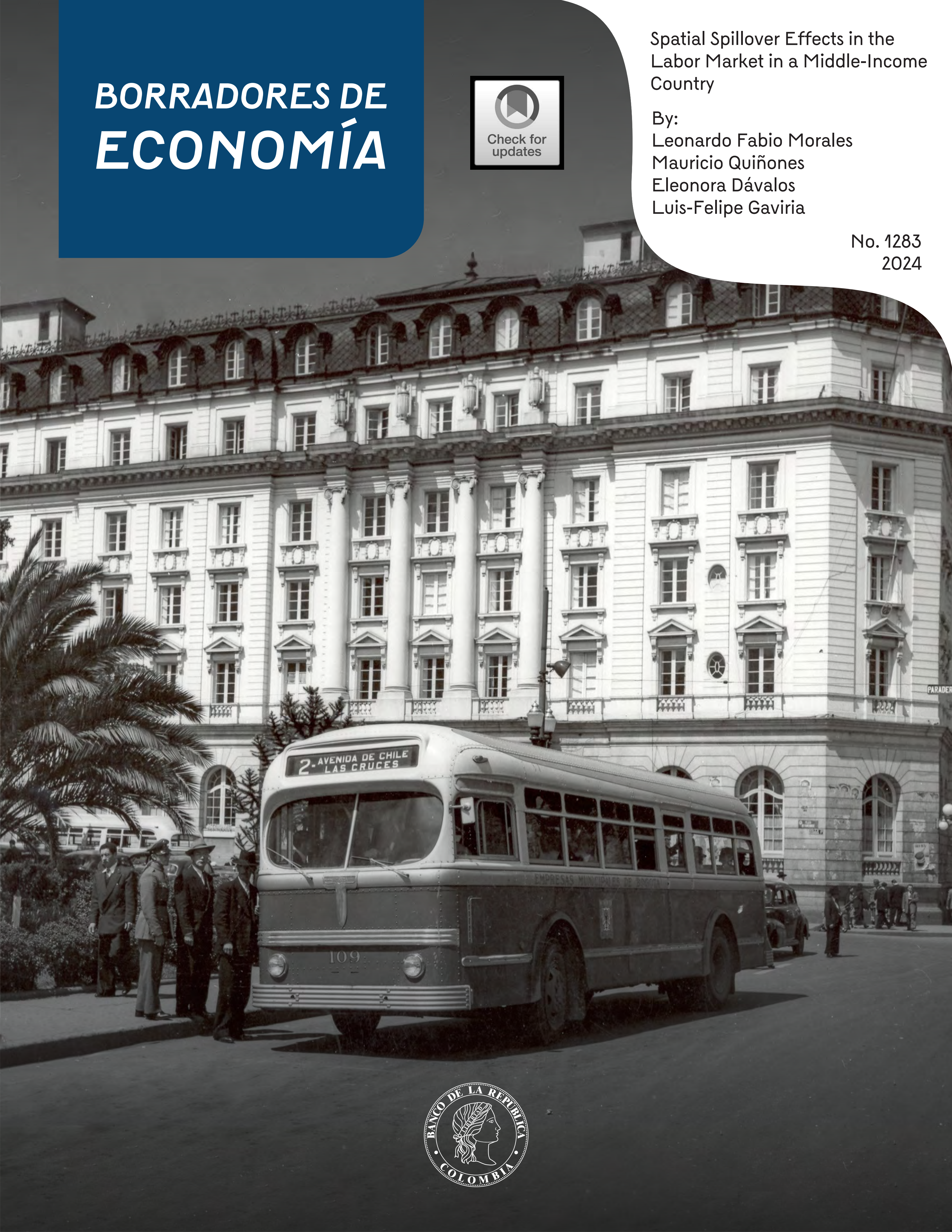
BORRADORES DE ECONOMÍA

Spatial Spillover Effects in the Labor Market in a Middle-Income Country

By:
Leonardo Fabio Morales
Mauricio Quiñones
Eleonora Dávalos
Luis-Felipe Gaviria

No. 1283
2024



BANCO DE LA REPUBLICA
COLOMBIA

Spatial Spillover Effects in the Labor Market in a Middle-Income Country[1]

Leonardo Fabio Morales, Mauricio Quiñones, Eleonora Dávalos, Luis-Felipe Gaviria

Abstract

Most macroeconomic labor literature on estimating matching functions does not consider spatial spillover effects. However, job search and vacancy-filling processes often involve neighboring locations, as local workers can search for and fill vacancies in nearby labor markets. We estimate a spatial spillover model using annual data for a middle-income country in Latin America. Our findings show that unemployment has a positive spatial spillover effect because an increase in the labor supply raises the probability of filling a vacancy. In contrast, vacancies have a negative spillover effect because local and neighboring vacancies compete to be filled by workers in both markets.

Keywords: Matching Function, Spatial Spillovers, Spatial Econometrics

JEL Codes: J61, J64, R12, R14.

[1] Leonardo Fabio Morales: Banco de la República lmoralzu@banrep.gov.co. Mauricio Quiñones: Pontificia Universidad Javeriana mauricio.quinones@javerianacali.edu.co. Eleonora Dávalos: Universidad EAFIT edavalosa@eafit.edu.co. Luis-Felipe Gaviria: Universidad EAFIT lfgaviriag@eafit.edu.co.

Efectos de derrame espacial en el mercado laboral de un país de renta media[2]

Leonardo-Fabio Morales, Mauricio Quiñones, Eleonora Dávalos, Luis-Felipe Gaviria

Resumen

La mayor parte de la literatura laboral sobre la estimación de funciones de emparejamiento no considera los efectos de derrame espacial. Sin embargo, los procesos de búsqueda de empleo y de ocupación de vacantes a menudo involucran ubicaciones vecinas, ya que los trabajadores locales pueden buscar y cubrir vacantes en mercados laborales cercanos. En este trabajo se estima un modelo de derrame espacial utilizando datos anuales para Colombia. Nuestros hallazgos muestran que el desempleo tiene un efecto de derrame espacial positivo porque un aumento en la oferta laboral aumenta la probabilidad de ocupar una vacante. En contraste, las vacantes tienen un efecto de derrame negativo porque las vacantes locales y vecinas compiten por ser ocupadas por trabajadores de ambos mercados.

Palabras clave: Función de Emparejamiento, Efectos Espaciales, Econometría Espacial

Códigos JEL: J61, J64, R12, R14.

[2] Leonardo Fabio Morales: Banco de la República lmoralzu@banrep.gov.co. Mauricio Quiñones: Pontificia Universidad Javeriana mauricio.quinones@javerianacali.edu.co. Eleonora Dávalos: Universidad EAFIT edavalosa@eafit.edu.co. Luis-Felipe Gaviria: Universidad EAFIT lfgaviriag@eafit.edu.co.

1 Introduction

Unemployment is frequently measured by cities or metropolitan areas (MA) because these spatial units are considered local labor markets. The stock of available open jobs is also measured by computing vacancies from firms in a city or MA. However, measuring these stocks at local labor markets ignores that spatial interactions can influence job search and vacancy-filling processes. Workers can search for jobs and find a match in neighboring cities or MA, while workers from nearby labor markets can fill local vacancies. This spatial interrelationship of generating new hires in the labor market occurs due to commuting between and across cities. The growing integration between prominent towns and their smaller neighbors explains a critical portion of this labor market dynamics. Therefore, a spatial perspective at the sub-national level is relevant for studying labor markets' search and matching processes.

The matching function is the fundamental theoretical tool to understand the generation of matches in the labor market. This function is an essential component of unemployment equilibrium models, frequently used to explain frictions and mismatches in labor markets (Pissarides, 2001). The function's inputs are unemployment and vacancy stocks, which determine new matches (hires). Yet, the matching function in its most standard form assumes that local labor markets are isolated, so there is no interaction between labor markets across the space (for a comprehensive survey of the literature, see Petrongolo and Pissarides, (2001)). However, recent literature on frictions and labor market search assesses spatial interactions in forming new employer-employee matches by including spatial spillover effects in matching function estimations (Burgess & Profit, 2001; Haller & Heuermann, 2016; Higashi, 2018; Ilmakunnas & Pesola, 2003; Kosfeld, 2007). The underlying question of this approach is whether the formation of new matches in a labor market is associated not only with local unemployment and vacancies but also with vacancies and unemployment in neighboring labor markets. Spatial spillovers from nearby labor markets might affect the ability of a local labor market to generate new hires because workers from neighboring locations can search for jobs in local labor markets and thus fill local vacancies, and local workers might search for jobs in neighboring communities and fill the vacancies produced in those markets.

The first effort to introduce spatial spillovers into unemployment equilibrium models posited that unemployment and vacancies in neighboring areas affected the formation of matches in local labor markets (Burda & Profit, 1996). This search model rationalized the existence of spatial spillovers in a search and matching framework, predicting that the effects of unemployment and vacancies across districts depended on the parameters of the theoretical model. In an assessment of the implications of their model, Burda and Profit (1996) find that unemployment spillover effects of closer districts are favorable for local labor markets.

The spillover effects of unemployment and vacancies are contingent on the dependent variable used for the matching function. Most empirical studies on matching functions use hires or the outflow of unemployed-to-employment.[1] Burgess and Profit (2001) illustrate how spillover estimation changes depending on the dependent variable of the matching function. Using a spatial lag of covariates model (SLX) the authors found that neighboring unemployment increases local hires, but decreases the outflow from local unemployment. They did not find any spillover effect coming from neighboring vacancies. Other studies that used the outflow from unemployment as a dependent variable found negative spillovers of unemployment (Ilmakunnas & Pesola, 2003; Kosfeld, 2007).

Recent studies incorporating state-of-the-art spatial econometric techniques tested several spatial effects specifications and used different spatial weight matrices. Higashi (2018) estimated an augmented matching function using SLX, spatial Durbin model (SDM), and spatial Durbin error model (SDEM) for labor markets in Japan. He used the traditional information criteria, AIC and BIC, to decide across specifications but did not use specialized statistical criteria to choose the best specification-matrix combination. In contrast to previous studies, Haller and Heuermann (2016) implemented the Bayesian criteria approach to choose their preferred combination of specification-matrix. In their study, they estimated augmented matching functions for spatial effects models using data from Germany. They argued that a positive spillover effect of unemployment and a negative spillover effect of vacancies should be expected from the search and matching framework. Unemployment in the neighboring labor market should increase local hires because a larger labor supply in the neighboring market increases the probability that those unemployed fill local open positions. Likewise, more vacancies in nearby markets should lower the likelihood of a local match because the probability that local unemployed find jobs in the neighboring district, instead of theirs, increases with the number of open positions there. The authors found a positive spillover effect of unemployment and a small positive spillover effect of vacancies. They explained that the latter result is due to the endogeneity of vacancies in neighboring regions.

We contribute to the literature in the following ways. Most of the literature on spatial spillovers in matching functions focuses on advanced economies' labor markets. We study the case of Colombia, a non-European middle-income economy. As Haller and Heuermann (2016), we use the Bayesian selection approach, a data-driven selection method of the best specification for the weight matrix. One of our main contributions to the literature is implementing a correction for endogeneity bias, which is available in the literature, to estimate spillover effects in matching functions. We provide evidence of the validity of our instrument and argue that our estimation is consistent and robust to the simultaneity bias. We show a comprehensive set of robustness checks that confirm our main findings, including the estimation of the SLX model with the optimal distance decay parameter proposed by Halleck Vega and Elhorst (2015).

This study estimates an SDEM for an augmented matching function. Using annual data for 1120 contiguous municipalities in Colombia between 2009 and 2019, we aim to understand how unemployment and vacancies in other cities affect hires in a local labor market. To control for the possible endogeneity bias that comes from unemployment, vacancies, and hires being simultaneously determined, we use a two-stage procedure that allows us to instrument possible endogenous variables (Ferreira-Neto, 2023; Halleck Vega & Elhorst, 2016; Rios, 2017; Zeilstra & Elhorst, 2014) . The SDEM includes unemployment and vacancy rates for a network of the neighboring labor markets defined throughout a weight matrix. Even though most of the literature assumes the relevant network of the labor market by imposing the weight matrices, we use a Bayesian selection approach to select the best matrix specification.

We find that local unemployment and vacancy rates positively affect hiring rates. As previous literature has pointed out, we find a positive spatial spillover of the unemployment rate on generating new matches (Burgess & Profit, 2001; Haller & Heuermann, 2016). This finding is consistent with the intuition of search and matching models, identifying that some of the unemployed workers in nearby cities will search for jobs in neighboring local markets, increasing the labor supply and facilitating the filling of local open job positions. In contrast, the neighboring market's vacancy rate negatively affects the local hiring rate. We argue that this result is also in line with equilibrium unemployment models; local vacancies will compete to be filled with vacancies in neighboring markets as the probability of a local worker being matched with a foreign vacancy increases with the number of open positions in the neighboring markets (Haller & Heuermann, 2016).

Our findings imply that a spatial perspective at the sub-national level should be considered in analyzing labor markets and formulating economic policies. Colombia is an example of many countries where economic policies affecting local labor markets are primarily determined at the municipality level. The effects of these policies would have effects beyond municipalities' borders. Therefore, policy coordination is a requirement for better results. These policies should promote the ease of moving to jobs (Gobillon & Selod, 2019) and improve efficiency through active search initiatives. Implementing targeted active labor market policies could also address regional disparities (Haller & Heuermann, 2016). Local public employment centers in small cities might connect workers with job opportunities in nearby regions. Concurrently, they could offer firm information about well-suited local candidates for their vacant positions.

2 Theoretical framework

This section presents a microeconomic rationale for a matching function and expands the framework to include spatial spillover effects. The basic model assumes that, in a local labor market j, unemployed workers, U_j, search for jobs and send one application to each of the vacancies available, V_j. Vacancies receive one or more applications and randomly select a worker to form a match. Frictions, such as lack of information or mismatch, can be modeled in various ways. However, because multiple workers can apply for the same vacancy and only one can fill it, some will remain unemployed, and some vacancies will remain unfilled. According to Petrongolo and Pissarides (2001), the expression for the matches generated in a period is derived from the probability that a vacancy remains without any application, which is given by the expression $\left(1-\frac{1}{V_j}\right)^{U_j}$. This expression is the complement of the probability that a vacancy receives an application, and given that there are V_j vacancies, it is raised to the power of U_j because each worker sends an application. The following equation represents the total matches generated:

$$H_j = V_j \left[1 - \left(1 - \frac{1}{V_j}\right)^{U_j}\right] \tag{1}$$

The variable H_j denotes the number of hires, equivalent to the product of the stock of vacancies and the probability of receiving an application. As shown in equation (2), the matching function in equation (1) is commonly estimated as a Cobb-Douglas hiring function concerning unemployment and vacancies. The literature estimates matching functions using vacancies, hires, and unemployment levels. Some studies use unemployment and vacancy rates or indexes (Blanchard & Diamond, 1989; Burgess, 1993; Layard, Nickell, & Jackman, 1991; Mumford & Smith, 1999). In this paper, we prefer the latter approach to the former because we include labor markets of different sizes in our regressions. Expressing all variables in relative terms could enhance comparability.

$$H_j = \alpha + \beta_U U_j + \beta_V V_j + u_j \tag{2}$$

Neither equation (1) nor equation (2) considers any spillover effects, assuming that the local labor market is isolated from other markets, with no influence from neighboring markets on the matching process. However, this assumption may be too strong, as job search and vacancy-filling processes often involve workers and vacancies from neighboring locations. Local workers can fill vacancies in neighboring areas, and neighboring workers can fill local vacancies. To account for these spillover effects, we propose the following augmented matching function:

$$H_j = V_j \left[1 - \left(1 - \frac{1}{V_j}\right)^{K_j}\right] \tag{3}$$

$$K_j = \left\{U_j - \left[s_j\left(V_{-j}\right) * U_j\right] + s_{-j} * U_{-j}\right\}$$

Equation (3) is similar to equation (1), but the exponent U_j has been replaced by the function K_j, which represents the number of unemployed workers searching for jobs in the local market. The determination of K_j is a process that considers the share of local unemployed workers searching for jobs in neighboring markets and the share of workers from neighboring markets searching for jobs in the local market. The first term in the K_j function accounts for the number of local unemployed workers searching in the local market, adjusted for the local workers who search in the neighboring market instead. This adjustment is calculated using the expression $s_j(V_{-j}) * U_j$, where $s_j(V_{-j})$ is the share of local workers searching outside, which itself depends on the number of vacancies in neighboring markets. The last term in the K_j function represents the share of workers from neighboring locations who search for jobs in the local market and not in their location. This term is calculated using the share of outside searchers from neighboring locations multiplied by the number of unemployed workers in those locations. Summarizing, as it is described in equation (3), a matching function that incorporates spillover effects must be a function not only of unemployment and vacancies in the local market but in nearby markets as well.

To estimate a modified version of equation (3), we follow the convention in the literature by assuming a Cobb-Douglas function. Our dependent variable is the hiring rate, while our explanatory variables include the local and neighboring unemployment and vacancy rates. We express all variables in rates because local and neighboring markets can differ significantly in size, with large cities often surrounded by medium or small towns. We specify the model in a log-linear form, common in the literature, as previous studies have found that a log-linear approximation is a good fit for aggregate matching functions (Petrongolo & Pissarides, 2001). Thus, the empirical equation we estimate can be represented as:

$$\ln(Hr_{jt}) = \alpha + \beta^U Ur_{jt} + \beta^V Vr_{jt} + \theta^U Ur_{-jt} + \theta^V Vr_{-jt} + u_{jt}, \tag{4}$$

where Hr_{jt}, Ur_{jt}, Vr_{jt} stand for local hiring, unemployment, and vacancy rates, respectively. On the other hand, Ur_{-jt} and Vr_{-jt} represent aggregated measures of neighboring unemployment and vacancy rates. As detailed in section 5, empirical estimation, we construct these aggregates using standard methods from spatial econometric models.

3 Data and context on Colombian labor markets

Colombia is a country divided into 32 *departamentos*, political units equivalent to states or provinces. States in Colombia are divided into municipalities, the most local form of government. In this research, municipalities (i.e., cities) are our units of analysis. According to the Colombian National Department of Statistics (DANE), as of 2023, there are 1123 municipalities, considering in these numbers 18 territories with special status.[2] Out of the 1123 municipalities, four units have been created since 2007; we only have complete information for one. Therefore, the total number of municipalities included in our estimations is 1120, excluding the three recently created municipalities with incomplete information. This last number virtually covers all Colombian land.

In this paper, we use three primary data sources. First, we use data on labor market outcomes from the DANE's official household survey, the GEIH (acronym in Spanish). We also use administrative records from the social security Contributions in Colombia. Finally, we use Colombian Census data for two different censuses. These data sources allow the construction of the variables of interest. Our definition of a labor market is each one of the 1120 municipalities for several reasons. A municipality is the most granular geographical unit for which the available data sources allow the construction of the variables of interest for the estimation. Exploring the matching process in labor markets at the lowest level of available geographical data is beneficial for the primary purpose of this research: studying the spillover effects of changes in unemployment and vacancies across labor markets. Having small geographical units helps to identify spillover effects that occur from workers' commuting practices from one city to another.

Using larger geographical units such as MA could be problematic because the definition of MA in Colombia is unclear and outdated, sometimes depending upon historical political agreements reached for some cities and not others. In Colombia, there are only seven official MAs, which are comprised of more than one city; nevertheless, plenty of well-connected cities share borders and do not belong to an official MA cluster. It is unclear why cities are not included in some MA clusters where commuting is a well-established practice between them and their neighboring towns.[3] Furthermore, quantifying the spillover effects within pre-established MA is still of interest for this research; this is especially relevant because economic policies are primarily determined at the municipality level, regardless of whether the city belongs to an MA cluster or not.

The GEIH randomly surveys an average of 23,000 Colombian households monthly. Despite being carried out monthly, its results are representative quarterly for 23 MA in Colombia (with a total of 40 cities) and annually for the 23 states in which these 40 major cities are located. From the GEIH, we can extract the annual unemployment rate for these 40 main cities and their 23 states but not for the other municipalities (1080) or states (9). For them, we need to carry out an interpolation process to fill in the yearly missing values for this variable. We explain this process in detail below.

3.1 Interpolation of labor market outcomes

To estimate annual unemployment for the rest of the municipalities for which the GEIH is not representative, we use the Colombian Population censuses of 2005 and 2018. The census information allows the computation of the unemployment rate for 2005 and 2018 for all municipalities. With this information, we interpolate the years between censuses for municipalities not covered in the household survey (1080) using the annual variation of the departmental unemployment rate. Given that we only have these rates for the 23 states covered by GEIH, we only interpolate for the municipalities within each of these states in the first part of the procedure. The process is as follows: from the Population Census, we have an initial value of the unemployment rate for each municipality in 2005. We multiply this value by the annual growth of the unemployment rate of the respective department to fill in the missing value in 2006. We carry out this filling process for each year between 2006 and 2017. In 2018, we used again the official municipal unemployment rate for all municipalities. Consequently, we use this value as the baseline for the interpolation for 2019, following the same procedure described above.

To interpolate the unemployment rate between censuses for small municipalities in remote states with no representation in the GEIH (i.e., municipalities in 9 states not covered in this survey), we use the same procedure, but the reference for the interpolation is the annual unemployment growth rate of the aggregated rural area for the whole country, for which this survey is representative. Thus, with the process described above, we can compute the annual unemployment rate for 1120 municipalities in Colombia during the 2010-2019 period. Our estimation period begins in 2009; nevertheless, we start the linear interpolation in 2005 because this procedure requires an initial unemployment value. In addition to unemployment, population censuses allow a reliable calculation of the employed and out-of-the-labor market population at the municipality level. Using these populations, we compute the respective unemployment, employment, and labor force rates, among other indicators. For these variables, we use the same interpolation process.

3.2 Vacancies

We use administrative records from the social security Contributions in Colombia. More specifically, we use an information system managed by the Ministry of Health, in which all firms must register the payroll taxes, payments, and pension contributions by employers in every municipality of Colombia; this information system is known as PILA (its acronym in Spanish). The PILA allows the construction of an employer-employee panel from which we can compute hiring levels. Similarly, we use the information from PILA to construct the stock of vacancies using the methodology in Morales and Lobo (2020); in this methodology, the stock of vacancies is computed as the number of open positions that are supported by the level of hiring and separations observed within each firm. We aggregate this information by municipalities for the 2010-2019 period. We construct vacancy and hiring rates with the stock of vacancies and the flow of hires. The former is the number of vacancies in a municipality as the share of all job positions (total employment plus vacancies); the latter is the total hires in a city as a proportion of total employment.

Our estimation sample is an annual panel of 1120 municipalities in Colombia for the 2010-2019 period. Summary statistics of the variables we use in the analysis are presented in Appendix A.1.[4] Average unemployment, vacancy, and hiring rates by municipality are 11%, 3.4%, and 5.6%, respectively. We acknowledge that we do not directly observe the vacancy rate, which introduces measurement error in the estimations; this might result in efficiency issues. Nevertheless, in robustness checks, we estimate a matching function using a subsample of the panel for which there is an alternative measurement of vacancies collected by the Public Employment Service Bureau of Colombia (2015-2019). The paper's main findings hold in the estimations with this alternative vacancy rate.

4 Empirical estimation

4.1 Specification and estimation

To quantify spatial spillover effects in the formation of new hires, we need to perform exploration tests to corroborate that there are spatial effects in the variables we want to model. After that, we need to find a specification for the spatial model and a weight matrix. In addition, our estimation can be subject to endogeneity bias because the hiring, unemployment, and vacancy rates are simultaneously determined in the local labor market. Regarding the exploration tests, we perform the traditional global and local Moran's I tests to detect spatial effects of the dependent and independent variables. The global Moran's I for hiring, unemployment, and vacancy rates indicates spatial autocorrelation for all three variables (see Appendix A.2).

The main results for local Moran's I (one of the local indicators of spatial association—LISA) for 2009 and 2019 indicate a cluster of municipalities with high (low) unemployment surrounded by other cities with high (low) unemployment; the same is valid for vacancies and hires (see Appendix A.3). The first line of maps presents the municipal hiring rate. In 2009, the country's western region (the Pacific region) showed a clustering pattern with high hiring rates, while the Amazon region in the south presented a clustering pattern with low hiring rates. In 2019, the western cluster disappeared while the pattern persisted in the Amazon region. The second line of maps presents the municipal unemployment rate. The Pacific region showed a widespread pattern of high clustering. This result is consistent with the theoretical framework, which suggests a direct relationship between hires and unemployment. The third line of maps shows the municipal vacancy rate. As with previous results, there is a persistent pattern of high vacancy rate clustering in the Pacific region. There is also a low vacancy rate clustering in the Andean region (central region).

Based on this evidence, we cannot reject the hypothesis of spatial autocorrelation in our variables of interest. We also use the Lagrange multiplier test (LM), standard and robust, to shed light on the possible econometric specifications of the models (Anselin, Bera, Florax, & Yoon, 1996). These tests start from a non-spatial estimation and test the non-spatial model against an autoregressive spatial or spatial error model. The results of these tests are presented in Appendix A.4. We reject the hypothesis of the non-existence of spatial effects in the error term and the spatial lag of the dependent variable.[5]

The specification that would allow the estimation of the spatial spillover effects of unemployment and vacancy rates is a model that includes a spatial lag, that is, the multiplication of each of these variables with a spatial weight matrix defining the network of neighboring labor markets. A model with these characteristics is represented in equation (5), where Hr_{it}, Ur_{it}, and Vr_{it}, represent the hiring, unemployment, and vacancy rates of municipality i at year t, respectively. Coefficients β^U and β^V capture the effect of unemployment and vacancies on the local hiring rate. The specification in equation (5) is better known as an SLX model. It includes the interaction between independent variables and the weight matrix. This specification allows estimating the effect of neighboring unemployment and vacancy rates on local hiring rates, which are captured by the parameters θ^U and θ^V, respectively. Finally, the estimation includes municipality and time-fixed effects (c_i, α_t) and an idiosyncratic error term v_{it}.

$$Hr_{it} = \beta^U Ur_{it} + \beta^V Vr_{it} + \sum_{j=1}^{N} w_{ij}\, Ur_{jt}\theta^U + \sum_{j=1}^{N} w_{ij}\, Vr_{jt}\theta^V + c_i + \alpha_t + v_{it} \tag{5}$$

The results of the LM tests indicate that the specification should include spatial autocorrelation terms in either the dependent variable or the error term. According to the suggestions of the LM tests, two possible specifications for estimating a spatial matching function would be a spatial autoregressive model (SAR) or a spatial error model (SEM). However, these two models do not include spatial interaction of unemployment and vacancies. SDM and SDEM are comprehensive specifications that include spatial lags of the independent variables. The former starts from the specification in equation (5) but includes a spatial lag term of the dependent variable. The latter reproduces the structure of the SLX model as well; however, the error term is defined as a function of its spatial lag. These two specifications are represented in equations (6) and (7), respectively.

Equation 6: Spatial Durbin model (SDM)

$$Hr_{it} = \rho \sum_{j=1}^{N} w_{ij}\, Hr_{jt} + \beta^U Ur_{it} + \beta^V Vr_{it} + \sum_{j=1}^{N} w_{ij}\, Ur_{jt}\theta^U + \sum_{j=1}^{N} w_{ij}\, Vr_{jt}\theta^V + c_i + \alpha_t + v_{it} \tag{6}$$

Equation 7: Spatial Durbin error model (SDEM)

$$Hr_{jt} = \beta^U Ur_{it} + \beta^V Vr_{it} + \sum_{j=1}^{N} w_{ij}\, Ur_{jt}\theta^U + \sum_{j=1}^{N} w_{ij}\, Vr_{jt}\theta^V + c_i + \alpha_t + v_{it} \tag{7}$$

$$v_{it} = \delta \sum_{j=1}^{N} w_{ij} v_{jt} + \varepsilon_{it}$$

The SDM and SDEM nest the SAR and SEM models, respectively; therefore, it is possible to test whether the data support the more comprehensive specification. This assessment is a simple Wald test of the joint significance of the parameters θ^U and θ^V, respectively. As shown in the next section, we reject the null hypothesis that $\theta^U = \theta^V = 0$ for all models we estimated using a variety of weight matrices.

In this paper, we choose the SDEM for several reasons; in the SDEM model, the spillover effects are local, meaning that they circumscribe only to the network defined by the weight matrix (Elhorst, 2014). Most of the literature that estimates spillover effects in labor markets argues that local spillover should be more consistent with theoretical frameworks where these effects are explained by commuting patterns of workers from neighboring locations (Ferreira-Neto, 2023; Halleck Vega & Elhorst, 2016; Patacchini & Zenou, 2007). In addition, previous literature has reported evidence on the local nature of the spillover effects using Bayesian selection methods; for instance, Haller and Heuermann (2016) for Germany. The spillover effects depicted in equation (3) of our theoretical framework are based on seekers filling vacancies in neighboring labor markets. This situation will likely occur within a network where commuting is possible. Therefore, it is more natural in our framework to assume that spillovers are local. We also estimate SLX models as a robustness check.

The SDEM is also a more convenient specification for endogeneity bias correction; it allows a two-stage instrumental variables procedure to tackle this problem. In this paper, we estimate SDEM models using a two-stage procedure, similar to the estimation strategy proposed by Ferreira-Neto (2023), which in turn is based on contributions by other studies (Halleck Vega & Elhorst, 2016; Rios, 2017; Zeilstra & Elhorst, 2014). In the first stage of the procedure, we regress endogenous variables as a function of instrumental variables and fixed effects. In the second stage, we used the prediction of the second stage to estimate equations for the SLX and the SDEM, equations (5) and (7), respectively.

For all previous reasons, we estimate SDEM and SLX specifications; nevertheless, traditional information criteria cannot be used to compare SDEM and SLX models that use different weight matrices to define the relevant spatial network. Because of this, we use modern Bayesian comparison approaches to compare these specifications and indicate the weight matrices most supported by the data. In the next section, we use these techniques to select the best relevant network definition in terms of the weight matrices. Finally, as an alternative way of choosing the more suitable network definition, we estimate an SLX model with the optimal distance decay parameter proposed by Halleck Vega and Elhorst (2015) in a robustness check. The main findings of the research hold regardless of the model or the network selection strategy.

4.2 Instrumental variables

In estimating the matching function represented in equations (5) and (7), we have four endogenous variables; therefore, we need at least four instrumental variables to identify the model. We construct instruments based on national-level measures of labor market indexes. For this purpose, we follow literature using Bartik-type instrumental variables (IVs); these IVs map national labor market indexes to local changes in labor demand (Autor & Duggan, 2003; Bartik, 1991; Blanchard & Katz, 1992; Bound & Holzer, 2000; Davis & Haltiwanger, 2014; Morales & Medina, 2019). The instruments are constructed by averaging labor-market indexes across non-neighboring labor markets, using weights representing the degree of similarity between the local labor market and those used to compute the weighted average. For our baseline estimation, we use four instruments. The general form of all of them follows this equation:

$$\tilde{I}_{jt} = \sum_{i \neq j} \overline{\omega}_{jit-1} * I_{it-1}, \tag{8}$$

where I_{it-1} is the local labor index in all labor markets different from local market j in the previous year; $\overline{\omega}_{jit-1}$ is the *i-j* entry of a row standardized weight matrix, which represents the inverse of a similarity index given by the squared of the difference between market i and market j in terms of their labor force participation rate in the previous year. Each element of this inverse similarity matrix (previous to standardization) can be represented as:

$$\omega_{jit-1} = 1\{d_{ij} > K\} * \frac{1}{\sqrt{\left(LFPR_{jt-1} - LFPR_{it-1}\right)^2}} \tag{9}$$

In equation (9), $LFPR$ stands for the labor force participation rate. In addition, the expression $1\{d_{ij} > K\}$ represents an indicator function that takes the value of 1 if the Euclidean distance, d, between the centroids of local labor markets i and j is greater than K, and zero otherwise. This last factor implies that no contiguous labor market is used to compute the weighted average; in the baseline estimation, we use K = 50 km, and in the results section, we show that our findings are robust to using larger distances. We use four labor market indexes to compute the instruments $\tilde{I}_{jt}$: The vacancy rate, the unemployment rate, the employment-to-population ratio, and the labor participation rates. All these variables lagged one year correspond to non-neighboring labor markets with centroids at least 50 km further away from the centroid of the local market.

5 Results

This section presents the results of the augmented matching functions to capture spatial spillover effects. Our baseline estimation is the two-stage procedure described in the previous section; nevertheless, we also present results for non-instrumental variables for comparison. In the first stage, we regress endogenous variables as a function of instrumental variables and fixed effects. In the second stage, we used the prediction of the first stage to estimate equations for the SLX and the SDEM models. In these equations (5) and (7), the hiring rate is a function of unemployment and vacancy rates and aggregated measures of the same variables in neighboring labor markets. We estimate each equation for eight different spatial weight matrices: queen contiguity matrices order 1, 2, and 3, and contiguity matrices for neighbors within 20 km, 40 km, 60 km, 80 km, and 100 km.[6] The comparison between spatial models using different weight matrices is not straightforward, and the election of the weight matrix is challenging. We tackle these difficulties using a Bayesian comparison approach.

We follow Costa da Silva, Elhorst, and Silveira Neto (2017) to compare across specifications (SDEM, SLX) and each of the eight weight matrices we consider in this study. We compute Bayesian posterior probabilities. The log-marginal likelihood and the posterior Bayesian probability can be calculated by integrating each model's vector of parameters over the parameter space. The duple of model-matrix with the highest Bayesian posterior probability is the one that most likely generated the sample (Dávalos & Morales, 2023). Therefore, we compute these probabilities for each combination specification-matrix couple and select the one with the highest probability as our preferred specification.

We present a two-stage non-spatial regression of the matching function in Appendix A.5. The coefficients of the semi-elasticities are 0.17 and 0.30 for local unemployment and vacancies, respectively. These elasticities are comparable to the ones we estimate using the two-stage procedure with the SDEM and the SLX models. Table 1 presents all the estimated models for the SDEM, and Table 2 presents all estimated models for the SLX specification. Finally, the results of estimating spatial non-instrumental variables specifications are presented in

Appendix A.6. For all specifications in Tables 1 and 2, we present at the bottom of each table the Bayesian posterior probabilities and the Wald test of the joint significance of the coefficients θ^U and θ^V, which allows testing whether the models can be nested to simpler versions with no spatial lags of the independent variables. In all specifications and for all matrices, we reject the hypothesis that these parameters are jointly not statistically significant. Finally, we also present the F-statistics of the first-stage regression for the endogenous variables at the bottom of Tables 1 and 2. In all cases, our instruments show they are valid in terms of the correlation with endogenous variables, with F-statistics of 35 and 29 for the unemployment and vacancy rates, respectively.

The local effects of unemployment and vacancy rates on hiring are similar across all two-stage specifications. In all the SDEM models, the spatial autocorrelation error term (δ) is statistically significant and positive, which constitutes evidence of spatial spillover effects in the determination of the local hiring rate, and it signals that SDEM should be preferred over the SLX model; nevertheless, we also consider SLX specification in our comparisons, and we let the Bayesian selection approach to confirm the SDEM as the preferred specification.

We compare the 16 models in Tables 1 and 2 using the Bayesian comparison approach following Costa da Silva et al. (2017). We calculate the Bayesian posterior probabilities for the SLX and SDEM models using various matrices. The log-marginal likelihood, which represents the model's probability, is obtained by integrating the model's parameters over the parameter space. The methodology involves finding the probabilities for different models and matrices, and the model-matrix pair with the highest probability is considered the most likely source of the sample. We obtain that the specification supported by the data is an SDEM with a matrix that defines as neighbors those local markets with a centroid inside a 20 km radius.[7]

Marginal effects can be obtained by taking the partial derivative of the expected value of the dependent variable with respect to a covariate; in the case of SDEM and SLX models, these are captured directly by the estimated coefficient (Elhorst, 2014; LeSage & Pace, 2010). They can be broken down into direct, indirect, and total effects, the latter being the sum of the first two. The direct effect is the impact of a change in a covariate on the dependent variable in the same spatial unit. In the SDEM and SLX models, this effect is captured by the coefficients of unemployment and vacancy rates. In contrast, the indirect effect is the impact on the dependent variable of other spatial units' covariates. In the SDEM and SLX models, the coefficients of the spatial lags of the unemployment and vacancy rates capture the indirect effects. We present the marginal effects of the preferred specification in Table 3.

On the one hand, the direct effect of an increment of 1pp in unemployment is an increase of the hiring rate by 15.7%; this is equivalent to a marginal rise of 0.84pp in the hiring rate for a 1pp increase in unemployment rate. On the other hand, an increment of 1pp in the vacancy rate is associated with an increase of 33% in the hiring rate, which is equivalent to a marginal rise of 1.8pp in the hiring rate given a 1pp increment in the vacancy rate. These findings align with most existing empirical and theoretical research on matching functions, suggesting that local unemployment and vacancies should positively affect the number of matches (Petrongolo & Pissarides, 2001).

Equilibrium unemployment models predict that an increase in the unemployment rate in neighboring regions is expected to result in more matches in a local market due to the larger labor supply, increasing the likelihood of successful matches. However, an increase in the number of job opportunities in neighboring regions (as indicated by a higher vacancy rate) is likely to decrease matches in the local labor market, as the probability of finding employment in those regions, instead of the local market, increases (Burda & Profit, 1996; Haller & Heuermann, 2016). In other words, vacancies in local and neighboring markets compete to find a suitable worker. The indirect effects, computed using the preferred specification, are presented in Table 3. The results show 7.3% and -19% semi-elasticities for unemployment and vacancies, respectively. Therefore, increments of 1pp in the neighboring markets' unemployment and vacancy rates would increase the local hiring rate by 0.4pp and -1pp, respectively. The total impact of unemployment and vacancy rates, the combination of direct and indirect effects, corresponds to semi-elasticities of 23% and 15%, respectively, given changes of 1pp, which correspond to marginal effects of 1.2pp and 0.86pp in the hiring rate given changes of 1pp in unemployment and vacancy rates, respectively.

We can summarize our results as a sizeable and positive effect of unemployment and vacancy rates in the formation of new hires and a considerable and significant spillover effect, which is positive in the case of unemployment and negative in the case of the vacancy rate. All these findings hold, as presented in Table 1 when using different matrices for the SDEM, in which case, the magnitude of the effects is greater than the one obtained using our preferred specification. As presented in Table 2, our findings are robust to using the SLX model for all different matrices used in this paper. In the case of the matrix supported by the Bayesian probabilities, the magnitudes of the coefficients estimated with the SLX model are similar to the ones estimated with our preferred specification. In the case of the SLX models with additional matrices, the effects have the same signs and larger magnitudes than the ones obtained with the preferred specification.

Our results are also robust to changes in instrumental variables. In Appendix A.7 and Appendix A.8, we present the results of our SDEM using variations in the construction of our instruments. Table A.7 shows the following variation: We use a different configuration for the expression $1\{d_{i,j} > K\}$ in equation (9). In this case, instead of using K = 50 km, we use K = 150 km; therefore, we exclude the weighted average from the computation of our instruments' labor markets inside a buffer of a 150 km radius. Results are consistent with our baseline estimation regarding the sign and significance of the coefficients; the magnitudes of the direct effects are relatively similar. Our main findings are robust to an additional configuration in the instrumental variables. In Appendix A.8, we present the results of the two-stage estimation using instruments that do not impose the restriction of excluding neighboring labor markets for their computation (i.e., $K = 0$). As before, the results of this latter estimation are consistent with the ones in our baseline estimation regarding the coefficients' sign, significance, and magnitude.

Our results are also robust to using an alternative source of information for the vacancy rate. In Appendix A.9, we present the results of the estimation of the two-stage SDEM, as the one presented in equation (7); nevertheless, for computation of the vacancy rate, we use the vacancies provided by the Special Public Employment Service Unit, a government agency created for this purpose (Morales, Ospino, & Amaral, 2021). Unfortunately, these data are only available from 2015, reducing our estimation sample for these robustness checks. The results of these estimations confirm our main findings; for instance, for the same weight matrix of our baseline estimation, the direct effect of an increment of 1pp in the unemployment (vacancy) rate is an increase in the hiring rate of 11% (18%). In the case of the indirect effect, an increment of 1pp in the unemployment (vacancy) rate is associated with a change of 8% (-7%) in the hiring rate. The magnitudes of these effects are smaller than our baseline estimation, but they are still comparable.

In a final robustness check, we use a particular advantage from the SLX model: it provides an alternative way of selecting the spatial weight matrix. We follow the methodology presented in Halleck Vega and Elhorst (2015), where the authors allow for parameterizing the spatial weight matrix $\mathbf{W}$. They model each entry of the weight matrix $w_{ij} = 1/d_{ij}^{\gamma}$, where d_{ij} is the distance between two spatial units, and γ is the optimal distance decay parameter, estimated jointly with the rest of the parameters in the regression. Appendix A.10 presents the results of the SLX model with the optimal distance decay parameter. The main findings hold when we use this alternative estimation in the two-stage framework. The estimated direct and indirect effects coefficients have the same signs and are statistically significant with higher magnitudes.

Table 1: Spatial Durbin error model. Comparison of the estimation results

	Change on municipal log of hiring rate															
	Queen 1st		*Queen 2nd*		*Queen 3rd*		*d < 20 km*		*d < 40 km*		*d < 60 km*		*d < 80 km*		*d < 100 km*	
Variables																
Unemployment rate	0.135***	(0.019)	0.129***	(0.019)	0.126***	(0.019)	0.157***	(0.019)	0.142***	(0.019)	0.131***	(0.019)	0.130***	(0.019)	0.131***	(0.019)
Vacancy rate	0.352***	(0.042)	0.341***	(0.041)	0.352***	(0.042)	0.336***	(0.042)	0.321***	(0.042)	0.339***	(0.042)	0.339***	(0.042)	0.338***	(0.042)
W*Unemployment rate	0.262***	(0.040)	0.438***	(0.073)	0.462***	(0.070)	0.073***	(0.025)	0.205***	(0.036)	0.354***	(0.049)	0.437***	(0.064)	0.503***	(0.086)
W*Vacancy rate	-0.464***	(0.088)	-0.727***	(0.156)	-0.790***	(0.149)	-0.190***	(0.056)	-0.296***	(0.079)	-0.557***	(0.106)	-0.711***	(0.137)	-0.838***	(0.180)
W*Error	0.192***	(0.014)	0.420***	(0.021)	0.393***	(0.020)	0.108***	(0.012)	0.243***	(0.017)	0.359***	(0.020)	0.437***	(0.022)	0.484***	(0.024)
Constant	1.019***	(0.007)	1.008***	(0.007)	1.009***	(0.007)	1.028***	(0.007)	1.018***	(0.007)	1.012***	(0.007)	1.008***	(0.007)	1.008***	(0.007)
Observations	12,320		12,320		12,320		12,320		12,320		12,320		12,320		12,320	
Number of groups	1,120		1,120		1,120		1,120		1,120		1,120		1,120		1,120	
Year FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Area FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
F-Test unemployment rate instrument	35.31		35.31		35.31		35.31		35.31		35.31		35.31		35.31	
F-Test vacancy rate instrument	29.10		29.10		29.10		29.10		29.10		29.10		29.10		29.10	
Wald-Test pval	0.000		0.000		0.000		0.003		0.000		0.000		0.000		0.000	
Bayesian comparison probabilities																
Model probability	0.145		0.304		0.460		0.484		0.119		0.103		0.210		0.112	

Notes: Standard errors in parentheses. Results of equation (7) by SDEM. The outcome variable used is change on the natural logarithm of the hiring rate. The sample includes all Colombian contiguous municipalities with available data from 2010 to 2019. The year 2009 is lost after lagging the series. The spatial weight matrix Queen 1st was generated using a queen contiguity method, first-order neighbors. The spatial weight matrix Queen 2nd was generated using a queen second-order neighbors. The spatial weight matrix Queen 3rd was generated using a queen second-order neighbors with 1 for first-order neighbors and 0.5 for second-order neighbors. $d < 20$, $d < 40$, $d < 60$, $d < 80$, and $d < 100$ km refer to contiguity matrices for neighbors within 20, 40, 60, 80 and 100 km, respectively. Municipal and year fixed effects regressors are not shown. See Appendix A.1 for description and summary statistics for all the variables.

**** p<0.01, ** p<0.05, * p<0.10*

Table 2: Spatial lag of X model. Comparison of the estimation results

	Change on municipal log of hiring rate															
	Queen 1st		*Queen 2nd*		*Queen 3rd*		*d < 20 km*		*d < 40 km*		*d < 60 km*		*d < 80 km*		*d < 100 km*	
Variables																
Unemployment rate	0.135***	(0.019)	0.130***	(0.019)	0.126***	(0.019)	0.160***	(0.019)	0.141***	(0.019)	0.131***	(0.019)	0.131***	(0.019)	0.132***	(0.019)
Vacancy rate	0.346***	(0.043)	0.337***	(0.043)	0.347***	(0.043)	0.326***	(0.042)	0.318***	(0.043)	0.330***	(0.043)	0.332***	(0.043)	0.332***	(0.043)
W*Unemployment rate	0.281***	(0.038)	0.473***	(0.057)	0.476***	(0.055)	0.069***	(0.025)	0.227***	(0.034)	0.375***	(0.044)	0.475***	(0.054)	0.564***	(0.070)
W*Vacancy rate	-0.482***	(0.082)	-0.766***	(0.120)	-0.786***	(0.116)	-0.189***	(0.055)	-0.323***	(0.075)	-0.561***	(0.093)	-0.745***	(0.115)	-0.900***	(0.144)
Constant	1.031***	(0.007)	1.030***	(0.007)	1.029***	(0.007)	1.033***	(0.007)	1.031***	(0.007)	1.029***	(0.007)	1.029***	(0.007)	1.030***	(0.007)
Observations	12,320		12,320		12,320		12,320		12,320		12,320		12,320		12,320	
Number of groups	1,120		1,120		1,120		1,120		1,120		1,120		1,120		1,120	
Year FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
Area FE	Yes		Yes		Yes		Yes		Yes		Yes		Yes		Yes	
F-Test unemployment rate instrument	35.31		35.31		35.31		35.31		35.31		35.31		35.31		35.31	
F-Test vacancy rate instrument	29.10		29.10		29.10		29.10		29.10		29.10		29.10		29.10	
Wald-Test pval	0.000		0.000		0.000		0.003		0.000		0.000		0.000		0.000	
Bayesian comparison probabilities																
Model probability	0.000		0.000		0.000		0.000		0.000		0.000		0.000		0.000	

Notes: Standard errors in parentheses. Results of equation (5) by SLX. The outcome variable used is change on the natural logarithm of the hiring rate. The sample includes all Colombian contiguous municipalities with available data from 2010 to 2019. The year 2009 is lost after lagging the series. The spatial weight matrix Queen 1st was generated using a queen contiguity method, first-order neighbors. The spatial weight matrix Queen 2nd was generated using a queen second-order neighbors. The spatial weight matrix Queen 3rd was generated using a queen second-order neighbors with 1 for first-order neighbors and 0.5 for second-order neighbors. $d < 20$, $d < 40$, $d < 60$, $d < 80$, and $d < 100$ km refer to contiguity matrices for neighbors within 20, 40, 60, 80 and 100 km, respectively. Municipal and year fixed effects regressors are not shown. See Appendix A.1 for description and summary statistics for all the variables.
**** p<0.01, ** p<0.05, * p<0.10*

Table 3: Direct and indirect effects estimates based on the coefficient estimates of the spatial Durbin error model (SDEM) reported in the fourth column in Table 1

	Direct effect		**Indirect effect**		**Total effect**	
Variables						
Unemployment rate	0.157***	(0.019)	0.073***	(0.025)	0.23***	(0.007)
Vacancy rate	0.336***	(0.042)	-0.190***	(0.056)	0.146***	(0.055)

Standard errors in parentheses.
*** p <0.01, ** p <0.05, * p <0.10

6 Conclusions

This paper explores spatial interdependence's role in local labor markets in a middle-income country. Spillover effects in this context are expected because workers can search for jobs in neighboring labor markets, and workers from nearby cities can fill local vacancies. We quantified unemployment and vacancies' effects on hires across and within municipalities. We built a panel data set of 1120 municipalities in Colombia from different sources, allowing for estimating these spillover effects in the matching function framework, a standard theoretical tool that describes the generation of matches in the labor market.

We found robust evidence to support the existence of spatial spillovers in labor markets, which indicates that local hires are affected by unemployment and vacancies in neighboring municipalities. We find positive spatial spillover of the unemployment rate on the generation of new matches; workers search for jobs in neighboring labor markets and fill open positions there. In contrast, we find that the neighboring market's vacancy rate negatively affects the local hiring rate; local vacancies will compete to be filled with vacancies in neighboring markets as the probability of a local worker being matched with a foreign vacancy increases with the number of open positions in the neighboring markets. Our findings imply that, given the growing integration across cities in developing and middle-income countries, a spatial perspective at the sub-national level should be considered in analyzing their labor markets. Our results are robust to robustness checks, including changes in our instrumental variables, measurement of some key variables, model specification, and estimation method.

According to our findings, shocks that affect local labor markets have larger implications than their local effects; they will indirectly contribute to changes in the formation of new firm-employee matches in other markets. In this sense, adverse shocks in larger cities, for instance, will spread their consequences across neighboring cities. This potential scenario supports differential policies conditioned upon the size and development of labor markets. Policies on accessibility to ease people moving to jobs and improving connections between workers and jobs through "the circulation of information flows between firm and workers" will contribute to the consolidation of small markets (Gobillon & Selod, 2019). For instance, implementing targeted active labor market policies could address regional disparities (Haller & Heuermann, 2016). These policies could take advantage of the trend of job seekers seeking employment beyond their hometowns. Local public employment centers in small cities might connect workers with job opportunities in nearby regions. Concurrently, they could inform firms about well-suited local candidates for their vacant positions. There is still room in the literature for further research to characterize the efficiency of labor markets across cities and the role of structural factors determining the heterogeneity in markets' ability to generate quality employers-jobs matches.

Disclosure statement

There are no relevant financial or non-financial competing interests to report. The opinions contained in this document are the sole responsibility of the authors and do not commit Banco de la República or its Board of Directors.

Data availability

The data from the Colombian National Department of Statistics (DANE) is free available and online. The data from PILA (its acronym in Spanish) comes from the Ministry of Health and its access is restricted.

Funding statement

No funding was received.

References

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*(1), 77–104. https://doi.org/10.1016/0166-0462(95)02111-6

Autor, D. H., & Duggan, M. G. (2003). The Rise in the Disability Rolls and the Decline in Unemployment*. *The Quarterly Journal of Economics*, *118*(1), 157–206. https://doi.org/10.1162/00335530360535171

Bartik, T. (1991). Who Benefits from State and Local Economic Development Policies? *Upjohn Press*. https://doi.org/10.17848/9780585223940

Blanchard, O. J., & Diamond, P. A. (1989). *The Aggregate Matching Function* (Working Paper 3175). National Bureau of Economic Research. https://doi.org/10.3386/w3175

Blanchard, O. J., & Katz, L. F. (1992). Regional Evolutions. *Brookings Papers on Economic Activity*, *n 1*, 1–75.

Bound, J., & Holzer, H. J. (2000). Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s. *Journal of Labor Economics*, *18*(1), 20–54. https://doi.org/10.1086/209949

Burda, M. C., & Profit, S. (1996). Matching across space: Evidence on mobility in the Czech Republic. *Labour Economics*, *3*(3), 255–278. https://doi.org/10.1016/S0927-5371(96)00013-9

Burgess, S. (1993). A Model of Competition Between Unemployed and Employed Job Searchers: An Application to the Unemployment Outflow Rate in Britain. *The Economic Journal*, *103*(420), 1190–1204. https://doi.org/10.2307/2234245

Burgess, S., & Profit, S. (2001). Externalities in the matching of workers and firms in Britain. *Labour Economics*, *8*(3), 313–333. https://doi.org/10.1016/S0927-5371(01)00036-7

Costa-da-Silva, D. F., Elhorst, J. P., & Silveira-Neto, R. da M. (2017). Urban and rural population growth in a spatial panel of municipalities. *Regional Studies*, *51*(6), 894–908. https://doi.org/10.1080/00343404.2016.1144922

Dávalos, E., & Morales, L. F. (2023). Diffusion of crime control benefits: Forced eradication and coca crops in Colombia. *Spatial Economic Analysis*, *18*(3), 292–317. https://doi.org/10.1080/17421772.2022.2151642

Davis, S. J., & Haltiwanger, J. (2014). *Labor Market Fluidity and Economic Performance* (Working Paper 20479). National Bureau of Economic Research. https://doi.org/10.3386/w20479

Duranton, G. (2015). A Proposal to Delineate Metropolitan Areas in Colombia. *Desarrollo y Sociedad*, *75*, 223–264. https://doi.org/10.13043/dys.75.6

Elhorst, J. P. (2014). *Spatial Econometrics* (Vol. 479). Springer. https://link.springer.com/book/10.1007/978-3-642-40340-8

Ferreira-Neto, A. B. (2023). Do public libraries impact local labour markets? Evidence from Appalachia. *Spatial Economic Analysis*, *18*(2), 216–238. https://doi.org/10.1080/17421772.2022.2123552

Gobillon, L., & Selod, H. (2019). Spatial Mismatch, Poverty, and Vulnerable Populations. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (pp. 1–16). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36203-3_7-1

Halleck Vega, S., & Elhorst, J. P. (2015). The SLX Model. *Journal of Regional Science*, *55*(3), 339–363. https://doi.org/10.1111/jors.12188

Halleck Vega, S., & Elhorst, J. P. (2016). A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors. *Regional Science and Urban Economics*, *60*, 85–95. https://doi.org/10.1016/j.regsciurbeco.2016.07.002

Haller, P., & Heuermann, D. F. (2016). Job search and hiring in local labor markets: Spillovers in regional matching functions. *Regional Science and Urban Economics*, *60*, 125–138. https://doi.org/10.1016/j.regsciurbeco.2016.06.007

Higashi, Y. (2018). Spatial spillovers in job matching: Evidence from the Japanese local labor markets. *Journal of the Japanese and International Economies*, *50*, 1–15. https://doi.org/10.1016/j.jjie.2018.06.001

Ilmakunnas, P., & Pesola, H. (2003). Regional Labour Market Matching Functions and Efficiency Analysis. *LABOUR*, *17*(3), 413–437. https://doi.org/10.1111/1467-9914.00247

Kosfeld, R. (2007). Regional Spillovers and Spatial Heterogeneity in Matching Workers and Employers in Germany. *Jahrbücher für Nationalökonomie und Statistik*, *227*(3), 236–253. https://doi.org/10.1515/jbnst-2007-0303

Layard, R., Nickell, S. J., & Jackman, R. (1991). *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford University Press.

LeSage, J. P., & Pace, R. K. (2010). Spatial Econometric Models. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (pp. 355–376). Springer. https://doi.org/10.1007/978-3-642-03647-7_18

Morales, L. F., & Lobo, J. (2020). Estimating vacancies from firms' hiring behavior: The case of a developing economy. *Journal of Economic and Social Measurement*, *45*(2), 139–170. https://doi.org/10.3233/JEM-210473

Morales, L. F., & Medina, D. (2019). Labour market fluidity and employment outcomes in Colombia: Evidence from employer-employee linked data. *CEPAL Review*. https://hdl.handle.net/11362/44717

Morales, L. F., Ospino, C., & Amaral, N. (2021). Online Vacancies and its Role in Labor Market Performance. *Borradores de Economía; No. 1174*. https://repositorio.banrep.gov.co/handle/20.500.12134/10143

Mumford, K., & Smith, P. N. (1999). The Hiring Function Reconsidered: On Closing the Circle. *Oxford Bulletin of Economics and Statistics*, *61*(3), 343–364. https://doi.org/10.1111/1468-0084.00133

Patacchini, E., & Zenou, Y. (2007). *Ethnicity and Spatial Externalities in Crime* (SSRN Scholarly Paper 1133775). CEPR Discussion Paper No. DP6130. https://papers.ssrn.com/abstract=1133775

Petrongolo, B., & Pissarides, C. A. (2001). Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, *39*(2), 390–431. https://doi.org/10.1257/jel.39.2.390

Pissarides, C. A. (2001). Employment protection. *Labour Economics*, *8*(2), 131–159. https://doi.org/10.1016/S0927-5371(01)00032-X

Rios, V. (2017). What drives unemployment disparities in European regions? A dynamic spatial panel approach. *Regional Studies*, *51*(11), 1599–1611. https://doi.org/10.1080/00343404.2016.1216094

Zeilstra, A. S., & Elhorst, J. P. (2014). Integrated Analysis of Regional and National Unemployment Differentials in the European Union. *Regional Studies*, *48*(10), 1739–1755. https://doi.org/10.1080/00343404.2012.708404

Appendix A.

Appendix A.1. Summary statistics

Table A.1: Summary statistics

	Obs	**Mean**	**Std. Dev.**
Variables			
Hiring rate	12,320	5.640	88.730
Unemployment rate	12,320	10.952	9.257
Vacancy rate	12,320	3.413	4.994

Notes: The unemployment rate is the ratio of the unemployed to the labor force. The hiring rate is the ratio of total hires to employment. The vacancy rate is the ratio of vacancies to employment plus vacancies. The estimation sample is an annual panel of 1,120 Colombian municipalities for the 2010-2019 period.

Appendix A.2. Global Moran's I

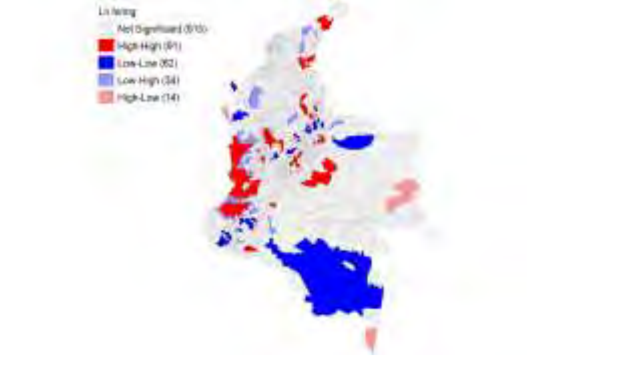
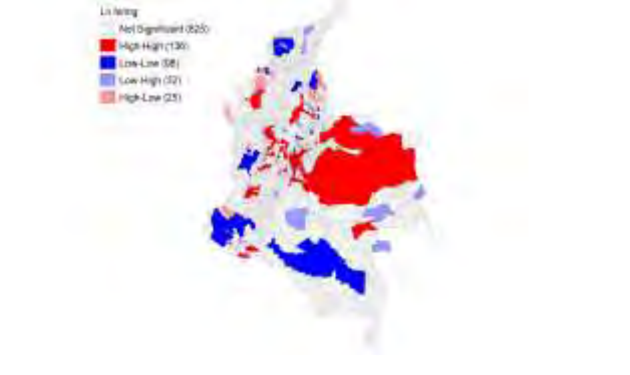
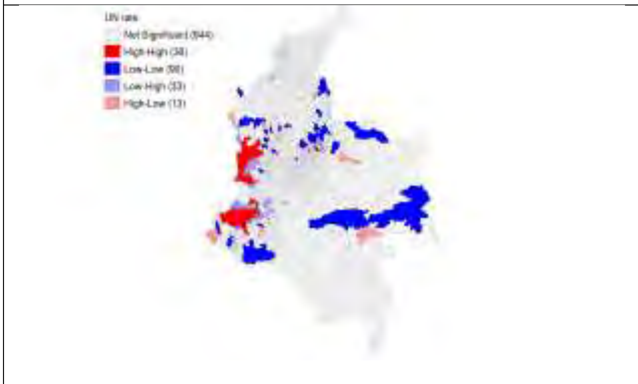
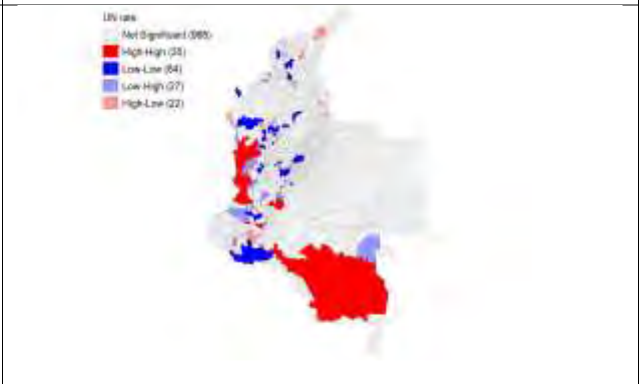
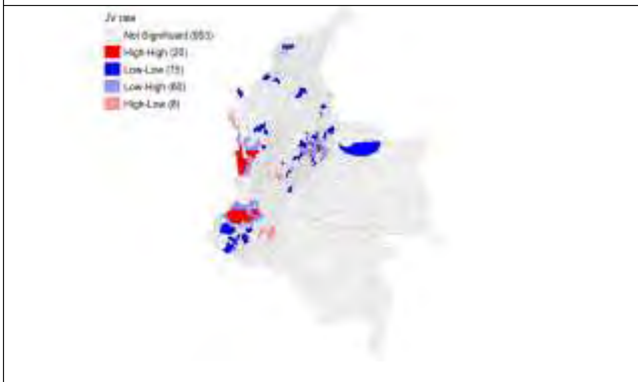
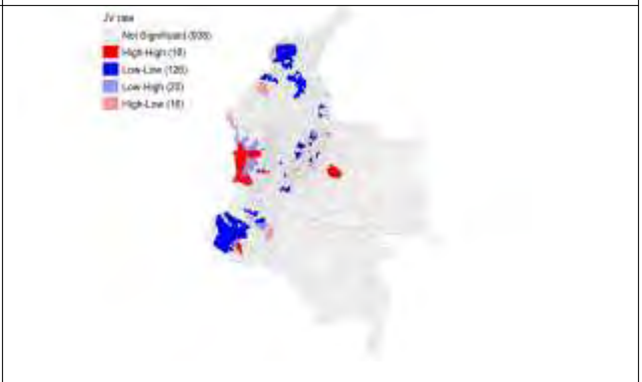
Table A.2: Global Moran's I value of log of hiring rate, unemployment rate, and job vacancy rate from 2009 to 2019

	Municipal level (1,120 municipalities)		
Year	*Log of hiring rate*	*Unemployment rate*	*Job vacancy rate*
2009	0.181***	0.319***	0.103***
2010	0.178***	0.345***	0.139***
2011	0.162***	0.305***	0.098***
2012	0.181***	0.357***	0.176***
2013	0.197***	0.305***	0.172***
2014	0.195***	0.330***	0.272***
2015	0.251***	0.260***	0.197***
2016	0.275***	0.270***	0.063***
2017	0.227***	0.203***	0.030**
2018	0.254***	0.228***	0.090***
2019	0.275***	0.259***	0.135***

***$p < 0.01$, **$p < 0.05$, *$p < 0.10$

Appendix A.3. Local Moran's I index of spatial association cluster map

Table A.3: Local indicator of spatial association cluster map of log of hiring rate, unemployment (UN) rate and job vacancy (JV) rate for 2009 and 2019

 2009	 2019
 2009	 2019
 2009	 2019

Notes: Significance level is 0.05.

Appendix A.4. Lagrange Multiplier test

Table A.4: Specific tests for spatial dependence, Lagrange Multiplier (LM)

Specification	**Pooled OLS**		**Spatial fixed effects**		**Time-period fixed effects**		**Spatial and time-period fixed effects**	
	Test statistic	*p value*	*Test statistic*	*p value*	*Test statistic*	*p value*	*Test statistic*	*p value*
LM spatial lag	2219.00	0.00	1887.10	0.00	1472.90	0.00	323.10	0.00
LM spatial error	2353.70	0.00	1960.30	0.00	1558.20	0.00	326.00	0.00
Robust LM spatial lag	16.40	0.00	5.50	0.02	12.20	0.00	4.50	0.04
Robust LM spatial error	151.10	0.00	78.80	0.00	97.60	0.00	7.40	0.01

Notes: LM tests are estimated using Matlab codes following Elhorst (2014a).

Appendix A.5. Non-spatial regression of the matching function

Table A.5: Non-spatial two-stage regression

Change on the municipal log of hiring rate		
	Coefficient	*Std. err.*
Variables		
Unemployment rate	0.166***	0.019
Vacancy rate	0.300***	0.041
Constant	-4.434***	0.145
Observations	12,320	
Number of groups	1,120	
Year FE	Yes	
Municipality FE	Yes	

**** p<0.01, ** p<0.05, * p<0.10*

Appendix A.6. Spatial non-instrumental regression of the matching function

Table A.6: Standard spatial regression for Spatial Durbin error model

	Change on municipal log of hiring rate							
	Queen 1st	*Queen 2nd*	*Queen 3rd*	*d < 20 km*	*d < 40 km*	*d < 60 km*	*d < 80 km*	*d < 100 km*
Variables								
Unemployment rate	0.013*** (0.001)	0.012*** (0.001)	0.012*** (0.001)	0.014*** (0.001)	0.012*** (0.001)	0.012*** (0.001)	0.012*** (0.001)	0.012*** (0.001)
Vacancy rate	0.044*** (0.002)	0.044*** (0.002)	0.044*** (0.002)	0.045*** (0.002)	0.044*** (0.002)	0.044*** (0.002)	0.045*** (0.002)	0.044*** (0.002)
W*Unemployment rate	0.013*** (0.002)	0.022*** (0.004)	0.023*** (0.004)	0.001 (0.002)	0.013*** (0.003)	0.018*** (0.004)	0.027*** (0.005)	0.030*** (0.005)
W*Vacancy rate	-0.016*** (0.005)	-0.013 (0.009)	-0.021** (0.008)	-0.011*** (0.004)	-0.015*** (0.005)	-0.023*** (0.008)	-0.036*** (0.011)	-0.030** (0.012)
W*Error	0.218*** (0.013)	0.437*** (0.020)	0.417*** (0.019)	0.137*** (0.012)	0.261*** (0.016)	0.380*** (0.019)	0.452*** (0.021)	0.491*** (0.022)
Constant	1.007*** (0.006)	0.996*** (0.006)	0.996*** (0.006)	1.015*** (0.006)	1.007*** (0.006)	1.001*** (0.006)	0.997*** (0.006)	0.997*** (0.006)
Observations	13,440	13,440	13,440	13,440	13,440	13,440	13,440	13,440
Number of groups	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wald-Test pval	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000

Notes: Standard errors in parentheses. Results of equation (7) by SDEM. The outcome variable used is change on the natural logarithm of the hiring rate. The sample includes all Colombian contiguous municipalities with available data from 2010 to 2019. The year 2009 is lost after lagging the series. The spatial weight matrix Queen 1st was generated using a queen contiguity method, first-order neighbors. The spatial weight matrix Queen 2nd was generated using a queen second-order neighbors. The spatial weight matrix Queen 3rd was generated using a queen second-order neighbors with 1 for first-order neighbors and 0.5 for second-order neighbors. $d < 20$, $d < 40$, $d < 60$, $d < 80$, and $d < 100$ km refer to contiguity matrices for neighbors within 20, 40, 60, 80 and 100 km, respectively. Municipal and year fixed effects regressors are not shown. See Appendix A.1 for description and summary statistics for all the variables.
*** p<0.01, ** p<0.05, * p<0.10

Appendix A.7. Robustness check, alternative instrumental variables 1. Spatial Durbin error model

Table A.7: Spatial Durbin error model robustness check. $K > 150$ km

	Change on municipal log of hiring rate							
	Queen 1st	*Queen 2nd*	*Queen 3rd*	*d < 20 km*	*d < 40 km*	*d < 60 km*	*d < 80 km*	*d < 100 km*
Variables								
Unemployment rate	0.079*** (0.028)	0.059** (0.027)	0.061** (0.028)	0.098*** (0.027)	0.070** (0.027)	0.065** (0.027)	0.060** (0.027)	0.066** (0.027)
Vacancy rate	0.335*** (0.061)	0.351*** (0.061)	0.350*** (0.061)	0.307*** (0.061)	0.339*** (0.061)	0.344*** (0.061)	0.344*** (0.061)	0.334*** (0.061)
W*Unemployment rate	0.175*** (0.057)	0.434*** (0.101)	0.377*** (0.097)	0.024 (0.030)	0.271*** (0.044)	0.356*** (0.058)	0.466*** (0.076)	0.527*** (0.106)
W*Vacancy rate	-0.310** (0.125)	-0.701*** (0.221)	-0.617*** (0.211)	-0.056 (0.067)	-0.458*** (0.102)	-0.597*** (0.133)	-0.759*** (0.165)	-0.881*** (0.226)
W*Error	0.211*** (0.014)	0.440*** (0.020)	0.415*** (0.020)	0.108*** (0.013)	0.255*** (0.017)	0.373*** (0.020)	0.455*** (0.022)	0.506*** (0.023)
Constant	1.029*** (0.007)	1.017*** (0.007)	1.018*** (0.007)	1.039*** (0.007)	1.027*** (0.007)	1.021*** (0.007)	1.017*** (0.007)	1.017*** (0.007)
Observations	12,320	12,320	12,320	12,320	12,320	12,320	12,320	12,320
Number of groups	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Test unemployment rate instrument	43.83	43.83	43.83	43.83	43.83	43.83	43.83	43.83
F-Test vacancy rate instrument	35.26	35.26	35.26	35.26	35.26	35.26	35.26	35.26
Wald-Test pval	0.003	0.000	0.000	0.700	0.000	0.000	0.000	0.000

Notes: Standard errors in parentheses. Results of equation (7) by SDEM. The outcome variable used is change on the natural logarithm of the hiring rate. The sample includes all Colombian contiguous municipalit with available data from 2010 to 2019. The year 2009 is lost after lagging the series. The spatial weight matrix Queen 1st was generated using a queen contiguity method, first-order neighbors. The spatial weight matrix Queen 2nd was generated using a queen second-order neighbors. The spatial weight matrix Queen 3rd was generated using a queen second-order neighbors with 1 for first-order neighbors and 0.5 for seconc order neighbors. $d < 20$, $d < 40$, $d < 60$, $d < 80$, and $d < 100$ km refer to contiguity matrices for neighbors within 20, 40, 60, 80 and 100 km, respectively. Municipal and year fixed effects regressors are not sho See Appendix A.1 for description and summary statistics for all the variables.
*** p<0.01, ** p<0.05, * p<0.10

Appendix A.8. Robustness check, alternative instrumental variables 2. Spatial Durbin error model

Table A.8: Spatial Durbin error model robustness check. No distance restriction

	Change on municipal log of hiring rate							
	Queen 1st	*Queen 2nd*	*Queen 3rd*	*d < 20 km*	*d < 40 km*	*d < 60 km*	*d < 80 km*	*d < 100 km*
Variables								
Unemployment rate	0.100*** (0.013)	0.089*** (0.013)	0.087*** (0.013)	0.116*** (0.013)	0.104*** (0.013)	0.092*** (0.013)	0.090*** (0.013)	0.092*** (0.013)
Vacancy rate	0.422*** (0.034)	0.423*** (0.034)	0.430*** (0.034)	0.418*** (0.034)	0.395*** (0.034)	0.419*** (0.034)	0.421*** (0.034)	0.419*** (0.034)
W*Unemployment rate	0.154*** (0.027)	0.310*** (0.048)	0.305*** (0.046)	0.059*** (0.019)	0.132*** (0.025)	0.237*** (0.034)	0.307*** (0.045)	0.351*** (0.057)
W*Vacancy rate	-0.270*** (0.069)	-0.526*** (0.119)	-0.531*** (0.114)	-0.170*** (0.048)	-0.164** (0.065)	-0.376*** (0.086)	-0.514*** (0.113)	-0.610*** (0.142)
W*Error	0.184*** (0.014)	0.401*** (0.021)	0.377*** (0.020)	0.103*** (0.013)	0.228*** (0.017)	0.342*** (0.021)	0.420*** (0.023)	0.468*** (0.024)
Constant	1.016*** (0.007)	1.005*** (0.007)	1.006*** (0.007)	1.023*** (0.007)	1.015*** (0.007)	1.010*** (0.007)	1.006*** (0.007)	1.005*** (0.007)
Observations	12,320	12,320	12,320	12,320	12,320	12,320	12,320	12,320
Number of groups	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Test unemployment rate instrument	49.19	49.19	49.19	49.19	49.19	49.19	49.19	49.19
F-Test vacancy rate instrument	29.65	29.65	29.65	29.65	29.65	29.65	29.65	29.65
Wald-Test pval	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000

Notes: Standard errors in parentheses. Results of equation (7) by SDEM. The outcome variable used is change on the natural logarithm of the hiring rate. The sample includes all Colombian contiguous municipaliti with available data from 2010 to 2019. The year 2009 is lost after lagging the series. The spatial weight matrix Queen 1st was generated using a queen contiguity method, first-order neighbors. The spatial weight matrix Queen 2nd was generated using a queen second-order neighbors. The spatial weight matrix Queen 3rd was generated using a queen second-order neighbors with 1 for first-order neighbors and 0.5 for second- order neighbors. $d < 20$, $d < 40$, $d < 60$, $d < 80$, and $d < 100$ km refer to contiguity matrices for neighbors within 20, 40, 60, 80 and 100 km, respectively. Municipal and year fixed effects regressors are not sho
See Appendix A.1 for description and summary statistics for all the variables.
*** p<0.01, ** p<0.05, * p<0.10

Appendix A.9. Robustness check, alternative measures. Spatial Durbin error model

Table A.9: Spatial Durbin error model robustness check. Alternative measures

	Change on municipal log of hiring rate							
	Queen 1st	*Queen 2nd*	*Queen 3rd*	*d < 20 km*	*d < 40 km*	*d < 60 km*	*d < 80 km*	*d < 100 km*
Variables								
Unemployment rate	0.088*** (0.008)	0.078*** (0.008)	0.077*** (0.008)	0.106*** (0.008)	0.084*** (0.008)	0.079*** (0.008)	0.078*** (0.008)	0.080*** (0.008)
Vacancy rate	0.191*** (0.020)	0.192*** (0.019)	0.193*** (0.019)	0.179*** (0.019)	0.184*** (0.019)	0.187*** (0.019)	0.190*** (0.019)	0.188*** (0.019)
W*Unemployment rate	0.109*** (0.017)	0.241*** (0.027)	0.231*** (0.026)	0.080*** (0.013)	0.156*** (0.015)	0.221*** (0.020)	0.273*** (0.025)	0.248*** (0.030)
W*Vacancy rate	-0.070* (0.037)	-0.170*** (0.056)	-0.160*** (0.054)	-0.066** (0.030)	-0.133*** (0.039)	-0.191*** (0.046)	-0.245*** (0.052)	-0.179*** (0.059)
W*Error	0.171*** (0.023)	0.420*** (0.034)	0.383*** (0.033)	0.110*** (0.020)	0.159*** (0.027)	0.257*** (0.035)	0.328*** (0.038)	0.428*** (0.039)
Constant	0.869*** (0.009)	0.852*** (0.009)	0.854*** (0.009)	0.874*** (0.009)	0.864*** (0.009)	0.858*** (0.009)	0.857*** (0.009)	0.858*** (0.009)
Observations	5,600	5,600	5,600	5,600	5,600	5,600	5,600	5,600
Number of groups	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
F-Test unemployment rate instrument	32.74	32.74	32.74	32.74	32.74	32.74	32.74	32.74
F-Test vacancy rate instrument	26.52	26.52	26.52	26.52	26.52	26.52	26.52	26.52
Wald-Test pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Standard errors in parentheses. Results of equation (7) by SDEM. The outcome variable used is change on the natural logarithm of the hiring rate. The sample includes all Colombian contiguous municipalit with available data from 2010 to 2019. The year 2009 is lost after lagging the series. The spatial weight matrix Queen 1st was generated using a queen contiguity method, first-order neighbors. The spatial weight matrix Queen 2nd was generated using a queen second-order neighbors. The spatial weight matrix Queen 3rd was generated using a queen second-order neighbors with 1 for first-order neighbors and 0.5 for seconc order neighbors. $d < 20$, $d < 40$, $d < 60$, $d < 80$, and $d < 100$ km refer to contiguity matrices for neighbors within 20, 40, 60, 80 and 100 km, respectively. Municipal and year fixed effects regressors are not sho
See Appendix A.1 for description and summary statistics for all the variables.
*** p<0.01, ** p<0.05, * p<0.10

Appendix A.10. Robustness check, SLX model with optimal distance decay parameter

Table A.10: SLX robustness check

	Change on municipal log of hiring rate			
	No restriction		*50 kms >*	
Variables				
Unemployment rate	0.101***	(0.013)	0.147***	(0.018)
Vacancy rate	0.425***	(0.033)	0.331***	(0.041)
W*Unemployment rate	1.491***	(0.201)	1.571***	(0.290)
W*Vacancy rate	-2.888***	(0.402)	-2.895***	(0.551)
Observations	12,320		12,320	
Number of groups	1,120		1,120	
Year FE	Yes		Yes	
Area FE	Yes		Yes	

Standard erros in parentheses

**** p<0.01, ** p<0.05, * p<0.10*

[1] Theoretical unemployment equilibrium models do not distinguish between these two variables (Pissarides, 2001). The concept of hires is often used to refer to new matches; the outflow from unemployment and hires does not necessarily coincide because new hires might come from previously out-of-the-labor-force workers.

[2] Legally, these are non-municipalized locations that correspond to territories under indigenous jurisdictions. However, DANE assigns them a municipal code to include them in its statistical system.

[3] Previous literature has paid attention to the limitations of MA definitions in Colombia (Duranton, 2015).

[4] All appendices are available online.

[5] For estimating the statistics in all these tests, we use a queen spatial contiguity matrix of order 1.

[6] The use of non-spatial matrices is a recent practice in the literature; for instance, infrastructure-based weights are an alternative approach to modeling the relationships across local labor markets (Haller & Heuermann, 2016). Given data limitations, we cannot explore this approach.

[7] Even though the selected matrix identifies as neighboring markets those municipalities within a 20 km radius, we find that the elasticities, estimated with the two-stage SDEM models, increase with the buffer radius allowed for including other labor markets in the network of neighbors, as it can be seen in Appendix A.10. This result can be expected because spillovers identified by the SDEM are local, therefore, by expanding the definition of the relevant network, the magnitude of the spillover effects increases.