Exploring Spatial Wage Curves: Evidence from Colombia's Formal and Informal Employment*

Miguel Santiago Pulido-Estupiñan†

February 7, 2024

Abstract

This paper estimates spatial wage curves for formal and informal workers in Colombia, using data from the Colombian Labor Force Survey *(GEIH)* for the period 2016-2019. Unlike previous Colombian wage curve studies, this analysis uses four different spatial matrices that focus on factors such as contiguity, distance, and intraregional economic activity. In addition, the study calculates informal-formal wage curve differentials by gender and education level, while also taking into account characteristics that would affect selection into formal and informal employment. Methodologically, the study employs a two-stage least squares (2SLS) approach using the lagged unemployment rate as an instrumental variable. The results show that informal workers, especially males and those with less education, are more sensitive compared to their formal, female and more educated counterparts. The study also shows that spatial spillovers have a positive effect on the wages of formal workers, while having a negative impact on informal workers.

Keywords: Spatial wage curves, formal workers, informal workers, Colombia, regional labor markets, wage elasticity.
JEL Classification: J31, J46, R12

*This research represents the culmination of my Master's in Economics journey. First and foremost, I would like to express my deepest gratitude to my advisor, Andrés Garcia-Suaza, for his guidance and support. I am also indebted to Professors Umberto Muratori, Paul Rodriguez, Jesus Otero, and Fernando Jaramillo for enriching this research with their valuable comments. The insights gained from the discussions during the microeconomic workshop, macroeconomic workshop, and brown bag seminar at the Universidad del Rosario were instrumental in shaping this work. I would like to thank my friends and colleagues Laura Clavijo, Santiago Fernandez, Silvia Ortiz, and Juan Andres Russy for their camaraderie and encouragement. Finally, I would like to express my deepest gratitude to my family, especially my mother, Liliana, whose constant support has been a source of strength and inspiration. All remaining errors are my own.

†Master's candidate in Economics, Department of Economics, Universidad del Rosario. Email: miguels.pulido@urosario.edu.co

1 Introduction

Imagine a scenario in which two people, equally qualified and with the same characteristics, are employed in different cities in Colombia. Why would one earn more than the other? Several factors could contribute to this disparity, but many economists contend that the answer lies in the intricate relationship between wages and unemployment rates-a subject that has captivated scholars for decades (Parra Castro & Váquiro Cuellar, 2017). Specifically, there has been considerable debate over whether high unemployment rates lead to a decrease or increase in the average wages received by workers (Guadagni, 2007).

In this context, wage flexibility emerges as a key and controversial concept. It is argued that greater wage flexibility - whether full or partial - would facilitate a more effective labor market response to adverse demand shocks, thereby reducing job losses relative to scenarios characterized by labor rigidity (Arango et al., 2010). This debate is situated within a theoretical framework in which neoclassical theory emphasizes the adaptability of wages to fluctuations in labor supply and demand. This contrasts with alternative theories, such as model of staggered contracts (Taylor, 1980), which assumes the existence of fixed wage contracts that are unaffected by external economic conditions.

The wage curve, extensively explored by Blanchflower and Oswald (1994, 1995, 2005), serves as an empirical tool for assessing wage flexibility and reveals an inverse relationship between wages and unemployment rates. Their findings suggest that wages are typically lower in regions with higher unemployment and higher in regions with lower unemployment. Alternatively, the efficiency wage theory extends these observations and provides a theoretical lens through which to understand the impact of the unemployment rate on wages [1]. According to this theory, in areas with high unemployment, firms are able to offer lower wages without risking high turnover because the threat of unemployment itself acts as a deterrent to poor performance. Conversely, in regions with low unemployment, firms must incentivize productivity and discourage complacency by offering higher wages.

Earlier literature focused primarily on the interplay between wages within a given region and the corresponding unemployment rate, treating regions as autonomous entities, recent research has broadened its scope. They now include the influence of unemployment rates in neighboring regions on wage dynamics (Elhorst et al., 2007; Karatas, 2017; Uchoa, 2019). These studies provide a deeper understanding of whether increases in unemployment rates in one region affect wages in neighboring regions. Such analyses provide economic insights into the intricate interdependencies that characterize wage fluctuations in nearby labor markets.

Moreover, labor markets in developing countries such as Colombia are characterized by a substantial prevalence of informal employment. Informal labor aggregates constitute a significant proportion, exceeding 50%, of the total employment landscape across Latin America (Jutting, Laiglesia, et al., 2009). In Colombia, informal employment has seen a notable increase and now accounts for approximately 58% of the total labor force (Departamento Administrativo Nacional de Estadísticas (DANE), 2023).

As Ramos et al. (2010) points out, informal employment can result from both involuntary exclusion from

[1] The full derivation of the efficiency wage theory can be found in the Appendix.

formal sector opportunities and voluntary opt-outs from structured formality to evade taxes and social contributions. Typically, informal employment involves individuals being trapped in unproductive and poorly protected jobs, making them significantly more vulnerable to local labor market fluctuations than their formal sector counterparts.

This paper aims to answer two main questions: First, it examines the relationship between wages and unemployment for both formal and informal workers in different urban centers in Colombia. Second, it examines the influence of spatial correlations on these wage and unemployment dynamics. To achieve this, the methodology used in this study employs a spatial wage curve approach that incorporates lagged regional unemployment rates as an instrument [2], using data from the Colombian Labor Force Survey from 2016 to 2019. In addition, I develop four different matrices based on factors such as contiguity, distance, and intraregional economic activities.

This study represents a significant contribution to the literature. In Colombia, no previous study has taken into account the spatial dimension while simultaneously taking into account the formal and informal labor force and disaggregating it by gender and level of education.

The results of this study show that informal workers experience a steeper spatial wage curve than formal workers, with values ranging from $[-0.139, -0.136]$ for informals and $[-0.0985, -0.0956]$ for formals. Moreover, the slope is more pronounced for male informal workers, with values ranging from $[-0.168, -0.165]$, than for female informal workers, with values ranging from $[-0.107, -0.105]$. Informal workers with lower levels of education, whose elasticity values range from $[-0.151, -0.147]$, are more vulnerable to wage fluctuations than their more educated counterparts; in fact, the wage curve is not statistically significant for informal workers with high levels of education. It is important to note that the magnitude and direction of spatial spillovers vary depending on the specific group considered and the matrix used for analysis.

The rest of this paper is organized as follows: Section 2 reviews the relevant literature; Section 3 provides a comprehensive overview of the data used in this study; Section 4 details the methodology used to analyze wage curve dynamics; Section 5 presents the main results, including model results; and finally, Section 6 draws conclusions and discusses the implications of this research for the Colombian labor market.

2 Previous Literature

As explained by Blanchflower and Oswald (1995), the wage curve paradigm establishes the relationship between individual wage levels and regional unemployment rates. This empirical principle, rooted in labor economics, is expressed as:

$$\ln(w) = \beta \ln(u) + \Phi$$

[2]Most of the works regarding the wage curve use the lagged unemployment rate as an instrument. See Arango et al. (2010), Baltagi and Başkaya (2022), Baltagi and Rokicki (2014), Baltagi et al. (2017), Berg and Contreras (2004), Blanchflower and Oswald (1995), de Paula and Marques (2022), Karatas (2017), Lugo (2006), and Sánchez and Núnez (1998).

where $\ln(w)$ is the natural logarithm of real wages, $\ln(u)$ is the logarithm of the regional unemployment rate, and Φ represents other terms that may affect real wages. Blanchflower and Oswald (2005) have stimulated extensive research in over 40 countries. This study, covering both mature Western economies and developing transition economies, consistently find a statistically significant inverse relationship between individual wage levels and local unemployment rates. The coefficient β, an elasticity parameter, is typically close to -0.1.

Globally, Buettner (1999) pioneered maximum likelihood methods for estimating spatial panel models. However, this groundbreaking work omitted the inclusion of regional unemployment rate logarithms in the wage curve paradigm. Similarly, Pannenberg and Schwarze (2000) faced challenges in disentangling spatially determined unemployment effects due to multicollinearity. Longhi et al. (2006) contributed by incorporating spatially weighted unemployment rates into wage curve estimation. Baltagi et al. (2012) found the insignificance of spatially induced unemployment in the dynamic wage curve, attributing it to workers' myopia with respect to neighboring regions. Baltagi and Rokicki (2014) highlighted a gender asymmetry, with men's wage elasticity to unemployment exceeding that of women. Ramos et al. (2015) introduced spatially lagged unemployment rates and wage dynamics to elucidate regional wage determination.

Given the study's focus on formal and informal workers, it is important to consider the findings of previous research in emerging economies, particularly in Latin America. Some studies have found that public sector workers, women, youth, and less educated workers, especially in the informal sector, are more vulnerable than their counterparts (Berg & Contreras, 2004; Bucheli & González, 2007; Lugo, 2006). Other studies have applied spatial components to account for spillovers, finding gender differences and strong evidence of negative spatial spillovers (Baltagi et al., 2017; de Paula & Marques, 2022).

In the Colombian context, previous research has laid the groundwork for understanding the wage curve, with studies such as Sánchez and Núnez (1998), Parra Castro and Váquiro Cuellar (2017), Ramos et al. (2010), Saavedra-Arango (2016), and Arango et al. (2010) estimating negative correlations between wage levels and unemployment rates, often in line with efficiency wage theory. For a more complete breakdown of the Colombian studies see Table A1 in the appendix.

This study, however, is distinguished by three main contributions: first, it integrates the spatial dimension, taking into account both the formal and the informal segments of the labor market; second, it significantly expands the geographic scope by including data from 31 cities, going beyond the typical 13-city focus in previous research; third, it applies the 2SLS method to the Mincer equation with a spatial component and uses the lagged regional unemployment rate as an instrument.

3 Data

The dataset used in this study comes from the Colombian Labor Force Survey (*Gran Encuesta Integrada de Hogares-GEIH*), conducted by the *Departamento Administrativo Nacional de Estadísticas (DANE)*, which covers the period from 2016 to 2019. This survey collects a wide range of demographic and employment-related information, including gender, age, marital status, educational level, working conditions, and sources of income. It provides data at the national, regional, and departmental levels, including the capital cities of each department. Notably, Colombia has 32 departments, each with its own capital city; however, one of these departments is an island and is excluded from this study.

The dataset is carefully filtered to the noninstitutional working age population, which includes individuals 15 years of age and older. Missing values within the dataset were systematically removed, resulting in a comprehensive dataset derived from 31 selected cities. The total number of observations exceeds 1 million.

It is important to note that the data structure is not a panel, but a repeated cross-section. This distinction arises because the composition of individuals surveyed in the GEIH changes over the years.

In order to distinguish between formal and informal employment strata within the GEIH dataset, a new variable was created through a systematic process. Initially, attention was focused on two key questions [3]. The first question sought to determine whether individuals were actively contributing to a pension fund and served as the basis for classification. This classification methodology was then refined using the second question, which sought to identify the entity responsible for covering an individual's monthly pension fund contributions. These responses allowed individuals to be classified as formal or informal workers.

The calculation of real hourly earnings followed a sequential approach. First, nominal monthly earnings were quantified through specific survey questions [4]. Next, information on the number of hours worked per week was obtained through another question [5]. The combination of these responses allowed the determination of the nominal hourly wage. Finally, the real hourly wage was determined by applying the Consumer Price Index (CPI), with the base year set at December 2018. Additionally, the detailed nature of the GEIH dataset facilitated the direct calculation of regional unemployment rates.

[3] Questions are: are you currently contributing to a pension fund? and who pays monthly for the pension fund membership?

[4] Questions are: first, before discounts, how much did you earn last month at this job?, second, what was the net profit or net fees from this activity, business, profession, or farm last month?, and how many months does what you received correspond to?

[5] How many hours a week do you normally work in that job?

Table 1: Descriptive Statistics GEIH 2016-2019

	Formal		**Informal**	
	Mean	**SD**	**Mean**	**SD**
Panel A. Males				
Real wage (log real COP)	8.68	0.68	8.11	0.73
Years of education	5.21	0.96	4.37	1.23
Age	38.16	11.73	41.04	15.51
Observations	234496		298437	
Panel B. Females				
Real wage (log real COP)	8.72	0.68	7.85	0.86
Years of education	5.54	0.79	4.51	1.24
Age	37.30	11.07	41.54	14.42
Observations	201614		265998	

Notes: Years of education are set to 1 for illiterate, 2 for preschool, 3 for basic primary, 4 for basic secondary, 5 for secondary (ending with a high school diploma), and 6 for at least a bachelor's degree and higher. Real hourly earnings are reported in Colombian Pesos (COP), and observations represent the number of data points for each category (formal and informal).

Table 1 presents descriptive statistics of the labor market in Colombia specially for hourly wages, age, and educational level for males and females disaggregated by their work status. It is important to note that the average real wage of women in the formal sector is higher than that of their male counterparts. This difference might be counterintuitive in the broader context of gender wage differentials, where men typically earn more than women. However, this finding is consistent with other reports by the DANE [6].

4 Methodology

4.1 Model

The basic model for estimating the standard wage curve, primarily anchored within the framework enunciated by Blanchflower and Oswald (1995), takes the following form:

$$LogW_{irt} = \alpha + \beta LogU_{rt} + \gamma X'_{irt} + \mu_r + \varphi_t + \vartheta_{irt} \tag{1}$$

[6] See *Brecha salarial de Género en Colombia 2022*: https://www.dane.gov.co/files/investigaciones/notas-estadisticas/dic-brecha-salarail-genero-2022-v3.pdf

In this equation, $LogW_{irt}$ represents the natural logarithm of real hourly wages for an individual i in region r at year t, and $LogU_{rt}$ measures the natural logarithm of the unemployment rate within region r at time t. The matrix X'_{irt} includes control variables, such as age, age squared, gender, marital status, educational attainment, industry classification, employment location, firm size, employment position, and other relevant dimensions [7]. The region fixed effect is denoted as μ_r, while φ_t is the time fixed effect [8]. The error term is denoted as ϑ_{irt}.

However, as emphasized by Baltagi and Başkaya (2022), the omission of spillover effects emanating from neighboring labor markets could potentially introduce bias and inconsistency into estimates that depend on the wage curve framework. Consequently, the standard wage equation needs to be augmented by the inclusion of a spatially mediated term that accounts for the cumulative influence of the weighted unemployment rates prevailing in neighboring regions, as follows:

$$LogW_{irt} = \alpha + \beta LogU_{rt} + \theta \sum_{j \neq r}^{J} \omega_{rj} LogU_{jt} + \gamma X'_{irt} + \mu_r + \varphi_t + \vartheta_{irt} \tag{2}$$

Note that the only extension of equation (1) is the inclusion of the term $\theta \sum_{j \neq r}^{J} \omega_{jt} LogU_{jt}$, which captures a composite measure of unemployment rates in other regions weighted by spatial linkages.

The coefficient β reflects the elasticity that characterizes real hourly wages within region r in response to fluctuations in its own unemployment rate. In contrast, the elasticity of real hourly wages within the same region with respect to unemployment rates in neighboring regions depends on the magnitude of θ and the configuration of the weight matrix. Negative values for β indicate that an increase in the unemployment rate in region r could potentially lower the earnings of individuals in that region.

The central focus of this paper is to investigate the differences in spatial wage curves between formal and informal workers within the Colombian jurisdiction. To this end, I estimate the wage curve regressions given in equation (2) for formal and informal workers separately. This is achieved in the following equation:

$$LogW^S_{irt} = \alpha^S + \beta^S LogU_{rt} + \theta^S \sum_{j \neq r}^{J} \omega_{rj} LogU_{jt} + \gamma^S X'_{irt} + \mu^S_r + \varphi^S_t + \vartheta^S_{irt} \tag{3}$$

Here, the superscript S denotes the individual's particular employment status, an attribute that can take the values *formal* or *informal*. I expect that $|\beta^{informal}| > |\beta^{formal}|$, implying that variations in local unemployment rates have a more pronounced effect on informal workers' wages than on their formal counterparts.

Addressing potential endogeneity in the explanatory variables is a critical aspect of ensuring the robustness of the estimation framework. As emphasized by Sánchez and Nuñez (1994), the unemployment rate could not only affect the wage level, but also be affected by it, potentially leading to endogeneity. This situation undermines the principle of strict exogeneity, which requires that the independent variables are not correlated with the error term. To counteract this, I use instrumental variables, specifically the lagged regional unemployment rate as an

[7] Other dimensions such as house characteristics, access to the Internet, and total number of persons in the household.

[8] Time fixed effect is a combination of annual fixed effects and seasonal/monthly fixed effects.

instrument to correct for the presumed endogeneity of the unemployment rate with respect to the wage level.

To rigorously test the validity of the instrument, the lagged regional unemployment rate, I will present the *Kleibergen-Paap Wald F statistic*, which assesses the risk of weak instrumentation. In addition, I will use the *Kleibergen-Paap LM statistic* to test for underidentification of the equation to ensure that the instrument is appropriately identified and relevant.

4.2 Spatial Weight Matrices: Capturing the Spatial Dimensions of the Colombian Wage Curve

In order to comprehensively analyze the spatial dimensions inherent in the Colombian wage curve, four different spatial weight matrices have been formulated. These matrices use the concepts of geographic proximity, road proximity, contiguity, and economic activity as basic building blocks.

The central role of these matrices lies in the spatial wage curve model, where the goal is to understand how wages in a given region are affected not only by local unemployment rates, but also by unemployment rates in neighboring regions. Incorporating these matrices into the model allows the cumulative influence of unemployment rates in neighboring areas to be captured, providing a more accurate representation of wage dynamics in Colombia.

In the following subsections, I will delve into the specifics of each spatial weight matrix, explaining how they are constructed and their unique characteristics.

4.2.1 Distance Matrix

This 31x31 matrix, called Ω_1, places a strong emphasis on the inverse distances between the centroids of the 31 cities under study. Formally, each entry of Ω_1 is defined as:

$$\omega_{rs} = \begin{cases} (d_{rs})^{-1} & \text{if} \quad r \neq s \\ 0 & \text{if} \quad r = s \end{cases}$$

Where $(d_{rs})^{-1}$ represents the inverse of the distance calculated using latitude and longitude coordinates between the centroids of regions r and s. This practice is widely used in spatial analysis, emphasizing the central role of proximity in shaping spatial interactions. Figure 1 panel a will allow for a better understanding of the matrix Ω_1.

4.2.2 Road Distance Matrix

Referred to as Ω_2, this 31x31 matrix incorporates road-based distances using data from Google Maps. However, due to the unique geographic characteristics of the Colombian landscape, an assumption is made. Specifically, for the cities of Leticia, Mitú, and Inírida, it is well known that travel by car is not feasible. Consequently, for these particular cases, the geographic distances from the previous matrix are retained in the formulation of Ω_2 to maintain consistency. Formally, each entry of Ω_2 is defined as follows:

$$\omega_{rs} = \begin{cases} (rd_{rs})^{-1} & \text{if} \quad r \neq s \\ 0 & \text{if} \quad r = s \end{cases}$$

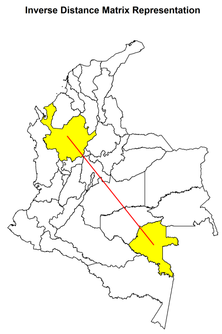
With $(rd_{rs})^{-1}$ representing the inverse of the road-based distance between regions r and s, while emphasizing the practical considerations for these specific cities. Figure 1 panel b will allow for a better understanding of the matrix Ω_2.

4.2.3 Contiguity Matrix

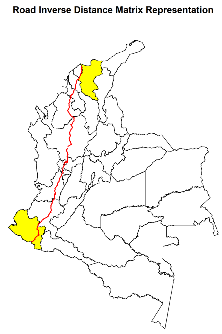
This 31x31 matrix, labeled Ω_3, introduces temporal constancy by capturing spillover influences between regions that share territorial boundaries. However, a critical assumption guides this construction. Given the observable geographic separation and the lack of shared borders among the 31 cities, an alternative approach is necessary. This approach aligns the departmental boundaries with the perimeters of each city, allowing for the juxtaposition of departmental boundaries. Formally, each entry of Ω_3 takes the following form:

$$\omega_{rs} = \begin{cases} c_{rs} & \text{if} \quad r \neq s \\ 0 & \text{if} \quad r = s \end{cases}$$

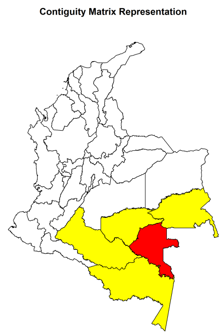
Where c_{rs} takes the value of 0 for non-contiguous regions and a value of 1 for the regions r and s that share contiguous boundaries. This approach allows interpolation of spatial interactions between partitions only if they share contiguous boundaries. Figure 1 panel c will allow for a better understanding of the matrix Ω_3.



(a) Inverse Distance Map



(b) Road Inverse Distance Map



(c) Contiguity Map

Figure 1: Spatial analysis of Colombia. (a) Shows the inverse distance relationships between Vaupés and Antioquia. (b) Details the critical transportation links, especially within the *Troncal de Occidente*. Finally, (c) presents the contiguity map, delineating the immediate neighboring regions that share borders with Vaupés.

4.2.4 Economic Activity Matrix

I will now present the economic activity matrix, a crucial component of this analysis, based on the work of Haddad et al. (2016). This matrix, labelled Ω_4, is constructed as an interregional input-output matrix for Colombia, focusing on the year 2012. The authors provide a detailed analysis of the intraregional and interregional shares of the average total output multipliers, providing valuable insights into regional economic relationships.

The economic activity matrix serves as a support for this study, allowing the complex economic interactions between Colombian regions to be taken into account. By incorporating this matrix into this analytical framework, I consider the broader regional economic context alongside local factors, providing a more comprehensive understanding of wage dynamics.

Figure 2: Heatmap of Interregional Economic Activity

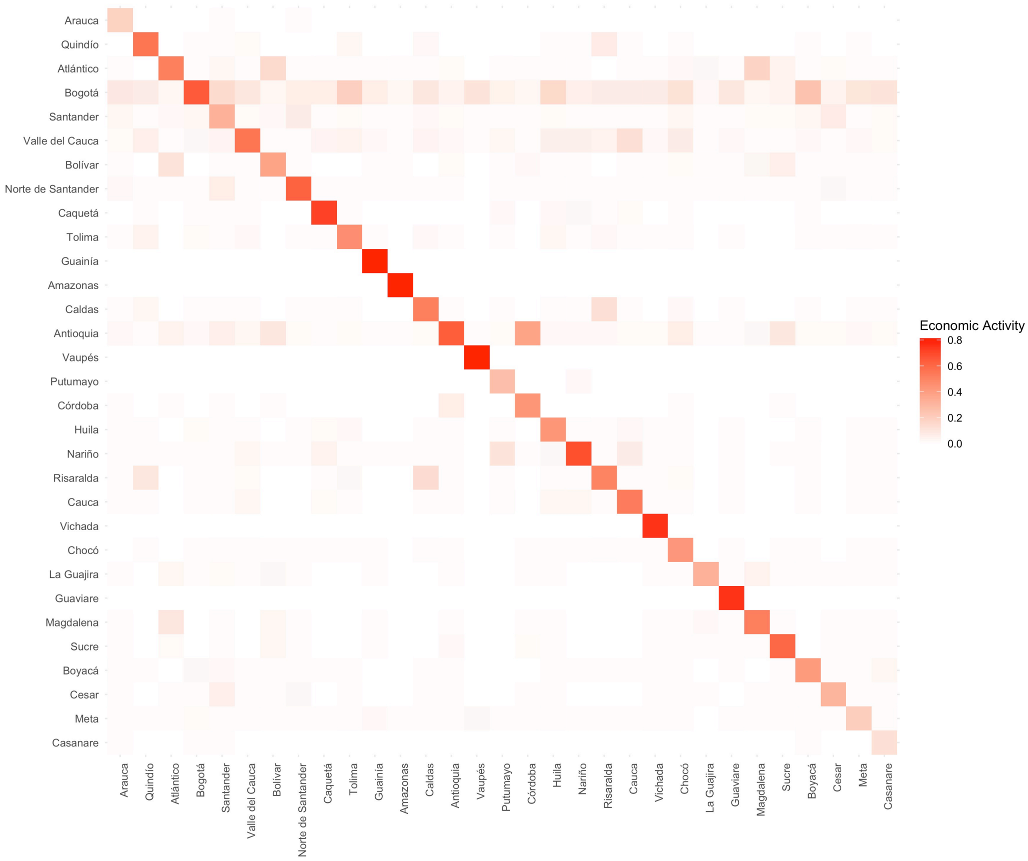


Figure 2 visualizes levels of economic activity within Colombian regions. Each square on the grid corresponds to economic activity between two regions, with darker shades indicating higher levels of activity. The most striking feature is the dark diagonal line across the matrix, indicating that the highest levels of economic activity occur within each region itself. In particular, Bogotá stands out with a notably dark square on this diagonal, indicating a high level of intraregional economic activity. This suggests that Bogotá, as the capital and a major urban center, has a vibrant economy with significant internal economic interactions, reflecting its central role in the national economy.

It is important to note that all four matrices discussed above have undergone a process of row normalization, which ensures that each row in the matrices adds up to one, thus facilitating comparative analysis. However, while this row normalization improves the analytical framework, it is imperative to recognise that they are inherently unable to account for social proximities between cities, which can exist independently of territorial contiguity, geographical distance and economic relationship.

5 Estimation Results

5.1 Standard Wage Curve

Table 2 presents the empirical results of the standard wage curve from 2016 to 2019, given in equation (1). The difference between columns (1) and (2) is the methodology used, one using ordinary least squares (OLS) and the other using two-stage least squares (2SLS) to address potential endogeneity, both including regional and time fixed effects. Column (1) shows a negative elasticity of -0.0321 between regional wages and the unemployment rate, meaning that a 1% increase in the unemployment rate would reduce average hourly wages by about 1.88 Colombian Pesos (COP). This elasticity differs in magnitude from previous values found for different time periods and fewer cities in the Colombian context. For example, values found using OLS in Saavedra-Arango (2016) and Sánchez and Núnez (1998) are closer to -0.1.

In contrast, when instrumenting with the lagged unemployment rate to address endogeneity in column (2), the elasticity increases in absolute value to -0.1046 and remains statistically significant. This implies a stronger negative relationship between unemployment and wages. In monetary terms, this elasticity translates into an approximate decrease of 6.12 COP in the average hourly wage for a 1% increase in the unemployment rate. It is important to note that this approximation does not take into account spatial spillovers and the values found in this approximation are similar in magnitude to those found by Arango et al. (2010) and Ramos et al. (2010) for Colombia and those found by Blanchflower and Oswald (2005) for other emerging economies.

Table 2: Standard Wage Curve 2016-2019

	OLS	**2SLS**
	(1)	(2)
$LogU_{rt}$	-0.0321***	-0.1046***
	(0.00459)	(0.0170)
R^2	0.456	0.456
Kleibergen-Paap Wald F Stat		38000
P-value Kleibergen-Paap LM Stat		0.000
Time Fixed Effects	YES	YES
Region Fixed Effects	YES	YES
Observations	1,017,219	996,484

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. In FE-2SLS specification, the logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

5.2 Spatial Wage Curve

Table 3 presents the results for the spatial wage curve from 2016 to 2019 for Colombia, where each column represents a different spatial weight matrix using the 2SLS estimation method. Column (1) shows the estimation results based on the contiguity matrix. The results show that: first, the elasticity of hourly wages with respect to the region's own unemployment rate is -0.104; second, the response of hourly wages to the unemployment rate of neighboring regions is -0.000172.

Columns (2) and (3) present the results based on inverse distance and inverse road distance, respectively. The results show an elasticity of -0.105 and -0.106 with respect to the local unemployment rate. When the effect of the neighboring unemployment rate is considered, the estimates are 0.000449 and 0.003690, respectively.

Finally, the estimates using the economic weight matrix are presented in column (4). The elasticity of real wages with respect to the local unemployment rate is -0.105, while the estimate with respect to the weighted neighborhood unemployment rate is 0.000007.

Analyzing the impact of unemployment on wages using different spatial matrices, we observe different effects. Using the contiguity matrix, a 1% increase in the region's own unemployment rate leads to a decrease in hourly wages of about 6.08 COP, and an additional decrease of about 0.01 COP due to neighboring regions' unemployment rates. The inverse distance matrix shows a similar decrease of about 6.14 COP in hourly wages for the region's unemployment, but an increase of 0.03 COP due to neighboring unemployment. The inverse road distance matrix results in a larger decrease of about 6.20 COP in hourly wages for the region's unemployment, with an increase of 0.22 COP from neighboring unemployment rates. Finally, using the economic weight matrix, the impact of the region's unemployment is again a decrease of about 6.14 COP in hourly wages, with a marginal increase of 0.0004 COP from the neighboring regions' unemployment rate.

Note that the coefficient on $LogU_{rt}$ is negative and statistically significant in all specifications. However, the coefficients on the spatial spillover term, $\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$, vary from positive to negative depending on the weighted matrix, but none are statistically significant.

Table 3: Spatial Wage Curve 2016-2019

	2SLS Contiguity	**2SLS Distance**	**2SLS Road Distance**	**2SLS Economic**
	(1)	(2)	(3)	(4)
$LogU_{rt}$	-0.104***	-0.105***	-0.106***	-0.105***
	(0.0173)	(0.0172)	(0.0171)	(0.0172)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.000172	0.000449	0.003690	0.000007
	(0.000689)	(0.000765)	(0.003460)	(0.000836)
R^2	0.456	0.456	0.456	0.456
Kleibergen-Paap Wald F Stat	36000	37000	37000	37000
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES
Observations	996,484	996,484	996,484	996,484

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

5.3 Spatial Wage Curve along Formal-Informal workers

Table 4 gives the estimation results for the spatial wage curve taking into account the formality-informality division for all four matrices. In doing so, different patterns emerge. First, it is worth noting that, regardless of the specification used, the hourly wages of informal workers are more sensitive to the different variations in the local unemployment rate than their counterparts. These results are consistent and in line with those of Baltagi and Başkaya (2022) and Baltagi and Rokicki (2014) for European countries, Baltagi et al. (2017) and de Paula and Marques (2022) for Brazil, and Ramos et al. (2010) for Colombia, who find a steeper wage curve for informal workers.

Using the intraregional economic activity matrix, column (7) shows that the elasticity of wages with respect to one's region's unemployment rate is -0.137 for informal workers, and column (8) shows the elasticity for formal workers with a value of -0.0969. This means that for a 1% increase in the unemployment rate, hourly wages decrease by about 5.71 COP and 7.72 COP, respectively. Using the inverse road distance matrix, the elasticities are -0.138 and -0.0985 for informal and formal workers, respectively. This translates into a decrease in hourly wages of about 5.76 COP for informal workers and 7.85 COP for formal workers.

Table 4: Spatial Wage Curve along Formal-Informal Workers 2016-2019

	2SLS Contiguity		**2SLS Distance**		**2SLS Road Distance**		**2SLS Economic**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Informal	**Formal**	**Informal**	**Formal**	**Informal**	**Formal**	**Informal**	**Formal**
$LogU_{rt}$	-0.136***	-0.0971***	-0.139***	-0.0956***	-0.138***	-0.0985***	-0.137***	-0.0969***
	(0.0250)	(0.0256)	(0.0247)	(0.0254)	(0.0245)	(0.0253)	(0.0248)	(0.0255)
$\sum_{j \neq r}^{J} \omega_{rj} LogU_{jt}$	-0.00181*	0.00231***	-0.000748	0.00241**	-0.00598	0.0170***	-0.00177	0.00295***
	(0.00103)	(0.000858)	(0.00113)	(0.000961)	(0.00511)	(0.00432)	(0.00124)	(0.00104)
R^2	0.298	0.489	0.298	0.489	0.298	0.489	0.298	0.489
Kleibergen-Paap Wald F Stat	23000	13000	24000	14000	24000	13000	24000	13000
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Observations	553,244	427,599	553,244	427,599	553,244	427,599	553,244	427,599

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

In terms of spatial spillovers, I find different effects of the unemployment rate in other regions on the wages of formal and informal workers. For formal workers, the estimates are positive and statistically significant for all four different matrices. For informal workers, however, the estimates are negative, but significant only when the contiguity weight matrix is used. Columns (3) and (4) of Table 4 show the estimates when the inverse distance matrix is used, the magnitude of the spillovers is -0.000748 for informal workers and 0.00241 for formal workers. When the contiguity matrix is used, the values for the spatial spillovers are shown in columns (1) and (2) with values of −0.00181 and 0.00231 for informal and formal workers, respectively.

In real terms, for formal workers, the results indicate an increase in hourly wages of about 0.19 COP (using the inverse distance matrix) and 0.18 COP (using the contiguity matrix) for a 1% increase in the unemployment rate in neighboring regions. For informal workers, however, the results lead to a decrease in hourly wages of about 0.03 COP (using the inverse distance matrix) and 0.08 COP (using the contiguity matrix).

These findings provide valuable insights into the functioning of Colombia's labor markets. First, they underscore the important role of informal employment in labor market adjustments during periods of macroeconomic shocks that lead to increased unemployment rates. Interpreting the estimated wage elasticities as a measure of labor market flexibility, the results suggest that the wages of informal workers in Colombia are more responsive to fluctuations in unemployment rates than their formal counterparts, a pattern previously documented by Baltagi et al. (2013) and Baltagi and Başkaya (2022). In addition, I show that this labor market flexibility in Colombia has a spatial component. More precisely, the wages of informal workers in a given region exhibit a negative response to changes in the unemployment rate in neighboring regions, while formal workers exhibit a positive response.

As a possible mechanism, in the case of informal labor markets, which are characterized by lower barriers to entry and greater wage flexibility, an increase in unemployment in nearby regions often leads to increased labor supply spilling over into these markets. As a result, informal employers, faced with a larger pool of available workers, may take advantage of this situation to lower wages, driven by increased competition among workers seeking employment. Conversely, the formal labor market, characterized by higher barriers to entry, stronger regulation, and more robust labor protections, responds differently under similar circumstances. When unemployment rises in neighboring regions, the formal sector in unaffected areas does not necessarily experience a corresponding increase in labor supply. In fact, it may experience an increase in demand for formal employment due to the relative stability and security of the sector. This is particularly pronounced when job losses in neighboring areas are concentrated in their formal sectors, thereby increasing the attractiveness of formal employment in unaffected areas. As a result, employers in these unaffected formal markets may increase wages to attract and retain skilled workers.

5.4 Informal and Formal Spatial Wage Curves By Gender

It is essential to consider gender as a critical variable in the study of wage differentials across geographic regions. This approach is based on the understanding that economic behavior and outcomes may differ significantly between men and women due to a variety of social, cultural, and economic factors. In Colombia, this aspect is particularly relevant as previous studies that do not include the spatial component have presented contrasting results on the gender sensitivity of wage curves. On the one hand, there are some studies that found that females are more sensitive than their male counterpart (Parra Castro & Váquiro Cuellar, 2017; Sánchez & Núnez, 1998) but other studies found that males are more vulnerable than females (Arango et al., 2010; Ramos et al., 2010).

These different views underscore the importance of including a spatial component in the analysis to better understand how geographic factors interact with gender in influencing wages. To this end, this study examines the spatial wage curve elasticities for men and women, with a particular focus on formal and informal workers, as defined in Table 5 and Table 6, respectively.

For the contiguity matrix, not taking into account the formal-informal status, the elasticities with respect to the local region unemployment are presented in column (1) of Table 5 and Table 6. The values of these elasticities are -0.152 for men and -0.0539 for women, and both are statistically significant. However, when considering spatial spillovers, symbolized by $\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$, both are negative but not statistically significant.

Taking into account the worker status and using the contiguity matrix, the estimates for formal and informal males are given in columns (2) and (3) of Table 5. For formal male workers, the elasticity with respect to the local unemployment rate is -0.156, implying a decrease in hourly wages of about 12.41 COP, and 0.00279 with respect to neighboring regions, implying an increase in hourly wages of about 0.22 COP. For informal male workers, the elasticity is -0.165 for their own unemployment rate, resulting in a decrease of about 7.47 COP in hourly wages, and -0.00226 for spatial spillover, resulting in an additional decrease of about 0.10 COP in hourly wages, all statistically significant.

In contrast, for formal and informal female workers, the estimates in columns (2) and (3) of Table 6 are lower in absolute value compared to their male counterparts. The values are -0.105 and -0.0291 for the elasticity with respect to local unemployment. This translates into a decrease in hourly wages of about 3.96 COP for informal workers and 2.33 COP for formal workers. With respect to neighboring regions, the elasticity values are -0.00108 for informal and 0.00159 for formal workers, leading to an additional decrease of about 0.04 COP in hourly wages for informal workers and an increase of about 0.13 COP for formal workers. Notably, only the elasticity for informal workers with respect to their own unemployment rate is statistically significant.

The same pattern is shown for the weight matrix of intraregional economic activities. For example, when ignoring the worker status of females, column (10) of Table 6, the values of the estimates are -0.0540 and -0.000149, while for males, column (10) of Table 5, the values are -0.152 and 0.000260 for the elasticity with respect to local unemployment rate and for other regions unemployment rate, respectively. Furthermore, when Table 5 and Table 6 are compared, the results indicate that regardless of the spatial weight matrix used and the worker status, men have much higher sensitivity to local unemployment rates than female workers consistent with previous findings (Baltagi et al., 2012; Baltagi & Rokicki, 2014; Baltagi et al., 2017; de Paula & Marques, 2022; Karatas, 2017).

Finally, using the weight matrix of inverse distance and inverse road distance, and controlling for formal status as shown in columns (6) and (9) of Table 5 for men and Table 6 for women, the effects on the local unemployment rate are negative but statistically significant only for men; when spatial spillovers are taken into account, the elasticities with respect to neighboring regions are positive but not statistically significant for formal female workers.

Table 5: Informal and Formal Spatial Wage Curve For Men

	2SLS Contiguity			2SLS Distance			2SLS Road Distance			2SLS Economic		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.152***	-0.165***	-0.156***	-0.153***	-0.168***	-0.155***	-0.154***	-0.167***	-0.159***	-0.152***	-0.167***	-0.156***
	(0.0220)	(0.0312)	(0.0333)	(0.0218)	(0.0308)	(0.0331)	(0.0216)	(0.0305)	(0.0330)	(0.0218)	(0.0310)	(0.0332)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.000097	-0.00226*	0.00279**	0.000723	-0.00114	0.00331***	0.00516	-0.00873	0.0220***	0.000260	-0.00219	0.00380***
	(0.000880)	(0.00130)	(0.00113)	(0.000977)	(0.00143)	(0.00127)	(0.00443)	(0.00645)	(0.00576)	(0.00107)	(0.00156)	(0.00137)
R^2	0.438	0.294	0.496	0.438	0.294	0.496	0.438	0.294	0.496	0.438	0.294	0.496
Kleibergen-Paap Wald F Stat	20000	13000	7790.95	21000	13000	7926.84	21000	13000	7814.47	21000	13000	7873.21
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	530,714	292,346	229,880	530,714	292,346	229,880	530,714	292,346	229,880	530,714	292,346	229,880

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 6: Informal and Formal Spatial Wage Curve For Females

	2SLS Contiguity			2SLS Distance			2SLS Road Distance			2SLS Economic		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.0539**	-0.105***	-0.0291	-0.0545**	-0.107***	-0.0272	-0.0549**	-0.106***	-0.0294	-0.0540**	-0.106***	-0.0285
	(0.0271)	(0.0394)	(0.0394)	(0.0268)	(0.0389)	(0.0390)	(0.0267)	(0.0386)	(0.0389)	(0.0269)	(0.0391)	(0.0392)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.000161	-0.00108	0.00159	0.000173	-0.000112	0.00114	0.00189	-0.00336	0.0102	-0.000149	-0.00105	0.00176
	(0.00106)	(0.00159)	(0.00130)	(0.00118)	(0.00176)	(0.00146)	(0.00532)	(0.00793)	(0.00648)	(0.00129)	(0.00192)	(0.00158)
R^2	0.487	0.293	0.490	0.487	0.293	0.490	0.487	0.293	0.490	0.487	0.293	0.490
Kleibergen-Paap Wald F Stat	16000	11000	5509.72	16000	11000	5646.7	16000	11000	5556.14	16000	11000	5584.13
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	465,770	260,898	197,719	465,770	260,898	197,719	465,770	260,898	197,719	465,770	260,898	197,719

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

5.5 Informal and Formal Spatial Wage Curves by Education Level

Table 7 and Table 8 present the results of the wage curve for formal-informal workers by education level, taking into account the spatial component.

Estimates are shown in Table 7 for less educated workers. There is a significant negative relationship between local unemployment rates and wages in all four matrices. This relationship is particularly pronounced for informal workers, with estimates ranging from -0.147 to -0.151, underscoring the vulnerability of this group's wages to local unemployment fluctuations, while for the formal sector within this education group, estimates are lower in absolute terms, ranging from -0.130 to -0.131, all of which are also statistically significant. Taking into account the economic activity matrix, the elasticity values for less educated informal and formal workers are -0.148 and -0.131, respectively, as shown in columns (11) and (12). This means that for every 1% increase in the unemployment rate, the hourly wages of less educated informal workers decrease by about 5.04 COP and the hourly wages of less educated formal workers decrease by about 6.14 COP.

Turning to the more educated workers in Table 8, the direct impact of regional unemployment on wages appears to be significant in all matrices for all workers and formal workers, but not for informal workers. However, the values are smaller in magnitude compared to their counterparts. Looking at the inverse distance matrix, the elasticity value for highly educated informal workers is -0.0694, implying a decrease in hourly wages of about 4.78 COP, while for formal workers it is -0.0725, implying a decrease in hourly wages of about 7.61 COP, as shown in columns (5) and (6).

The spatial spillovers for the less educated workers show interesting patterns. The estimates are negative but not statistically significant for the full sample, but when informal status is taken into account, the estimates remain negative but are significant only when using the spatial matrix with contiguity and economic weight, as shown in columns (2) and (11) of Table 7, respectively. For formal and less educated workers, the estimates are positive and significant for all four matrices. For the contiguity matrix, the spatial spillover values for less educated informal and formal workers are -0.00227 and 0.00338, respectively, as shown in columns (2) and (3). In real terms, this means that for every 1% increase in the unemployment rate in neighboring regions, there is an additional decrease of about 0.08 COP in hourly wages for less-educated informal workers and an increase of about 0.16 COP in hourly wages for less-educated formal workers.

In contrast, for more educated workers, the effects on neighborhood unemployment are not statistically significant for the full sample, informal workers, or formal workers. However, the estimates are significant only when the inverse road distance matrix is used for the full sample and for formal workers, as shown in columns (7) and (9) of Table 8. The elasticity with respect to the unemployment rate in neighboring regions using the road distance matrix is 0.0161, as shown in column (9). This means that for every 1% increase in the unemployment rate in neighboring regions, the hourly wages of highly educated formal workers increase by about 1.69 COP.

Table 7: Informal and Formal Spatial Wage Curve Along Lower Educated Workers

	2SLS Contiguity			2SLS Distance			2SLS Road Distance			2SLS Economic		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.125***	-0.147***	-0.131***	-0.127***	-0.151***	-0.130***	-0.127***	-0.149***	-0.131***	-0.126***	-0.148***	-0.131***
	(0.0213)	(0.0277)	(0.0353)	(0.0211)	(0.0273)	(0.0350)	(0.0209)	(0.0271)	(0.0350)	(0.0212)	(0.0274)	(0.0351)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.000733	-0.00227**	0.00338***	0.000320	-0.000921	0.00383***	-0.000215	-0.00804	0.0184***	-0.000610	-0.00226*	0.00390***
	(0.000850)	(0.00114)	(0.00107)	(0.000943)	(0.00126)	(0.00120)	(0.00428)	(0.00569)	(0.00531)	(0.00103)	(0.00137)	(0.00129)
R^2	0.244	0.176	0.233	0.244	0.176	0.233	0.244	0.176	0.233	0.244	0.176	0.233
Kleibergen-Paap Wald F Stat	24000	21000	4990.45	25000	21000	5082.47	25000	21000	4988.16	25000	21000	5044.26
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	626,278	431,251	185,885	626,278	431,251	185,885	626,278	431,251	185,885	626,278	431,251	185,885

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table 8: Informal and Formal Spatial Wage Curve Along Upper Educated Workers

	2SLS Contiguity			2SLS Distance			2SLS Road Distance			2SLS Economic		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.0594**	-0.0687	-0.0746**	-0.0585**	-0.0694	-0.0725**	-0.0613**	-0.0692	-0.0769**	-0.0592**	-0.0688	-0.0747**
	(0.0294)	(0.0594)	(0.0350)	(0.0291)	(0.0584)	(0.0346)	(0.0289)	(0.0583)	(0.0345)	(0.0292)	(0.0589)	(0.0348)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	0.000762	-0.000617	0.00175	0.000518	-0.000462	0.00128	0.00985*	-0.00229	0.0161**	0.000968	-0.000826	0.00243
	(0.00115)	(0.00244)	(0.00124)	(0.00128)	(0.00268)	(0.00138)	(0.00580)	(0.0121)	(0.00626)	(0.00140)	(0.00293)	(0.00150)
R^2	0.440	0.346	0.433	0.440	0.346	0.433	0.440	0.346	0.433	0.440	0.346	0.433
Kleibergen-Paap Wald F Stat	12000	3673.92	8497.14	12000	3808.37	8689.81	12000	3722.64	8577.66	12000	3755.98	8608.36
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	370,206	121,993	241,714	370,206	121,993	241,714	370,206	121,993	241,714	370,206	121,993	241,714

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

6 Robustness Check

In this section, I extend the analysis by including the Inverse Mills Ratio *(IMR)*[9] in equation (3). The IMR is derived from the first stage selection model based on the Heckman correction method and is used to correct for

[9]For relevant studies that have utilized the IMR in similar contexts, see Arango et al., 2010; Baltagi and Başkaya, 2022; Baltagi et al., 2017; Gonzales-Martinez, 2018.

potential sample selection bias in the econometric model. The Heckman correction is particularly important because it can lead to biased and inconsistent estimates if left unaddressed.

The need for this correction arises from the potential for non-random selection into labor market segments. For example, individuals may self-select into employment based on unobserved characteristics that also affect wage levels, creating a correlation between these unobserved factors and the error term in the wage equation. The revised model incorporating the IMR is specified as follows:

$$LogW_{irt}^{S} = \alpha^{S} + \beta^{S} LogU_{rt} + \theta^{S} \sum_{j \neq r}^{J} \omega_{rj} LogU_{jt} + \gamma^{S} X'_{irt} + \delta^{S} IMR_{irt} + \mu_{r}^{S} + \varphi_{t}^{S} + \vartheta_{irt}^{S} \tag{4}$$

Table A2 - Table A6, included in the appendix, show the spatial wage curve relationships for both informal and formal workers, with adjustments made for selection into employment status. This detailed analysis spans all four spatial weight matrices.

It is important to highlight that the Inverse Mills Ratios, denoted as *IMR*, emerge as statistically significant in the second stage regressions in all the models we tested. The consistent significance of the *IMR* coefficients underscores that process selection into employment status, as well as the distinction between informality and formality in the labor market, is not a random event.

It is also worth noting that despite the inclusion of the *IMR* to account for selection bias, the estimates observed in this section do not deviate significantly from those presented in the previous section of this study.

7 Conclusions

The empirical exploration undertaken in this study has revealed intricate relationships within spatial wage dynamics, underscoring the multifaceted interplay between wages and unemployment rates across multiple dimensions. These dimensions span different spatial matrices, employment categories, gender distinctions, and educational classifications, providing a comprehensive picture of regional wage trends.

First, this analysis confirms a consistent inverse relationship between wages and local unemployment rates. Interestingly, the elasticity of this relationship appears to strengthen when we account for potential endogeneity, presenting a nuanced view of the conventional wage curve that differs from previous periods studied in the Colombian context.

Second, the findings on spatial spillovers are consistent with the negative relationship between wages and neighborhood unemployment rates commonly observed in international studies (Baltagi & Başkaya, 2022; Ramos et al., 2015). However, despite the predominant focus on negative spatial spillovers in the existing literature, this analysis, along with a few other studies, highlights scenarios where positive spillovers occur (Baltagi et al., 2012; Baltagi & Rokicki, 2014; de Paula & Marques, 2022). This divergence underscores the need for further research to understand the conditions that foster such positive effects.

Third, when the estimation is further stratified by employment type, a striking revelation emerges: informal workers exhibit a pronounced sensitivity to unemployment fluctuations compared to their formal counterparts. Their wages are not only more responsive to shifts in regional unemployment, but also reflect a negative elasticity with respect to neighboring unemployment rate.

Fourth, adding a gender lens to the narrative, male informal workers stand out as particularly vulnerable to unemployment fluctuations. The spatial spillover patterns unravel a tapestry of interrelated influences that affect male and female workers differently. Underlying these trends are potential socioeconomic and geographic determinants that could influence wage dynamics.

Fifth, another layer of complexity unfolds when focusing on educational attainment. Less educated workers, especially those in the informal sector, show increased vulnerability to local unemployment shocks. In contrast, their highly educated counterparts, especially those in the informal sector, appear impervious to these fluctuations. More intriguingly, the latter group appears to benefit from spillover effects, especially when unemployment escalates in adjacent areas.

Overall, this study underscores the profound complexity of spatial wage dynamics. The elasticity of the wage-unemployment relationship varies across methodologies, while certain segments, such as male informal workers, emerge as particularly vulnerable. Moreover, the instrumental role of road connectivity and economic intraregional activities in shaping regional wage dynamics suggests broader policy implications, not only for improving regional connectivity, but also for promoting economic resilience.

Ultimately, these findings have important policy implications. Recognizing and addressing the vulnerabilities of specific groups can pave the way for more targeted labor market strategies. At the same time, the central role of infrastructure, especially roads, points to the far-reaching benefits of strategic investments that can promote regional economic stability and growth.

References

Arango, L. E., Obando, N., & Posada, C. E. (2010). *Sensibilidad de los salarios al desempleo regional en Colombia: Nuevas estimaciones de la curva de salarios* (tech. rep.). Banco de la República. Bogotá, Colombia. https://doi.org/10.32468/be.590

Baltagi, B. H., Baskaya, Y. S., & Hulagu, T. (2013). How different are the wage curves for formal and informal workers? Evidence from Turkey. *Papers in Regional Science*, *92*(2), 271–283.

Baltagi, B. H., & Başkaya, Y. S. (2022). Spatial wage curves for formal and informal workers in Turkey. *Journal of Spatial Econometrics*, *3*(1), 3.

Baltagi, B. H., Blien, U., & Wolf, K. (2012). A dynamic spatial panel data approach to the German wage curve. *Economic Modelling*, *29*(1), 12–21.

Baltagi, B. H., & Rokicki, B. (2014). The spatial Polish wage curve with gender effects: Evidence from the Polish Labor Survey. *Regional Science and Urban Economics*, *49*, 36–47.

Baltagi, B. H., Rokicki, B., & De Souza, K. B. (2017). The Brazilian wage curve: New evidence from the National Household Survey. *Empirical Economics*, *53*(1), 267–286. https://doi.org/10.1007/s00181-016-1105-5

Berg, J., & Contreras, D. (2004). Political-Economic regime and the wage curve: Evidence from Chile, 1957–96. *International Review of Applied Economics*, *18*(2), 151–165.

Blanchflower, D. G., & Oswald, A. J. (1994). *The wage curve*. MIT press.

Blanchflower, D. G., & Oswald, A. J. (1995). An introduction to the wage curve. *Journal of economic perspectives*, *9*(3), 153–167.

Blanchflower, D. G., & Oswald, A. J. (2005). The wage curve reloaded.

Bucheli, M., & González, C. (2007). An estimation of the wage curve for Uruguay. *Documento de Trabajo/FCS-DE; 11/07*.

Buettner, T. (1999). The effect of unemployment, aggregate wages, and spatial contiguity on local wages: An investigation with German district level data. *Papers in Regional Science*, *78*(1), 47–67.

Departamento Administrativo Nacional de Estadísticas (DANE). (2023). *Boletín Gran Encuesta Integrada de Hogares (GEIH) abril-junio 2023* (tech. rep.). https://www.dane.gov.co/files/operaciones/GEIH/bol-GEIHEISS-abr-jun2023.pdf

de Paula, G. C., & Marques, A. M. (2022). Brazilian wage curve: Further evidence based on spatial interactions in times of recession, 2012–2019. *The Indian Journal of Labour Economics*, *65*(3), 689–708.

Elhorst, J. P., Blien, U., & Wolf, K. (2007). New evidence on the wage curve: A spatial panel approach. *International Regional Science Review*, *30*(2), 173–191.

Gonzales-Martinez, R. (2018). The Wage Curve, Once More with Feeling: Bayesian Model Averaging of Heckit Models.

Guadagni, G. (2007). La curva de los salarios. *Boletín de lecturas sociales y económicas*, *2*(7), 76–79.

Haddad, E. A., Faria, W. R., Galvis-Aponte, L. A., & Hahn-De-Castro, L. W. (2016). Interregional input-output matrix for Colombia, 2012. *Borradores de Economia*, (923).

Jutting, J., Laiglesia, J. R. d., et al. (2009). *Is informal normal?: Towards more and better jobs in developing countries*. Development Centre of the Organisation for Economic Co-operation; Development.

Karatas, H. M. (2017). The Turkish spatial wage curve. *Econometrics*, *5*(3), 37.

Longhi, S., Nijkamp, P., & Poot, J. (2006). Spatial heterogeneity and the wage curve revisited. *Journal of Regional Science*, *46*(4), 707–731.

Lugo, D. C. (2006). Curva salarial: Una aplicación para el caso de México, 1993-2002. *Estudios económicos*, 233–273.

Pannenberg, M., & Schwarze, J. (2000). Wage dynamics and unemployment in Germany: Evidence from regional panel data. *Labour*, *14*(4), 645–655.

Parra Castro, M. A., & Váquiro Cuellar, K. L. (2017). *Curva de salarios: Sensibilidad del nivel de los salarios frente al nivel de desempleo en Colombia, 2013 – 2015*. https://doi.org/10.22518/issn.2462-909X

Ramos, R., Duque, J. C., & Surinach, J. (2010). Is the wage curve formal or informal? evidence for Colombia. *Economics Letters*, *109*(2), 63–65.

Ramos, R., Nicodemo, C., & Sanromá, E. (2015). A spatial panel wage curve for Spain. *Letters in Spatial and Resource Sciences*, *8*(2), 125–139.

Saavedra-Arango, J. (2016). Nuevas estimaciones de la curva de salarios para Colombia: 2007-2015.

Sánchez, F., & Nuñez, J. (1994). La curva de salarios para Colombia. *Estimaciones de las relaciones entre desempleo, la inflación y los ingresos laborales*, *1996*.

Sánchez, F., & Núnez, J. (1998). La curva de salarios para Colombia. (3).

Taylor, J. B. (1980). Aggregate dynamics and staggered contracts. *Journal of political economy*, *88*(1), 1–23.

Uchoa, C. F. A. (2019). An estimate of spatial wage curve for Brazilian northeast region. *Nexos Econômicos*, *13*(1), 63–79.

Appendices

A Tables

Table A1: Colombian Literature on the Wage Curve.

Author(s)	**Period**	**Cities**	**Formal-Informal**	**Methodology**	**Elasticity Value**
Sánchez and Núnez, 1998	1984-1996	7	No	OLS	-0.13
Ramos et al., 2010	2002-2006	13	Yes	OLS	All: -0.07 Informals: -0.17 Formals: -0.06
Arango et al., 2010	1984-2009	7 and 13	No	OLS	[-0.086, -0.093]
Saavedra-Arango, 2016	2007-2015	3	No	OLS	[-0.6, -1.4]
Parra Castro and Váquiro Cuellar, 2017	2013-2015	13	No	2SLS	-0.116

Table A2: Spatial Wage Curve along Formal-Informal Workers 2016-2019. IMR

	2SLS Contiguity		**2SLS Distance**		**2SLS Road Distance**		**2SLS Economic**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Informal	**Formal**	**Informal**	**Formal**	**Informal**	**Formal**	**Informal**	**Formal**
$LogU_{rt}$	-0.135***	-0.0961***	-0.138***	-0.0946***	-0.137***	-0.0975***	-0.136***	-0.0958***
	(0.0250)	(0.0256)	(0.0247)	(0.0254)	(0.0245)	(0.0253)	(0.0248)	(0.0255)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.00183*	0.00226***	-0.000778	0.00234**	-0.00608	0.0169***	-0.00181	0.00287***
	(0.00103)	(0.000857)	(0.00113)	(0.000960)	(0.00511)	(0.00432)	(0.00124)	(0.00104)
IMR	-2.051***	-5.790***	-2.051***	-5.790***	-2.051***	-5.790***	-2.051***	-5.789***
	(0.170)	(0.393)	(0.170)	(0.393)	(0.170)	(0.393)	(0.170)	(0.393)
R^2	0.299	0.490	0.299	0.490	0.299	0.490	0.299	0.490
Kleibergen-Paap Wald F Stat	23000	13000	24000	14000	24000	13000	24000	13000
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES
Observations	553,244	427,599	553,244	427,599	553,244	427,599	553,244	427,599

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table A3: Informal and Formal Spatial Wage Curve Along Males. IMR

	2SLS Contiguity			**2SLS Distance**			**2SLS Road Distance**			**2SLS Economic**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.151***	-0.164***	-0.155***	-0.152***	-0.167***	-0.154***	-0.153***	-0.166***	-0.158***	-0.152***	-0.166***	-0.155***
	(0.0220)	(0.0312)	(0.0333)	(0.0218)	(0.0308)	(0.0331)	(0.0216)	(0.0305)	(0.0330)	(0.0218)	(0.0310)	(0.0332)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.000112	-0.00227*	0.00278**	0.000697	-0.00116	0.00328***	0.00534	-0.00867	0.0220***	0.000233	-0.00221	0.00376***
	(0.000880)	(0.00130)	(0.00113)	(0.000976)	(0.00143)	(0.00127)	(0.00443)	(0.00645)	(0.00576)	(0.00107)	(0.00156)	(0.00137)
IMR	-4.560***	-1.753***	-5.259***	-4.560***	-1.753***	-5.259***	-4.560***	-1.753***	-5.259***	-4.560***	-1.753***	-5.258***
	(0.300)	(0.267)	(0.657)	(0.300)	(0.267)	(0.657)	(0.300)	(0.267)	(0.657)	(0.300)	(0.267)	(0.657)
R^2	0.439	0.294	0.497	0.439	0.294	0.497	0.439	0.294	0.497	0.439	0.294	0.497
Kleibergen-Paap Wald F Stat	20000	13000	7790.96	21000	13000	7925.85	21000	13000	7814.49	21000	13000	7873.22
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	530,714	292,346	229,880	530,714	292,346	229,880	530,714	292,346	229,880	530,714	292,346	229,880

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table A4: Informal and Formal Spatial Wage Curve Along Females. IMR

	2SLS Contiguity			**2SLS Distance**			**2SLS Road Distance**			**2SLS Economic**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.0541**	-0.105***	-0.0279	-0.0547**	-0.107***	-0.0260	-0.0551**	-0.106***	-0.0283	-0.0542**	-0.106***	-0.0273
	(0.0271)	(0.0394)	(0.0394)	(0.0268)	(0.0389)	(0.0390)	(0.0266)	(0.0386)	(0.0389)	(0.0269)	(0.0391)	(0.0392)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.000157	-0.00107	0.00156	0.000175	-0.000105	0.00110	0.00194	-0.00334	0.0102	-0.000155	-0.00105	0.00171
	(0.00106)	(0.00159)	(0.00130)	(0.00118)	(0.00176)	(0.00146)	(0.00532)	(0.00793)	(0.00648)	(0.00129)	(0.00192)	(0.00158)
IMR	-2.438***	-0.569*	-3.956***	-2.438***	-0.569*	-3.957***	-2.438***	-0.569*	-3.958***	-2.438***	-0.569*	-3.956***
	(0.295)	(0.300)	(0.788)	(0.295)	(0.300)	(0.788)	(0.295)	(0.300)	(0.788)	(0.295)	(0.300)	(0.788)
R^2	0.487	0.293	0.490	0.487	0.293	0.490	0.487	0.293	0.490	0.487	0.293	0.490
Kleibergen-Paap Wald F Stat	16000	11000	5509.15	16000	11000	5646.12	16000	11000	5555.69	16000	11000	5583.54
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	465,770	260,898	197,719	465,770	260,898	197,719	465,770	260,898	197,719	465,770	260,898	197,719

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table A5: Informal and Formal Spatial Wage Curve Along Lower Educated Workers. IMR

	2SLS Contiguity			2SLS Distance			2SLS Road Distance			2SLS Economic		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.122***	-0.147***	-0.128***	-0.124***	-0.151***	-0.126***	-0.124***	-0.149***	-0.128***	-0.123***	-0.148***	-0.127***
	(0.0210)	(0.0272)	(0.0350)	(0.0208)	(0.0269)	(0.0347)	(0.0207)	(0.0266)	(0.0347)	(0.0209)	(0.0270)	(0.0348)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	-0.00104	-0.00245**	0.00305***	0.000021	-0.00107	0.00350***	-0.00150	-0.00880	0.0173***	-0.000900	-0.00233*	0.00354***
	(0.000838)	(0.00112)	(0.00105)	(0.000931)	(0.00124)	(0.00118)	(0.00422)	(0.00560)	(0.00525)	(0.00102)	(0.00135)	(0.00128)
IMR	9.017***	11.24***	5.162***	9.017***	11.24***	5.162***	9.017***	11.24***	5.162***	9.017***	11.24***	5.162***
	(0.0924)	(0.141)	(0.0839)	(0.0924)	(0.141)	(0.0839)	(0.0924)	(0.141)	(0.0839)	(0.0924)	(0.141)	(0.0839)
R^2	0.265	0.202	0.253	0.265	0.202	0.253	0.265	0.202	0.253	0.265	0.202	0.253
Kleibergen-Paap Wald F Stat	24000	21000	4991.12	25000	21000	5083.15	25000	21000	4988.85	25000	21000	5044.94
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	626,278	431,251	185,885	626,278	431,251	185,885	626,278	431,251	185,885	626,278	431,251	185,885

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

Table A6: Informal and Formal Spatial Wage Curve Along Upper Educated Workers. IMR

	2SLS Contiguity			2SLS Distance			2SLS Road Distance			2SLS Economic		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**	**All**	**Informal**	**Formal**
$LogU_{rt}$	-0.0612**	-0.0881	-0.0694**	-0.0606**	-0.0887	-0.0676**	-0.0632**	-0.0888	-0.0718**	-0.0611**	-0.0880	-0.0696**
	(0.0291)	(0.0586)	(0.0347)	(0.0288)	(0.0576)	(0.0343)	(0.0286)	(0.0575)	(0.0342)	(0.0289)	(0.0581)	(0.0345)
$\sum_{j\neq r}^{J} \omega_{rj} LogU_{jt}$	0.000593	-0.000333	0.00146	0.000436	-0.000117	0.00105	0.00903	0.0000933	0.0145**	0.000740	-0.000588	0.00208
	(0.00114)	(0.00241)	(0.00123)	(0.00127)	(0.00264)	(0.00137)	(0.00575)	(0.0120)	(0.00621)	(0.00139)	(0.00289)	(0.00149)
IMR	6.273***	9.293***	4.941***	6.273***	9.293***	4.942***	6.273***	9.293***	4.941***	6.273***	9.293***	4.941***
	(0.0756)	(0.156)	(0.0811)	(0.0756)	(0.156)	(0.0811)	(0.0756)	(0.156)	(0.0811)	(0.0756)	(0.156)	(0.0811)
R^2	0.450	0.365	0.442	0.450	0.365	0.442	0.450	0.365	0.442	0.450	0.365	0.442
Kleibergen-Paap Wald F Stat	12000	3673.85	8497.18	12000	3808.29	8689.87	12000	3722.53	8577.71	12000	3755.93	8608.40
P-value Kleibergen-Paap LM Stat	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Region Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	370,206	121,993	241,714	370,206	121,993	241,714	370,206	121,993	241,714	370,206	121,993	241,714

Notes: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Robust standard errors in parentheses. The logarithm of unemployment rate by region in the previous period is used as instruments. Robust Kleibergen-Paap Wald F statistics suggest that the weak instruments null hypothesis is rejected. The null hypothesis of the Kleibergen-Paap LM statistic is that the equation is underidentified. Both regressions include controls as the individual characteristics are age, gender, marital status, education level, social security registration, individual's tenure, employment location, industry classification, occupational group, permanency of employment, firm size, and employment position.

B Efficiency Wages

Define E_i as worker's effort, w_i as the wage by the firm, and w_R as the wage in other firms. This is:

$$E_i = e(w_i, w_R) \quad e_{w_i} > 0 \quad e_{w_R} < 0 \tag{5}$$

The total magnitude of efficiency units of firm i, L_i, is obtained by multiplying the number of the firm's employees, N_i, by the effort of the workers, E_i; that is:

$$L_i = E_i N_i \tag{6}$$

The firm's profit is:

$$\pi_i = p_i AF(E_i N_i) - w_i N_i \tag{7}$$

where p_i is the price of good i and A is the technology. Thus, to maximize profit, the firm chooses N_i and w_i:

$$\frac{\partial \pi_i}{\partial N_i} = p_i A E_i F_L(E_i N_i) - w_i = 0 \tag{8}$$

$$\frac{\partial \pi_i}{\partial w_i} = p_i A N_i F_L(E_i N_i) e_{w_i}(w_i, w_R) - N_i = 0 \tag{9}$$

This implies:

$$\frac{w_i e_w(w_i, w_R)}{e(w_i, w_R)} = 1 \tag{10}$$

which means that the firm should find the wage for which the elasticity of the effort function is equal to 1. The firm must increase its wage to the extent that effort is increasing faster than that and, therefore, the wage per unit effort is falling.

For a wage-unemployment curve with a negative slope, several assumptions are made. First, it is assumed that the economy consists of two regions, 1 and 2, and that agents are risk-neutral; they obtain utility from income, which, for the case of region 1, is denoted by w_1, and dis-utility from effort, e. Utility function is given by:

$$V = w_1 - e \tag{11}$$

Let α be the probability that individuals who act with idleness will not be detected. Therefore, a probability $(1 - \alpha)$ that an individual who makes no effort will be detected, in which case he will be fired and will have to make an effort e to find another job. The expected utility of a laid-off worker is:

$$\bar{V} = (w_1 - e)\gamma(u_1) + \lambda_1[1 - \gamma(u_1)], \quad \gamma'(u_1) < 0, \tag{12}$$

where $\gamma(u_1)$ is the probability of finding a job, which depends inversely on the unemployment rate u_1 prevailing in the local market. Equation 12 is a convex combination of the utility of working by exerting effort, $w_1 - e$, and the joint value of unemployment insurance and leisure, λ_1. The unemployment rate in region 2 is u_2. This region, like region 1, is affected by shocks to labor demand.

Assume that workers separate from firms at a constant rate k, the rate of job destruction, and that new hires equal $\gamma(l_1 - n_1)$, where l_1 is the economically active population and n_1 is the employed population. In a steady state equilibrium it must be satisfied that $kn_1 = \gamma(l_1 - n_1)$. Thus, defining the unemployment rate as $u_1 = 1 - n_1/l_1$, we arrive at $\gamma = kn_1/u_1l_1$, which shows that the probability of finding a job depends inversely on the unemployment rate of the region.

The above model suggests the existence of a convex wage curve with a negative slope. If both regions have the same unemployment benefit ($\lambda_1 = \lambda_2$), they will have the same wage curve. This can be obtained by equating the expected utility of effort and inactivity. That is:

$$w_1 - e = \alpha w_1 + (1 - \alpha)\{(w_1 - e)\gamma(u_1) + \lambda_1[1 - \gamma(u_1)]\} \tag{13}$$

$$w_1 = e + \lambda_1 + \frac{\alpha e}{(1 - \alpha)[1 - \gamma(u_1)]} \tag{14}$$

$$\tag{15}$$

resulting in a negative relationship between wage rate and unemployment rate.