic sectors are defined according to *Classificação Nacional de Atividades Econômicas Domiciliar*, using ISIC Rev. 3, 2-digit level. Sampling weights were used.

Appendix Table A3. OLS results for individual wage equation

	(1) Full Sample	(2) Formal Workers	(3) Informal Workers
Male	0.227***	0.226***	0.207***
	(0.008)	(0.009)	(0.005)
Age	0.070***	0.063***	0.067***
	(0.002)	(0.002)	(0.002)
Age squared	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)
Married	0.130***	0.121***	0.127***
	(0.004)	(0.004)	(0.005)
Elementary school	0.218***	0.210***	0.218***
	(0.007)	(0.006)	(0.008)
High school	0.511***	0.495***	0.479***
	(0.010)	(0.007)	(0.012)
University	1.019***	1.002***	0.993***
	(0.013)	(0.013)	(0.019)
Post-graduate degree	1.069***	1.072***	1.014***
	(0.027)	(0.035)	(0.027)
Preta (Black)	−0.140***	−0.149***	−0.123***
	(0.006)	(0.007)	(0.006)
Amarela (Asian)	0.104***	0.100***	0.121***
	(0.014)	(0.018)	(0.013)
Parda (Mixed White/Black)	−0.117***	−0.121***	−0.105***
	(0.004)	(0.005)	(0.003)
Indigenous	−0.115***	−0.108***	−0.114***
	(0.009)	(0.011)	(0.012)
Managers	0.717***	0.708***	0.643***
	(0.016)	(0.014)	(0.017)
Professionals	0.698***	0.720***	0.652***
	(0.010)	(0.015)	(0.012)
Technicians	0.393***	0.370***	0.418***
	(0.005)	(0.004)	(0.008)
Clerks	0.164***	0.132***	0.163***
	(0.007)	(0.007)	(0.009)
Service and Sales Workers	−0.044***	−0.066***	−0.046***
	(0.004)	(0.004)	(0.006)
Repair and Maintenance Workers	0.136***	0.203***	0.012*
	(0.007)	(0.007)	(0.007)
Electricity, Gas and Water Supply	0.266***	0.276***	0.208***
	(0.023)	(0.024)	(0.022)
Construction	−0.094***	−0.079***	0.025***
	(0.013)	(0.014)	(0.009)
Wholesale and Retail Trade; Repair	−0.093***	−0.080***	−0.010
	(0.008)	(0.009)	(0.009)
Hotels and Restaurants	−0.160***	−0.145***	−0.067***
	(0.012)	(0.012)	(0.014)
Transport, Storage and Communications	0.048***	0.050**	0.108***
	(0.018)	(0.021)	(0.009)
Financial Intermediation	0.320***	0.346***	0.226***
	(0.041)	(0.053)	(0.019)
Real Estate, Renting and Business Activities	−0.077***	−0.072***	−0.004
	(0.009)	(0.009)	(0.013)
Public Administration, Defense and Social Security	0.185***	0.087***	0.442***
	(0.032)	(0.026)	(0.030)
Education	−0.097***	−0.092***	0.061***
	(0.020)	(0.016)	(0.021)
Health and Social Work	−0.015	−0.045***	0.183***
	(0.012)	(0.012)	(0.018)
Other Community, Social and Personal Service Activities	−0.113***	−0.082***	−0.006
	(0.015)	(0.016)	(0.013)
Private Households with Employees	−0.264***	−0.214***	−0.143***
	(0.019)	(0.019)	(0.019)
City FE	✓	✓	✓
Observations	2,408,145	1,586,873	821,272
R^2	0.58	0.58	0.57

Notes: Standard errors in brackets, clustered at city level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Reference groups: At most literacy for education, White for ethnicity, industrial production and services workers for occupation, and manufacturing for economic sector. All regressions use individual sampling weights. Computed using Stata's `reghdfe` routine.

Appendix Table A4. OLS Results for Effect of City Size and Informality with Controls

	Basic			Split Density		Interaction	
	(1) Pooled	(2) Formal	(3) Informal	(4) Formal	(5) Informal	(6) Formal	(7) Informal
Log Employment Density	0.0220^{***}	0.0218^{***}	0.0216^{***}			0.0157^{***}	0.0185^{***}
	(0.0055)	(0.0052)	(0.0061)			(0.0055)	(0.0065)
Log Formal Density				0.1025^{***}	0.0334		
				(0.0243)	(0.0248)		
Log Informal Density				-0.0720^{***}	-0.0110		
				(0.0219)	(0.0234)		
Informal Share						-0.0533^{***}	-0.0191
						(0.0133)	(0.0147)
Log Emp. Density × Inf. Share						-0.0213^{**}	-0.0113
						(0.0100)	(0.0108)
Controls	✓	✓	✓	✓	✓	✓	✓
State FE	✓	✓	✓	✓	✓	✓	✓
Number of Cities	198	198	198	198	198	198	198
R^2	0.93	0.93	0.92	0.93	0.92	0.93	0.92

Notes: Standard errors in parentheses. * $p < 0.10$,** $p < 0.05$,*** $p < 0.01$. Dependent variable is the respective set of estimated city fixed effects. Density and informal share are standardized. City-level controls are: longitude, latitude, altitude, altitude squared, distance to state capital and distance squared from Almeida and Carneiro (2012).