



# Housing Informality and Labor Informality in Space: In Search of the Missing Links

Gustavo A. García<sup>1</sup> · Erika Raquel Badillo<sup>2</sup> · Juan Manuel Aristizábal<sup>3</sup>

Received: 27 June 2023 / Accepted: 28 January 2024 / Published online: 26 February 2024  
© The Author(s) 2024

## Abstract

This paper analyzes the relationship between informal housing and labor informality at the intraurban level, considering these two phenomena' simultaneity and spatial dimensions. Our analysis focuses on the context of a city in a developing country, Medellín (Colombia), characterized by significant housing precariousness and low employment quality, where space seems to play an essential role in understanding this relationship. Using data from 176 analytical regions in Medellín for 2017, we estimate a series of spatial simultaneous equation models that consider the potential cross-equation correlations in the error terms. The results show that these two types of urban informality are highly persistent in space, with noticeable spatial clusters observed in the peripheral and marginalized areas of the city. Additionally, the estimated econometric models reveal that precarious working conditions are key to explaining the spatial choice of housing and its characteristics, and vice versa. These findings emphasize the need for place-based policies that specifically target disadvantaged areas and help improve residents' working and housing conditions to address urban informality.

**Keywords** Informal housing · Slums · Labor informality · Spatial dependence · Spatial simultaneous equations model

**JEL Classification** O17 · O18 · R14 · C33 · J46

---

✉ Gustavo A. García  
ggarci24@eafit.edu.co

<sup>1</sup> School of Finance, Economics and Government, OMEGA Research Group, Universidad EAFIT, Medellín, Colombia

<sup>2</sup> Faculty of Economics, GINVECO Research Group, Universidad Autónoma Latinoamericana (Unaula), Medellín, Colombia

<sup>3</sup> Faculty of Accounting, Economic and Administrative Sciences, Universidad de Manizales, Manizales, Colombia

## Introduction

Nearly one out of eight people in the world live in slums (i.e., informal settlements with precarious dwellings, a lack of essential services, and insecure tenure),<sup>1</sup> which translates to approximately one billion individuals living in slum conditions (UN-Habitat, 2010, 2013, 2016a). Most people living in precarious conditions are found in developing countries, where roughly 863 million individuals, or one-third of their population, reside in informal settlements. Latin America is the world's second most urbanized region, with 20% to 30% of its urban population residing in informal and impoverished settlements and 60% of the region's poor individuals living in slums (Vargas et al., 2017). Since it is projected that poverty will continue to urbanize throughout the region, analyzing the determinants of housing informality is crucial for understanding the growth of this type of settlement and proposing government measures to improve the quality of life in informal settlements.

Several factors have been identified in the literature to explain the formation and proliferation of informal settlements. Notably, Roy et al. (2014) and Mahabir et al. (2016) identified factors such as local topography, street patterns, population dynamics, the politics of slums, and the informal economy. While there is ample evidence of the effects of local conditions and policies on the growth of slums (Alves, 2018; Cavalcanti et al., 2019; Marx et al., 2013), the available evidence of the effects of demographics and economics is scarce, particularly concerning how informal economy<sup>2</sup> relates to the formation and functioning of informal settlements.

According to International Labor Organization (ILO) (2018), Bonnet et al. (2019), and Deléchat and Medina (2021), 61% of all workers worldwide (around 2 billion workers) are informally employed, with the prevalence of labor informality being significantly higher in developing countries (90%), than in developed countries (18%). In Latin America, 60% of employees are engaged in informal activities, and 62% are extremely poor and vulnerable. The countries in this region with the highest rates of labor informality are Bolivia (79%), Paraguay (70%), Peru (67%), and Colombia (60%) (OECD/ILO, 2019). Labor informality has been reported to have considerable implications for development outcomes, as it is an obstacle to sustainable development due to its close relationship with economic growth, poverty, and inequality. Informal workers lack formal and social protection, are less educated, and work in small firms with low productivity (Del Carpio & Patrick, 2021). As a result, they are more susceptible to poverty and face significant barriers (or exclusion) to accessing the educational and financial systems (Deléchat & Medina, 2021; Roy et al., 2014).

<sup>1</sup> In this paper, the terms “slums”, “informal settlements”, and “housing informality” are used interchangeably, as is common in the literature (Cavalcanti et al., 2019; UN-Habitat, 2012). However, as stated by UN-Habitat (2003), Mahabir et al. (2016), and Azunre et al. (2021), an informal settlement (or housing informality) is one type of slum, which depends on the dwelling's degree of inadequacy (which we discuss in more detail below).

<sup>2</sup> The term “informal economy” refers to the operation of unregistered enterprises (i.e., those not registered at the corresponding chamber of commerce), as well as the lack of labor protections for employees (e.g., lack of an employment contract or health and pension plans) working in either informal or formal firms (Basu & Chau, 2015; Deléchat and Medina, 2021; ILO, 1993).

Concerning the relationship between slums and labor activities, recent research has shown that informal activities are a primary economic driver for the emergence of slum conditions (Cavalcanti, 2017, 2019). For instance, an important characteristic of informal workers is that they typically carry out their activities at home, blurring the boundaries between domestic and labor spaces, thus leading to precarious living conditions (Cavalcanti, 2017; Suárez et al., 2016). Additionally, informal workers often live in peripheral neighborhoods with fewer regulations, lower housing prices, and reduced costs of goods, services, and other commodities (Posada, 2018). Consequently, informal construction proliferates in informal neighborhoods, which results in a high concentration of low-income, informal workers in these informal settlements (Cavalcanti, 2019; Cavalcanti et al., 2019; UN-Habitat, 2011). This pattern of urban informality prevails in developing countries, particularly Latin American cities (Abramo, 2009; Bouillon, 2012; Maloney, 2004; UN-Habitat, 2012).

The spatial relationship between informal housing and informal employment aligns with Rallet's (2000) concept of organized proximity. In this regard, the relationship of proximity between these two types of informality may be constituted by institutionalized relationships, informal and tacit networks, and hierarchies, or even by the power of force and violence. Posada and Moreno-Monroy (2017) suggest common structural features in the relationship between informal housing and informal employment. Therefore, to understand the factors that influence the formation of slums and the nature of precarious work, it is essential to analyze the relationship between informal work activities and informal housing. Accordingly, this paper aims to provide new evidence on the relationship between informal housing and labor informality and the role of space in this relationship.

For our analysis, we use data from Medellín, the second-largest city in Colombia, which makes for an interesting case study because of its high levels of informal housing and labor informality and the marked spatial segregation of these two phenomena within the urban context. Our empirical analysis is carried out at the intraurban level, considering 176 analytical regions as our spatial units of analysis for 2017. We perform spatial analysis to examine the simultaneity of housing informality and labor informality in space. In addition, we estimate spatial simultaneous equations models to evaluate the simultaneous effects of housing informality and labor informality, considering the spatial dimension of these two phenomena. Such models are estimated using General Spatial Three-Stages Least Squares (GS3SLS), which is a system instrumental variable procedure that includes a Generalized Method of Moments (GMM) estimation when there is spatial interrelation in the error terms (Kelejian & Prucha, 2004). Also, this procedure allows us to address the endogeneity resulting from the simultaneous occurrence of the two types of informality and the inclusion of spatial lags in the models.

This paper contributes to the existing literature in three ways. First, unlike previous studies that have examined housing informality and labor informality separately (with few considering the spatial dimension of these two phenomena), we study them simultaneously. Second, by conducting a spatial analysis at the intraurban level, we gain a more detailed understanding of the spatial dynamics of informal housing and its relationship with labor informality. Third, our empirical approach considers the spatial spillover effects inherent in both types of informality, allowing

us to explore the role of precarious labor conditions in explaining precarious housing conditions, as well as the role of spatial spillover effects. This study is the first to rigorously analyze the spatial relationship between informal housing and labor informality. In addition, it provides empirical evidence and conclusions specific to the case of Medellín and offers conceptual and methodological insights into urban informality relevant to other cities in the Global South.

The rest of this paper is organized as follows. Section "[Related literature](#)" discussed the relevant literature on informal housing and its relationship with the informal economy. Section "[Data and descriptive statistics](#)" provides an overview of the study area and the data used in this study and presents some descriptive statistics. Sections "[Econometric model](#)" and "[Results](#)" explain the proposed empirical strategy and report the key findings. Finally, Section "[Conclusions](#)" summarizes the main findings and concludes this research.

## Related Literature

In this section, we present a literature review and discuss the key factors that influence the emergence and growth of informal housing, focusing on the literature about the informal economy.

A considerable body of work has investigated the factors contributing to slum development. Notably, the studies by Roy et al. (2014) and Mahabir et al. (2016) have emerged as notable contributions to this literature. Roy et al. (2014) performed a comprehensive review of various studies on slums and identified seven critical factors associated with the formation of informal settlements: population dynamics, economic growth, housing market dynamics, local topography, street pattern, politics of slums, and informal economy. Concerning the informal economy, the authors stated that a large and persistent informal sector (firms and workers) could foster stability within slums and attract more people to live in these areas. Hence, according to these authors, any discussion on informal settlements must consider their relationship with the informal economy.

Mahabir et al. (2016) also conducted an exhaustive literature review and identified four key factors influencing the growth of informal settlements: location, rural-to-urban migration, poor urban governance, and ill-designed policies. Although the authors did not directly mention the relationship between informal work and informal housing, they did highlight that the location choices made by slum dwellers are influenced, in part, by social and economic ties that may be related to similar income-generating activities. In other words, low-income informal workers may attract more informal workers, leading to more low-income people moving into the slums.

Regarding the literature in the Latin American context, we highlight the studies by Bouillon (2012), Posada and Moreno-Monroy (2019), and Alves (2021). In his study, Bouillon (2012) examined the difficulties and challenges facing the housing market in Latin America and the Caribbean. According to the author, the quantitative and qualitative housing deficit in this region, the barriers to accessing the financial system through mortgage loans, and the elevated costs of land and construction

are the main factors driving the formation of precarious settlements. Concerning the effects of labor informality, Bouillon (2012) argues that the inability of informal workers to document their income is the main obstacle to accessing a home loan program. In particular, the study finds that about 30% of households lack access to mortgage loans because they do not have sufficient income or cannot document their income. Thus, the growth of informal housing in Latin American cities is mainly caused by low-income populations' lack of access to the financial system and the absence of government policies to address this problem.

For their part, Posada and Moreno-Monroy (2019) studied the relationship between informal housing and informal employment in Sao Paulo (Brazil) and Bogotá (Colombia). The authors presented a set of stylized facts drawn from these cities and proposed a search model to analyze these types of informality within a unified framework. Among the stylized facts, the authors found that the distribution of informal work-related trips is more decentralized and, therefore, informal workers commute less frequently than formal workers. This pattern can be attributed to a higher prevalence of informal street or home activities. Regarding the housing market, the authors showed that informal settlements are more common in the city's periphery, while formal housing is concentrated in the city's center. By incorporating these stylized facts into a search model, the authors demonstrated that, during formalization, there is more intense competition for proximity to the city's center, where formal jobs are concentrated. This competition increases rental prices for formal housing and, thus, the displacement of informal workers to the city's periphery. Consequently, as the city expands to accommodate informal and formal workers, the prevalence of informal housing increases.

More recently, Alves (2021) investigated slum growth in Brazilian cities by estimating a spatial equilibrium model. His research contributes to the literature by demonstrating that urban wages rise as cities experience economic growth, attracting many low-income individuals who, given their limited human capital, are predominantly employed in the informal sector. However, as the costs of formal housing increase when compared to informal housing, a significant proportion of low-income individuals engaged in informal activities are compelled to reside in slum neighborhoods, thereby contributing to the expansion of such settlements.

Bonet et al. (2016) and Gallego et al. (2018) are among the few empirical studies in Colombia that have attempted to analyze informal housing and labor informality jointly. Bonet et al. (2016) studied housing informality and labor informality in Colombia's main cities, investigating their characteristics, dynamics, relationships, and the main factors influencing them. They conducted both aggregate analyses and analyses using household survey microdata. At the aggregate level, the authors found a positive relationship between housing informality and labor informality in the cities under analysis, which has persisted over time. Regarding the relationship between the two indicators of informality at the individual level, the study revealed that 30% of informal workers in Colombia live in informal settlements, while this percentage is 15% for formal workers. The analysis of microeconomic factors influencing informal housing and labor informality demonstrated that an informal worker increases the likelihood of living in an informal settlement. Their findings also showed that education plays an important role in reducing informality in both

housing and employment, while low-income levels and household size have a positive effect on informal housing.

For their part, Gallego et al. (2018) are the only authors who have examined informal housing and labor informality at the intraurban level. Even though they focused on investigating the determinants of labor informality at the intraurban level in Medellín (Colombia), they also established, for the first time, a relationship between informal housing and labor informality, considering the spatial dimension of informality. These authors showed the spatial simultaneity of labor informality and informal housing. In particular, they found that areas with high levels of informal housing also have high levels of labor informality and that urban informality is concentrated in the peripheral and impoverished areas of the city. Another important contribution of the authors is their consideration of spatial spillovers, an essential aspect of analyzing urban informality in space.

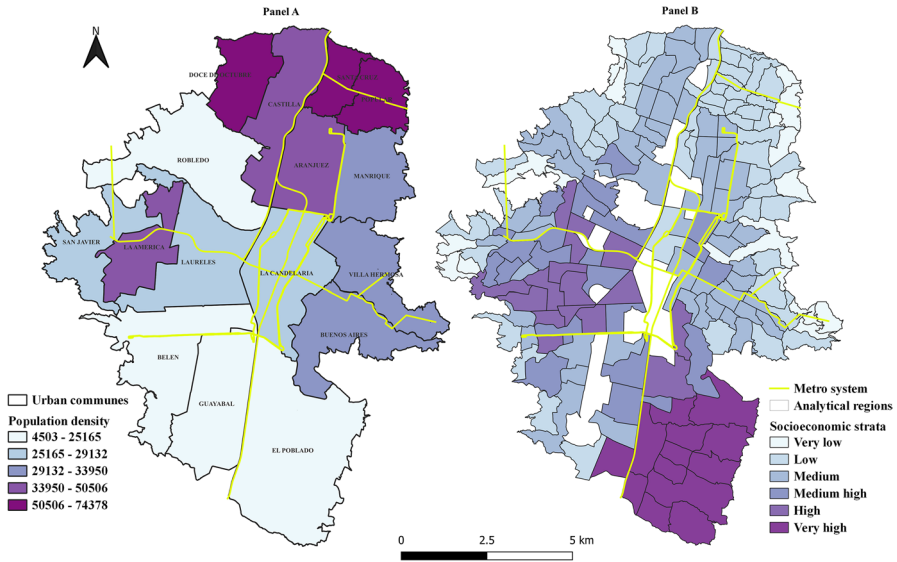
Despite the prevalence of informal housing and informal employment in Latin American cities, these two types of informality have mainly been studied separately, with little empirical evidence of the relationship between them. To address this gap in the literature, this study attempts to empirically investigate the simultaneity of informal housing and labor informality at the intraurban level. In addition, it seeks to examine the existence of spatial spillover effects, which are crucial in understanding this type of urban phenomenon.

## Data and Descriptive Statistics

### Study Area and Spatial Unit of Analysis

In this study, we use data from Medellín, a city in the region of Antioquia in northwestern Colombia and the second largest city in the country after Bogotá, the capital. The city has a population of 2.6 million people and spans an area of 380.64 km<sup>2</sup>, resulting in a population density of 6,749 inhabitants per km<sup>2</sup>. In terms of socioeconomic variables for 2017, Medellín is characterized by an important heterogeneity in the labor market, with unemployment and labor informality levels of 10.6% and 42%, respectively (Colombia: unemployment rate: 9.8%; labor informality rate: 47%) (DANE, 2023a). Regarding income inequality and poverty, Medellín presents a Gini coefficient of 0.52 and a poverty rate of 14.2% (Colombia: Gini coefficient: 0.51; poverty rate: 27%) (DANE, 2023b).

Our analysis focuses on the urban part of Medellín, which is divided into 16 districts, or communes, and 246 neighborhoods. Figure 1 shows the spatial distribution of population density and income levels across Medellín. As observed, the most densely populated areas are in the north and southwest of the city. Low and low-middle-income residents predominantly populate the northern part of the city, while wealthy residents primarily populate the southern part. Notably, Medellín stands out from other cities in Colombia and Latin America because of its public transportation system, the Metro system. This system has significantly improved the city's accessibility, especially for most remote and low-income residents (Bocarejo et al., 2014).



**Fig. 1** Study area (**Panel A**) Density and (**Panel B**) Distribution of income categories, Medellín. Notes: Population density (population/km<sup>2</sup>) by communes. Levels of income at the analytical regions level based on economic stratification in six categories (1 = very low, 6 = very high). Source: 2017 Quality-of-Life Survey for Medellín

We use analytical regions as the spatial unit of analysis. As mentioned above, Medellín has two levels of administrative divisions: communes (16) and neighborhoods (246). Communes are very large and internally heterogeneous regions, which poses challenges when using them as spatial units of analysis. In 2017, there was an average of 16 neighborhoods per commune in Medellín. The largest communes were *Robledo* and *El Poblado*, with 25 and 22 neighborhoods, respectively, and the smallest communes were *Guayabal* and *Santa Cruz*, with 9 and 11 neighborhoods, respectively (see Fig. 1). Using statistical inference based on large administrative regions, nonetheless, can lead to aggregation issues such as the ecological fallacy (Robinson, 1950), aggregation bias (Amrhein & Flowerdew, 1992; Fotheringham & Wong, 1991; Paelinck, 2000), and the Modifiable Areal Unit problem (MAUP) (Openshaw & Taylor, 1981).

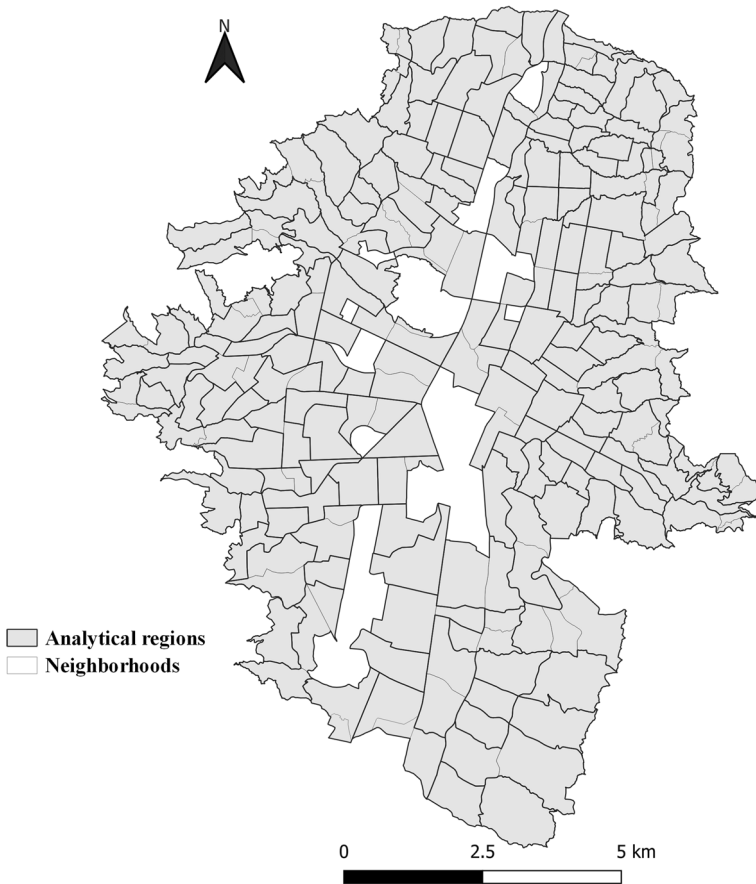
Employing neighborhoods as the spatial unit of analysis also entails certain limitations. First, statistical analysis at the neighborhood level may suffer from a lack of statistical validity because the data available for socio-economic studies in Medellín come from the Quality-of-Life Survey and are designed to be representative at the commune level. Second, calculating rates can be problematic due to the small number of households surveyed at the neighborhood level (Diehr, 1984). In 2017, for instance, the average number of households surveyed per neighborhood in Medellín was 113 (SD=74.4), with values ranging from 1 to 406, and 19% of neighborhoods had fewer than 50 households surveyed. Lastly, Weeks et al. (2007) noted that using administrative divisions at the neighborhood level could lead to spurious spatial autocorrelation issues.

An alternative approach to overcoming the limitations of using communes or neighborhoods as the spatial unit of analysis is to employ analytical regions. These regions are spatial units that meet specific criteria, such as size, shape, and internal homogeneity, making them suitable for studying socio-economic phenomena in space (Duque et al., 2013). In this study, we follow the strategy used by Duque et al. (2013) and Duque et al. (2015). According to these authors, each analytical region must satisfy two conditions: (1) it should be a homogeneous region in terms of a set of relevant socio-economic variables for the phenomenon under analysis (Duque et al., 2006, 2013), and (2) it should contain a minimum threshold surveyed households to ensure statistical representativeness (Duque et al., 2015). The first criterion contributes to minimizing aggregation bias, while the second requirement helps to solve the problem of small numbers of households surveyed (Diehr, 1984) and, in turn, reduces the impact of inaccuracies in geolocation (Duque et al., 2015).

To delineate the regions, we use the Max-P-regions algorithm using ClusterPy (Duque et al., 2011, 2013). This algorithm attempts to aggregate large regions into a maximum number of smaller regions, such that each new region is homogeneous concerning a set of socio-economic characteristics and satisfies a predefined minimum threshold of surveyed households. This approach is primarily a minimization process of intra-regional heterogeneity and aggregation bias, and it is the only algorithm that allows endogenizing the number of regions (Duque et al., 2013). Following the strategy used by Duque et al. (2013), who designed analytical regions to study intraurban poverty and calculate a slum index in Medellín, and which therefore fits our study quite well, we use information from the 2017 Quality-of-Life Survey for Medellín on variables related to house characteristics and households' members. In terms of the minimum threshold of surveyed households, we use the same established by Duque et al. (2013), set at a minimum of 100 surveyed households. According to the authors, this threshold attempts to capture the heterogeneity in the larger spatial units so that the new smaller regions created consider that heterogeneity. The authors also mention that the value of 100 surveyed households is oversampling to ensure statistical validity, but optimal threshold setting is an open question that requires further research. The number of analytical regions for Medellín was very similar to those found by Duque et al. (2013), where 246 neighborhoods were grouped into 176 analytical regions, with an average area of 0.521 km<sup>2</sup>. Figure 2 illustrates the comparison between analytical regions and the neighborhoods.

## Preliminary Evidence

The data used in this paper come from the 2017 Quality-of-Life Survey for Medellín. This annual cross-sectional survey provides individual-level information on household characteristics, demographics, education, social security, labor market dynamics, and poverty and socio-economic conditions indicators. The sample considered in this study includes individuals aged between 18 and 60 residing in the urban area of Medellín. It includes approximately 43,000 individuals and 12,000 households, an expanded sample of around 2.4 million individuals and 780,000 households. The data were aggregated at the analytical region level to calculate the rates for the variables of interest.



**Fig. 2** Spatial unit of analysis: 176 analytical regions

Regarding the definition of informal housing, we adopt the definition proposed by the Nations Human Settlements Program (UN-Habitat) (2003, 2006, 2016b), which includes some type of housing deprivation, such as low standards of urban services, insecure land tenure, or non-durable housing structures. According to this definition, slums are settlements that have one or more of the following characteristics: (1) walls that are not made of durable materials, (2) overcrowded inadequate living spaces (i.e., more than three people per room), or (3) lack access to either potable water or improved sanitation services.<sup>3</sup> This definition has been widely used in the literature and applied in

<sup>3</sup> UN-Habitat's definition of slum also includes a lack of home ownership; however, we do not consider this to indicate housing deprivation in the context of Medellín. This is because a significant number of both the middle-class and slum populations in the city do not own their homes. Notably, 35% of low-income individuals, 38% of middle-income individuals, and 28% of high-income individuals in Medellín are not homeowners. According to UN-Habitat (2006), the definition of a slum can be modified depending on each city's specific situation.

**Table 1** Descriptive statistics of informality, control and instrumental variables

	Mean	SD	P25	Median	P75	Min	Max
Housing informality (%)	6.83	6.11	1.93	5.74	10.79	0.00	26.45
Labor informality (%)	33.40	14.01	20.06	35.56	44.80	4.87	64.00
Control variables							
% Tertiary education	11.39	12.02	2.41	6.37	16.87	0.00	54.73
% Female head of household	15.63	3.95	12.90	15.30	17.58	7.79	34.14
% Children	5.02	2.58	3.12	4.69	6.83	0.00	14.13
Unemployment rate (%)	7.84	5.17	3.81	6.80	10.95	0.00	25.79
Population density	24.59	19.30	11.05	18.79	32.76	2.11	107.84
Instrumental variables							
Housing informality in 2011 (%)	8.40	9.38	0.00	5.90	12.26	0.00	51.72
Euclidean distance to CBD (km)	2.15	0.92	1.55	2.15	2.86	0.00	4.79

Population density = population/km<sup>2</sup>

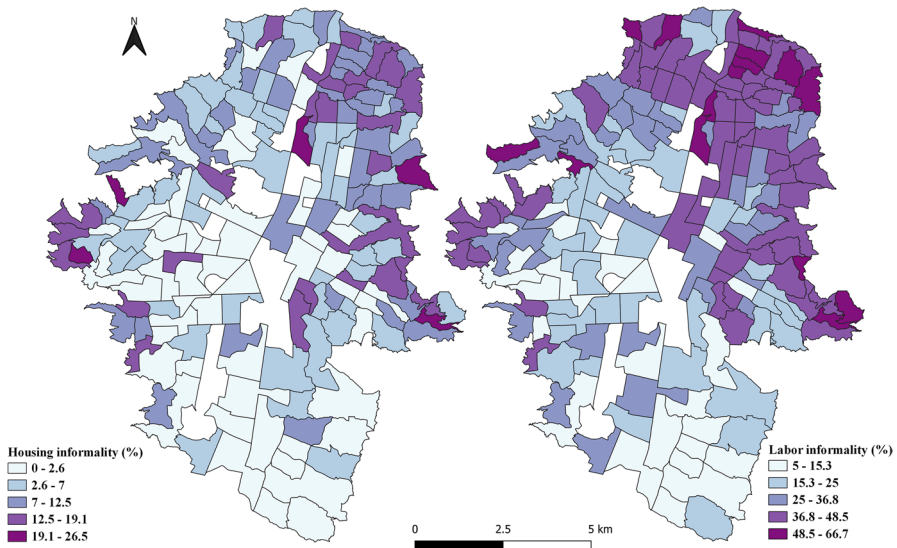
different countries. Some of these studies include those by Rains and Krishna (2020) in India; Brueckner (2013) in Indonesia; Aliu et al. (2021) in Nigeria; Duque et al. (2015) and Gallego et al. (2018) in Colombia; Galiani et al. (2017) in El Salvador, Mexico and Uruguay; and Cavalcanti et al. (2019) and Alves (2021) in Brazil.

As for the definition of labor informality, we opt for a legalistic definition that regards informality as the lack of mandated labor protection for workers. In particular, we define informal workers as those not covered by the health insurance and pension systems. This definition is commonly used in empirical studies on labor informality, as it is easy to implement using household surveys across different countries. Moreover, it is much broader, allowing for the inclusion of a more heterogeneous group of informal workers (García, 2019; Perry et al., 2007).

Using these definitions, we calculate the variables of interest related to informality at the analytical region level. On the one hand, we compute the informal housing rate as the proportion of households within each analytical region that exhibit any of the characteristics of a slum. On the other hand, the labor informality rate is determined by the ratio of workers not covered by the social security system to the total number of workers in each analytical region. By employing this approach, we found that, on average, 6.8% of dwellings in Medellín were classified as informal, with some regions having informal housing rates as high as 26%. In terms of labor informality, we find that the average labor informality rate in the analytical regions was 33%, with a maximum rate of 64% (see Table 1).

Figure 3 shows the spatial distribution of the two types of informality. As observed, there is a simultaneity of high levels of housing informality and labor informality in the city's most peripheral and poorest areas, specifically in the north-east and west. In these areas, the housing informality rates range from 19 to 26%, while the labor informality rates range from 48 to 66%.

To confirm the spatial simultaneity of the two types of informality, we compute a bivariate map of informal housing and labor informality. This map, shown in Fig. 4, confirms the strong spatial simultaneity of both types of informality. Regions with

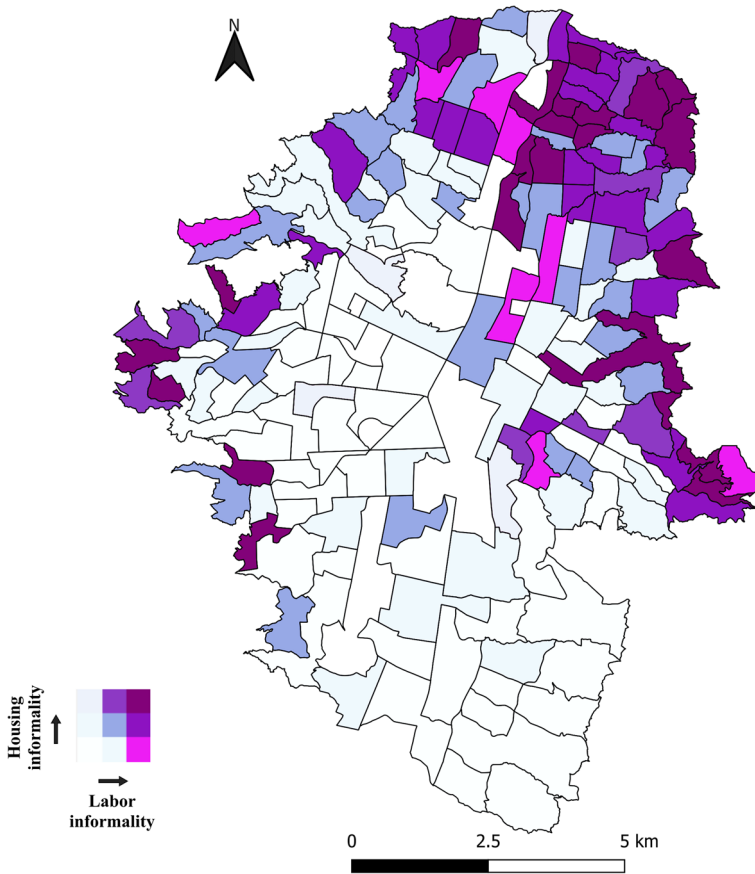


**Fig. 3** Spatial distribution of housing and labor informality, 2017

high levels of informal housing also exhibit high levels of labor informality, particularly in the city's eastern, western, and northern peripheries, which are densely populated by low-income individuals. Similar patterns of urban informality in peri-urban areas have been observed in other studies conducted in developing countries, such as that by Martínez-Jiménez et al. (2022).

Furthermore, to explore the spatial dependence and clustering processes between the two informality types, we compute a bivariate Local Indicator of Spatial Association (LISA) cluster map (Anselin et al., 2002). This bivariate LISA cluster map, depicted in Fig. 5, reveals local spatial correlation patterns at the analytical region level between housing informality and neighboring regions' average labor informality rate. These local statistics are vital because they show that the magnitude of the spatial relationship between informal housing and labor informality is unevenly distributed across space. Concerning clustering, the results show that the study area is dominated by spatial clusters (high-high and low-low locations) rather than by spatial outliers (high-low and low-high locations). Clustering processes of informality are observed, wherein areas with high levels of housing informality are surrounded by nearby areas with high levels of labor informality (hot spots of intraurban informality) in the north-east and certain eastern parts of the city. Conversely, areas with low levels of informal housing are surrounded by areas with low levels of labor informality (cold spots of intraurban informality) in the western and southern of the city. This north-south division of housing and labor conditions reveals the marked spatial segregation of these two phenomena at the intraurban level in Medellín, which is also associated with the socio-spatial segregation prevalent in the city (Villarraga et al., 2014).

We include socio-economic conditions related to human capital, demographics, and labor characteristics. Particularly, we consider the share of the population with tertiary education (% Tertiary education), the share of female heads of household (%

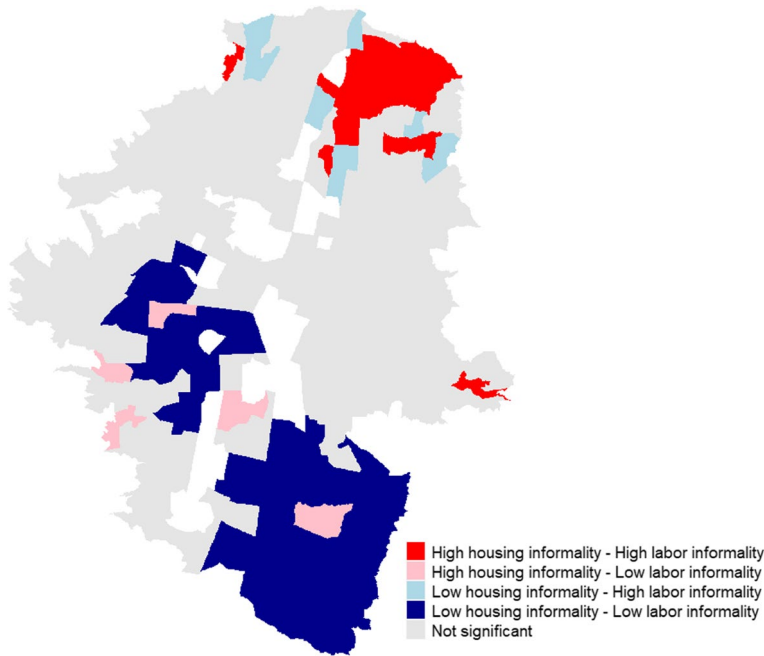


**Fig. 4** Bivariate map between housing informality and labor informality, 2017

Female heads of household), the share of children under six years of age (% Children), population density, and the unemployment rate. Table 1 presents the main descriptive statistics for these variables, and Fig. 6 displays their spatial distribution. As observed in the maps, the lowest levels of education and the highest levels of unemployment are concentrated in the peripheral and more informal areas of the city, which supports the existence of a relationship between urban informality and precarious socio-economic conditions. This relationship has also been documented by Roy et al. (2014), Villaraga et al. (2014), Mahabir et al. (2016), and Martínez-Jiménez et al. (2022).

## Econometric Model

To measure the relationship between housing informality and labor informality while considering the simultaneity of these two phenomena, we estimate the following system of two equations:



**Fig. 5** Bivariate LISA cluster map, 2017. Note: number of permutations=999 and significance level at 5%. First-order Queen criterion of contiguity was used to calculate the spatial weights

$$housing\ informality_r = \beta_{11}labor\ informality_r + \mathbf{x}'_r\boldsymbol{\omega}_1 + u_{1r} \tag{1}$$

$$labor\ informality_r = \beta_{21}housing\ informality_r + \mathbf{x}'_r\boldsymbol{\omega}_2 + u_{2r}, \tag{2}$$

where *housing informality<sub>r</sub>* and *labor informality<sub>r</sub>* represent the levels of housing and labor informality, respectively, in analytical region *r*, and they are the endogenous variables of the system. Term  $\mathbf{x}_r$  is a vector of control variables, which include the region’s human capital, demographics, and labor characteristics (see Table 1). Additionally,  $\boldsymbol{\omega}_1$  and  $\boldsymbol{\omega}_2$  are vectors of the coefficients to estimate, and  $u_{1r}$  and  $u_{2r}$  denote the stochastic errors. Our coefficients of interest to estimate are  $\beta_{11}$  and  $\beta_{21}$ , which represent the effect of labor informality on informal housing and vice versa, respectively.

To consider the spatial dimension of housing informality and labor informality, we estimate four types of spatial simultaneous models: the Spatial Lag model (SLM), the Spatial Error Model (SEM), the Spatial Autoregressive Combined model (SAC), and the Spatial Durbin Model (SDM). The SLM incorporates a spatial lag of the dependent variable as an explanatory factor; hence, the system is given by

$$housing\ informality_r = \beta_{11}labor\ informality_r + \rho_1Whousing\ informality_r + \mathbf{x}'_r\boldsymbol{\omega}_1 + u_{1r} \tag{3}$$

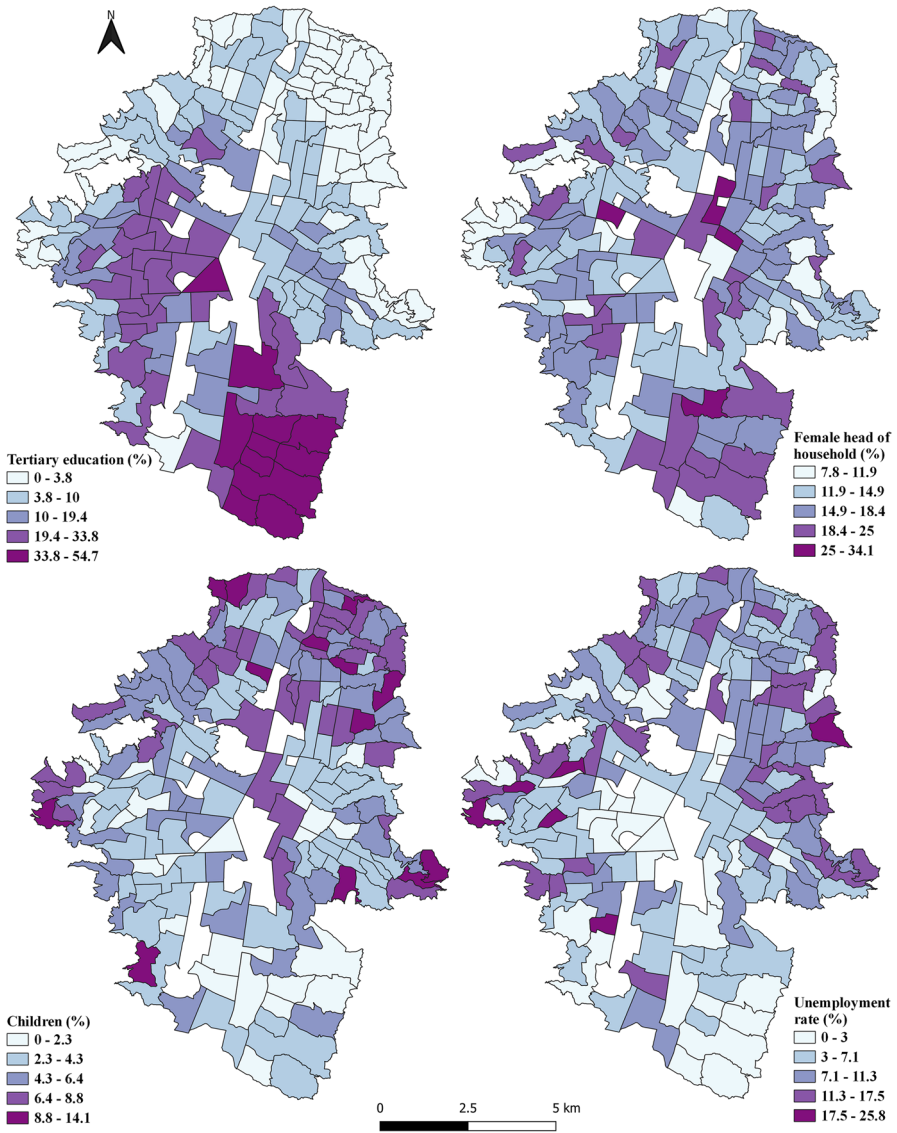


Fig. 6 Spatial distribution of the control variables

$$labor\ informality_r = \beta_{21}housing\ informality_r + \rho_2 Wlabor\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_2 + u_{2r} \tag{4}$$

For its part, the SEM incorporates the spatial dependence on stochastic disturbances, that is,

$$housing\ informality_r = \beta_{11}labor\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_1 + u_{1r} \tag{5}$$

$$u_{1r} = \gamma_1 W u_{1r} + \varepsilon_{1r} \quad (6)$$

$$labor\ informality_r = \beta_{21} housing\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_2 + u_{2r} \quad (7)$$

$$u_{2r} = \gamma_2 W u_{2r} + \varepsilon_{2r}, \quad (8)$$

where  $\varepsilon_{1r}$  and  $\varepsilon_{2r}$  are stochastic errors.

The SAC model combines the SLM and the SEM, including the spatial dependence on the dependent variable and stochastic disturbances. The structure of the SAC model is as follows:

$$housing\ informality_r = \beta_{11} labor\ informality_r + \rho_1 W housing\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_1 + u_{1r} \quad (9)$$

$$u_{1r} = \gamma_1 W u_{1r} + \varepsilon_{1r} \quad (10)$$

$$labor\ informality_r = \beta_{21} housing\ informality_r + \rho_2 W labor\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_2 + u_{2r} \quad (11)$$

$$u_{2r} = \gamma_2 W u_{2r} + \varepsilon_{2r} \quad (12)$$

Finally, the SDM incorporates the spatial lag of the dependent variable and the spatial lags of the independent variables as additional explanatory variables. The structure of this model is as follows:

$$housing\ informality_r = \beta_{11} labor\ informality_r + \lambda_1 W labor\ informality_r + \rho_1 W housing\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_1 + W \mathbf{x}'_r \boldsymbol{\theta}_1 + u_{1r} \quad (13)$$

$$labor\ informality_r = \beta_{21} housing\ informality_r + \lambda_2 W housing\ informality_r + \rho_2 W labor\ informality_r + \mathbf{x}'_r \boldsymbol{\omega}_2 + W \mathbf{x}'_r \boldsymbol{\theta}_1 + u_{2r} \quad (14)$$

In all the models,  $W$  is a matrix ( $r \times r$ ) of spatial connectivity between the analytical regions (Anselin, 1988), which, in our case, is a standardized first-order Queen matrix.

Estimating the proposed system of equations may face two problems related to endogeneity (Le Gallo & Fingleton, 2021). The first problem arises from including endogenous variables as regressors, that is, the inclusion of the dependent variable from the second equation (*labor informality*) in the first equation and the inclusion of the dependent variable from the first equation (*housing informality*) in the second equation. The second problem stems from including a spatial lag

in the model. To deal with these endogeneity problems, we use an Instrumental Variable (IV) approach.

To address the first endogeneity problem arising from the inclusion of endogenous variables as regressors, we use the Euclidean distance (in km) from the centroids of each analytical region to Medellín's Central Business District (CBD) as an instrument for the *labor informality* variable in the first equation (see Table 1 for some descriptive statistics). This instrument is based on the spatial mismatch hypothesis, which states that poor labor market outcomes, both in quantity and quality, can be attributed to intraurban disconnection (Kain, 1968). In other words, greater disconnection from or distance to job opportunities can lead to adverse labor market outcomes due to factors such as high commute costs, reduced effectiveness, and intensity of job searches, elevated costs associated with job searches, and employer discrimination against employees living in outlying and marginalized areas (Gobillon et al., 2007). Under this hypothesis, it is possible that the calculated instrument effectively identifies the effect of labor informality on housing informality since it is relevant to explaining labor informality but does not directly influence housing decisions (exclusion restriction).

Regarding the endogeneity problem related to the *housing informality* variable in the second equation, we use the level of housing informality lagged in time (2011) as an instrument. The idea behind this instrument is that the level of housing informality in each region in the past tends to explain the level of housing informality in subsequent periods. Indeed, using a time series approach, Bonet et al. (2016) confirmed the temporal inertia of housing informality in several cities in Colombia. According to Liu (2017), using lagged variables as instruments aims to avoid the simultaneity caused by local shocks. Therefore, the lagged level of housing informality could serve as a suitable instrument for relevance and homogeneity, as it explains the emergence of informal settlements without directly determining the current level of informal work in the regions. In the literature, lagged values of endogenous variables have been widely used as instruments in urban economics since the pioneering work of Ciccone and Hall (1996). Other applications of these lagged values can be found in studies such as those by Moretti (2004) and Liu (2017).

To address the second endogeneity problem associated with the inclusion of a spatial lag in the models, we follow the spatial econometric literature, which states that the spatial lags of a superior order of control variables are adequate instruments to correct this endogeneity issue (Kelejian & Robinson, 1993; Kelejian & Prucha, 1998; Fingleton & Le Gallo, 2010). According to López et al. (2020), spatial lags of order 2 are enough; thus, we use these variables as instrumental variables to tackle this endogeneity problem.

Another methodological aspect to be considered in estimating the proposed system of spatial simultaneous equations is the presence of cross-equation correlations in the error terms. This study assumes that unobserved regional characteristics influence informality in housing and labor markets. For instance, limited access to poor or informal public transportation may constrain individuals' ability to access formal employment opportunities and isolate certain parts of the city that could become informal settlements (Boisjoly et al., 2017; Posada & Moreno-Monroy, 2019; Vargas et al., 2017). Another relevant unobservable factor is the housing deficit, which is

quite evident in developing countries. This deficit limits access to public services, affecting health and skill accumulation, and incentivizes informal housing solutions, especially among the most economically disadvantaged segments of the population (Cavalcanti et al., 2019; Vargas et al., 2017).

In summary, our model is a system of spatial simultaneous equations that allows for correlation between the errors of the equations. We use the Generalized Spatial Three-Stage Least Square (GS3SLS) estimator proposed by Kelejian and Prucha (2004) to estimate this model. This procedure first corrects the endogeneity problem using a two-stage estimator, namely the GS2SLS method. Then, it considers the relationship between the equations in the system through the stochastic disturbances analogous to a Seemingly Unrelated Regression (SUR).

## Results

Table 2 presents the results of the non-spatial and spatial estimations.<sup>4</sup> For comprehensive estimates including all variables, direct and indirect effects, and first-stage estimates, please refer to Tables 3, 4, and 5 in the Appendix, respectively.

We begin by discussing the validity of the instruments, that is, that the instruments are exogenous and relevant. In the family of spatial linear models, there is not yet an exogeneity test of the instruments, such as the Sargan/Hansen test, that can be calculated in non-spatial linear models. In addition, the system of equations is exactly identified, which means that tests of overidentifying restrictions cannot be calculated. Therefore, we based the discussion on the exogeneity condition of the instruments on theoretical prediction and the literature review discussed in the previous section. Regarding the relevance of the instrument variables, we use the first-stage estimates to corroborate this condition. Column (1) in Table 5 in the Appendix shows that the level of housing informality lagged in time (2011) is relevant in explaining the current level of housing informality, as the estimated coefficient is positive and statistically significant. This result confirms the theoretical prediction made in the previous section, where it is mentioned that housing informality is a very persistent phenomenon (Bonet et al., 2016). In the case of the labor informality variable, which has the Euclidean distance of each analytical region to the CBD as its instrument variable, we observe in column (2) in Table 5 in the Appendix that the estimated coefficient is positive and statistically significant. This result confirms the spatial mismatch hypothesis, i.e., greater distance to the CBD or intraurban disconnection leads to poor labor market outcomes (Gobillon et al., 2007), which in our case is related to higher labor informality.

Considering the main estimates in Table 2, we observe that the estimated coefficients for housing and labor informality are positive and statistically significant, indicating that higher levels of labor informality positively impact the emergence of informal housing and vice versa. This finding confirms that these two urban informality

<sup>4</sup> The models are estimated using the function *spsur3s3ls* of the R package *spsur* (Angulo et al., 2020).

**Table 2** Estimates of system of spatial simultaneous equations

	Spatial models									
	Non-spatial model		SLM		SEM		SAC		SDM	
	Housing informality (1)	Labor informality (2)	Housing Informality (3)	Labor informality (4)	Housing informality (5)	Labor Informality (6)	Housing informality (7)	Labor informality (8)	Housing informality (9)	Labor informality (10)
A. Estimates										
Housing informality		1.009*** (0.1171)	0.339*** (0.0415)	1.021*** (0.1178)	0.251*** (0.0420)	0.788*** (0.1223)	0.300*** (0.0400)	0.987*** (0.1200)	0.382*** (0.0517)	1.011*** (0.1315)
Labor informality	0.345*** (0.0408)		-0.125 (0.1798)				-0.065 (0.1629)		0.223 (0.7538)	-0.434 (0.3891)
W Housing informality				0.118 (0.1191)				0.110 (0.1132)	-0.140 (0.1518)	0.465 (0.5339)
W Labor informality					-0.468*** (0.1458)	-0.069 (0.1499)		-0.174 (0.1650)		
$\rho$				0.628	0.356	0.652	0.343	0.632	0.343	0.649
R <sup>2</sup>	0.329	0.629	0.333	0.628	0.356	0.652	0.343	0.632	0.343	0.649
Breusch-Pagan test	30.47***	30.47***	30.47***	30.47***	13.13***	27.62***	27.62***	27.62***	35.85***	35.85***
N	176	176	176	176	176	176	176	176	176	176
LM tests of spatial dependence										
LM <sub>SLM</sub>					4.930*					
Robust LM <sub>SLM</sub>					0.573					
LM <sub>SEM</sub>					7.958**					
Robust LM <sub>SEM</sub>					3.601					
LM <sub>SAC</sub>					8.532*					

All models include the set of control variables mentioned in Table 1. Standard errors in parenthesis. Spatial models are estimated by Generalized Spatial Three-Stage Least Square (GS3SLS). We instrument labor informality and housing informality variables using the Euclidean distance (in km) from the centroids of each analytical region to the central business district (CBD) of Medellín and the level of housing informality lagged in time (2011), respectively. The inclusion of spatial lag in the models is instrument using the spatial lag of order 2 of control variables.  $W$  is a matrix ( $r \times r$ ) of spatial connectivity among analytical regions, which, in our case, is a standardized first-order Queen type. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

phenomena affect each other. To corroborate the presence of spatial dependence in this system of simultaneous equations, we compute the Lagrange Multiplier (LM)-lag and LM-error tests of spatial dependence (Mur et al., 2010). According to the results of these tests, which are reported at the bottom of Table 2, both simple LM tests are significant, suggesting the presence of spatial dependence in the model. The robust LM tests, for their part, help us understand what type of spatial dependence may be at work. The results show that the LM-lag robust to spatial dependence in error is still significant, while the LM-error robust to spatial dependence in lag becomes insignificant, indicating that when the spatial dependence of the error is considered, the spatial autocorrelation through the lagged dependent variable disappears.

Furthermore, we assessed the correlation between the error terms of the equations using the standard Breusch-Pagan test, which is employed to test the diagonality of the variance-covariance matrix of the disturbances across equations (Baltagi, 2011). As observed in the results presented in Table 2, the Breusch-Pagan test is statistically significant (in all the models), which rejects the null hypothesis of zero correlation between the two equations. This means it is important to consider the presence of cross-equation correlations in the error terms when estimating the system of equations related to informality.

Columns 3 to 10 of Table 2 report the estimates of the spatial models. The results for all models consistently indicate that labor informality has a positive and statistically significant effect on informal housing, suggesting that precarious labor conditions are a crucial driver of precarious housing conditions. Moreover, the estimates reveal that informal housing positively and significantly affects labor informality. In this regard, several studies have highlighted the negative effects of low-quality housing on multiple factors of individual and family well-being, including physical and mental health, stress indicators, family interactions, and work-related aspects (Cattaneo et al., 2009; Field, 2005; Galiani et al., 2017). In the case of the latter, the literature has shown that inadequate housing conditions are associated with a higher susceptibility to illness, increased time spent on cleaning tasks, reduced labor productivity, limited time available for work, and a greater propensity to engage in low-quality jobs (Cattaneo et al., 2009). These findings confirm the bidirectional causality between these two types of urban informality at the intraurban level in Medellín.

Since the point estimates do not account for the presence of indirect effects associated with the spatial structure of the models, we estimated both the direct and indirect effects, along with their statistical significance in the SLM and the SAC model (see Table 4 in the Appendix) following LeSage and Pace (2009), Elhorst (2010) and Halleck Vega and Elhorst (2015). The estimated direct effects of labor informality on informal housing reaffirm our previous findings that informal work activities are critical to the emergence of informal settlements. As noted in the literature review, the low wages earned by informal workers and their inability to document their income exclude them from accessing the financial system and formal home ownership opportunities, these workers are forced to rely on informal housing as a form of alternative housing.

In the SLM and the SAC model, we observed that the indirect effect (or spillover effect) of labor informality on informal housing is not statistically significant, indicating that the level of informal housing in one region is not affected by the level of

labor informality in neighboring regions. This result confirms the highly localized relationship between informal work activities and informal settlements, implying that many poor workers may find shelter and work opportunities within the slum economy. This is consistent with the findings of Marx et al. (2013) and van Ham and Manley (2015), who reported that although slum economies may offer a greater range of services and work opportunities, they face structural constraints that lead to less formal housing and slum growth.

## Conclusions

This study provides evidence of a positive relationship between informal housing and informal work activities at the intraurban level in Medellín (Colombia). Additionally, the descriptive analysis revealed that these two types of urban informality are simultaneously spatially persistent phenomena, with pronounced spatial clustering in the peripheral and marginalized areas of the city. According to these results, urban informality occurs locally and is concentrated in poor and peripheral areas.

Moreover, the estimated econometric models, which consider the spatial coexistence of these phenomena and specific unobservable local characteristics that simultaneously affect both types of informality, showed that labor informality positively affects the emergence of informal housing. This positive relationship indicates that slums may be the only housing solution for informal workers within a city, given the employment conditions that exclude them from accessing formal housing options (Burdett & Sudjic, 2011; Cavalcanti, 2019). Furthermore, individuals engaged in informal employment may find it more expensive to reside in formal housing rather than in informal housing. As a result, a significant proportion of low-income individuals involved in informal activities may be drawn to live in marginalized neighborhoods, thereby contributing to their expansion.

Regarding the spillover effects of labor informality on informal housing, the results indicate that the level of informal work in surrounding neighborhoods does not affect the level of informal housing. This may suggest a highly localized relationship between informal housing and labor informality due to the peculiarities of the informal work carried out within homes, leading to precarious housing conditions. This observation may support the idea that the lack of adequate transportation infrastructure in informal peripheries hinders access for a large part of the low-income population to areas where formal employment is concentrated. Hence, poor informal workers tend to concentrate in informal peripheral suburban belts and engage in more decentralized activities than formal workers (Posada & Moreno-Monroy, 2019; Suárez et al., 2016).

Our empirical findings highlight the importance of considering the working conditions of informal settlement dwellers when designing housing solutions for socio-economically disadvantaged citizens living in precarious conditions in peripheral areas. In this regard, within the Sustainable Development Goals (SDGs) framework, strategies have been developed to guarantee access to affordable housing financing options for informal workers. In addition, improving the likelihood of securing formal jobs is expected to reduce urban informality significantly.

**Appendix**

**Table 3** Estimates of system of spatial simultaneous equations

	Spatial models									
	Non-spatial					Spatial models				
	SLM		SEM		SAC		SDM			
Housing informality (1)	Labor informality (2)	Housing informality (3)	Labor informality (4)	Housing informality (5)	Labor Informality (6)	Housing informality (7)	Housing informality (8)	Housing informality (9)	Labor informality (10)	
Housing informality	0.713*** (0.1202)		1.021*** (0.1178)	1.021*** (0.1178)		1.021*** (0.1178)	0.844*** (0.1212)		0.970*** (0.1300)	
Labor informality	0.259 (0.0436)		0.339*** (0.0415)	0.339*** (0.0415)		0.262*** (0.0414)	0.375*** (0.0524)		0.375*** (0.0524)	
% Tertiary education	0.017 (0.0565)	-0.674*** (0.0769)	0.082 (0.0554)	-0.637 (0.0782)	0.082 (0.0554)	-0.612*** (0.0798)	0.022 (0.0565)	-0.654*** (0.0784)	0.109 (0.0690)	-0.602*** (0.0873)
% Female head of household	-0.106 (0.1001)	0.398** (0.1643)	-0.151 (0.1011)	0.389** (0.1674)	-0.151 (0.1011)	0.390** (0.1713)	-0.230* (0.0995)	0.483** (0.1666)	-0.160 (0.1324)	0.333* (0.2009)
% Children	0.269 (0.1743)	0.288 (0.2917)	0.219 (0.1741)	0.238 (0.2970)	0.219 (0.1741)	0.180 (0.3037)	0.258 (0.1656)	0.218 (0.2942)	0.170 (0.1958)	0.270 (0.3188)
Unemployment rate	0.102 (0.0795)	-0.220* (0.1319)	0.124 (0.0801)	-0.252* (0.1383)	0.124 (0.0801)	-0.252* (0.1383)	0.118 (0.0734)	-0.254 (0.1312)	0.118 (0.0894)	-0.230 (0.1432)
Population density	0.009 (0.0279)	0.082 (0.0460)	0.001 (0.0278)	0.068 (0.0468)	0.001 (0.0278)	0.068 (0.0479)	0.006 (0.0261)	0.074 (0.0462)	-0.003 (0.031)	0.074 (0.0488)
W Housing informality			-0.125 (0.1798)				-0.061 (0.1647)		0.220 (0.7627)	-0.503 (0.3816)
W labor informality			0.118 (0.1191)					0.122 (0.1113)	-0.150 (0.1537)	0.626 (0.5108)
W % Tertiary education									-0.007 (0.1923)	0.244 (0.4561)
W % Female head of household									-0.518* (0.2825)	0.647 (0.5153)

Table 3 (continued)

	Spatial models									
	Non-spatial					Spatial models				
	SLM		SEM		SAC		SDM			
Housing informality (1)	Labor informality (2)	Housing informality (3)	Labor informality (4)	Housing informality (5)	Labor Informality (6)	Housing informality (7)	Housing informality (8)	Housing informality (9)	Labor informality (10)	
<i>W</i> % Children rate										
<i>W</i> Unemployment rate										
<i>W</i> Population density										
$\rho$					-0.125 (0.1798)	0.118 (0.1191)	-0.500* (0.2194)	-0.177 (0.1697)		
R <sup>2</sup>	0.357	0.658	0.333	0.628	0.333	0.628	0.353	0.647	0.346	0.648
Breusch-Pagan test	8.543***		30.47***		30.47***		16.37***		32.08***	
N	176		176		176		176		176	
LM tests of spatial dependence										
LM <sub>SLM</sub>		4.930*								
Robust LM <sub>SLM</sub>		0.573								
LM <sub>SEM</sub>		7.958**								
Robust LM <sub>SEM</sub>		3.601								
LM <sub>SAC</sub>		8.532*								

All models include the set of control variables mentioned in Table 1. Standard errors in parenthesis. Spatial models are estimated by Generalized Spatial Three-Stage Least Square (GS3SLS). We instrument labor informality and housing informality variables using the Euclidean distance (in km) from the centroids of each analytical region to the central business district (CBD) of Medellín and the level of housing informality lagged in time (2011), respectively. The inclusion of spatial lag in the models is instrument using the spatial lag of order 2 of control variables. *W* is a matrix (*r* x *r*) of spatial connectivity among analytical regions, which, in our case, is a standardized first-order Queen type. \**p* < 0.1; \*\**p* < 0.05; \*\*\**p* < 0.01

**Table 4** Direct and indirect spillovers effects

	SLM			SAC		
	DE	IE	TE	DE	IE	TE
<b>Housing informality</b>						
Labor informality	0.306*** (0.0425)	-0.035 (0.0521)	0.270*** (0.0671)	0.263** (0.0404)	-0.015 (0.0447)	0.247** (0.0638)
% Tertiary education	0.054 (0.0571)	-0.006 (0.0141)	0.048 (0.0548)	0.022 (0.0575)	-0.001 (0.0108)	0.021 (0.0573)
% Female head of household	-0.137 (0.1031)	0.015 (0.0306)	-0.121 (0.0941)	-0.230** (0.1020)	0.013 (0.0419)	-0.217** (0.0997)
% Children	0.240 (0.1847)	-0.027 (0.0538)	0.212 (0.1789)	0.258 (0.1703)	-0.015 (0.0546)	0.243 (0.1695)
Unemployment rate	0.119 (0.0845)	-0.013 (0.0258)	0.105 (0.0770)	0.118 (0.0761)	-0.006 (0.0222)	0.111 (0.0732)
Population density	0.004 (0.0290)	-0.000 (0.0056)	0.004 (0.0267)	0.006 (0.0270)	-0.000 (0.0048)	0.005 (0.0269)
<b>Labor informality</b>						
Housing informality	0.895*** (0.1232)	0.134 (0.1457)	1.029*** (0.2190)	0.846*** (0.1242)	0.115 (0.1339)	0.961*** (0.2094)
% Tertiary education	-0.639*** (0.0793)	-0.095 (0.1028)	-0.735 (0.1378)	-0.656*** (0.0797)	-0.089 (0.0963)	-0.745*** (0.1274)
% Female head of household	0.390** (0.1655)	0.058 (0.0693)	0.449** (0.2016)	0.485** (0.1692)	0.065 (0.0795)	0.551** (0.2073)
% Children	0.238 (0.2857)	0.035 (0.0730)	0.274 (0.3397)	0.219 (0.1374)	0.029 (0.0518)	0.249 (0.1703)
Unemployment rate	-0.248* (0.1364)	-0.037 (0.0540)	-0.286* (0.1712)	-0.254* (0.0467)	-0.034 (0.0144)	-0.289* (0.0545)
Population density	0.073 (0.0478)	0.011 (0.0160)	0.084 (0.0576)	0.074 (0.0467)	0.010 (0.0144)	0.085 (0.0545)

*DE* Direct effects; *IE* Indirect effects; *TE* Total effects. Standard errors in parenthesis

\*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.1$

**Table 5** First stage estimates

	Housing informality (1)	Labor informality (2)
Housing informality in 2011 (%)	0.238*** (0.0467)	
Euclidean distance to CBD (km)		1.771** (0.6941)
% Tertiary education	-0.104** (0.0460)	-0.847*** (0.0756)
% Female head of household	-0.036 (0.0958)	0.443*** (0.1677)
% Children	0.200 (0.1727)	0.459 (0.2958)
Unemployment rate	0.031 (0.0768)	-0.216 (0.1344)
Population density	0.056** (0.0268)	0.103** (0.0465)
R <sup>2</sup>	0.428	0.665
N	176	176

Standard errors in parenthesis. \*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.1$

**Acknowledgements** Erika Badillo wishes to acknowledge the financial support from the Universidad Autónoma Latinoamericana (Unaula) Research Fund, code 35-000017, she also acknowledges for the support from the Scientific Colombia Program – EFI Alliance, code 60185, under agreement No. FP44842-220-2018.

**Funding** Open Access funding provided by Colombia Consortium

**Data Availability** The data and codes supporting the findings of this study are available at the following link <https://gusgarciaacruz.github.io/InformalHousingLabor>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abramo, P. (2009). Social innovation, reciprocity and the monetarization of territory in informal settlements in Latin American cities. Social innovation and territorial development. In D. MacCallum, S. V. Haddock, & F. Moulaert (Eds.), *Social innovation and territorial development* (1st ed., pp. 115–130). Routledge.
- Aliu, I., Akoteyon, I., & Soladoye, O. (2021). Living on the margins: Socio-spatial characterization of residential and water deprivations in Lagos informal settlements, Nigeria. *Habitat International*, 107, 102293.

- Alves, G. (2018). Determinants of slum formation: the role of local politics and policies. *Working paper, No 2018/06*. CAF.
- Alves, G. (2021). Slum growth in Brazilian cities. *Journal of Urban Economics*, 122, 103327.
- Amrhein, C., & Flowerdew, R. (1992). The effect of data aggregation on a Poisson regression model of Canadian migration. *Environment and Planning A*, 24(10), 1381–1391.
- Angulo, A., López, F. A., Minguez, R., & Mur, J. (2020). *Sponsor: Spatial seemingly unrelated regression models*. Software Manual. <http://github.com/rominsal/spsur>
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer.
- Anselin, L., Syabri, I., & Smirnov, O. (2002). Visualizing multivariate spatial correlation with dynamically linked windows. In L. Anselin & S. Rey (Eds.), *New tools for spatial data analysis: proceedings of the specialist meeting*. Center for Spatially Integrated Social Science (CSISS).
- Azunre, G., Amponsah, O., Takyi, S., & Mensah, H. (2021). Informality-sustainable city nexus: The place of informality in advancing sustainable Ghanaian Cities. *Sustainable Cities and Society*, 67, 102707.
- Baltagi, B. (2011). *Econometrics* (5th ed.). Springer.
- Basu, A. K., & Chau, N. H. (2015). Informal work in developing countries. In *International encyclopedia of the social & behavioral sciences: Second Edition*.
- Bocarejo, J., Portilla, I., Velásquez, J., Cruz, M., Peña, A., & Oviedo, D. (2014). An innovative transit system and its impact on low-income users: The case of the Metrocable in Medellín. *Journal of Transport Geography*, 39, 49–61.
- Boisjoly, G., Moreno-Monroy, A., & El-Geneidy, A. (2017). Informality and accessibility to jobs by public transit: Evidence from the São Paulo metropolitan region. *Journal of Transport Geography*, 64, 89–96.
- Bonet, J., Pérez-Valbuena, G., & Chirivía-Bonilla, E. J. (2016). Informalidad laboral y en la vivienda: primeros indicios para las principales ciudades colombianas. In *Documentos de trabajo sobre Economía Regional y Urbana* (no. 241). Banco de la Republica de Colombia.
- Bonnet, F., Vanek, J., & Chen, M. (2019). *Women and men in the informal economy – A statistical brief*. WIEGO.
- Bouillon, C. (2012). *Un espacio para el desarrollo: Los mercados de vivienda en América latina y el caribe*. Inter American Development Bank.
- Brueckner, J. (2013). Slums in developing countries: New evidence for Indonesia. *Journal of Housing Economics*, 22(4), 278–290.
- Burdett, R., & Sudjic, D. (Eds.). (2011). *Living in the Endless City: The Urban Age Project by the London School of Economics and Deutsche Bank's Alfred Herrhausen Society*. Phaidon.
- Cattaneo, M., Galiani, S., Gertler, P., Martinez, S., & Titiunik, R. (2009). Housing, health, and happiness. *American Economic Journal: Economic Policy*, 1(1), 75–105.
- Cavalcanti, A. (2017). Work, slums and informal settlements traditions: Architecture in the Favela do Telegrafo. *Traditional Dwellings and Settlements Review*, 28(2), 71–81.
- Cavalcanti, A. (2019). *Urban informality shaped by labor: addressing the spatial logics of favelas*. PhD dissertation, Delft University of Technology.
- Cavalcanti, T., Da Mata, D., & Santos, M. (2019). On the determinants of slum formation. *The Economic Journal*, 129(621), 1971–1991.
- Ciccone, A., & Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, 86(1), 54–70.
- DANE. (2023a). *Mercado Laboral* [Data file]. Retrieve from: <https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral>
- DANE. (2023b). *Pobreza y Desigualdad* [Data file]. Retrieve from: <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-moneteria>
- Del Carpio, M., & Patrick, C. (2021). Agglomeration and informality: Evidence from Peruvian establishments. *Journal of Regional Science*, 61(2), 442–471.
- Deléchat, C. & Medina, L. (2021). *The global informal workforce: priorities for inclusive growth*. International Monetary Fund.
- Diehr, P. (1984). Small area statistics: Large statistical problems. *American Journal of Public Health*, 74(4), 313–314.
- Duque, J., Artís, M., & Ramos, R. (2006). The ecological fallacy in a time series context: Evidence from Spanish regional unemployment rates. *Journal of Geographical Systems*, 8, 391–410.
- Duque, J., Dev, B., Betancourt, A., & Franco, J. (2011). ClusterPy: library of spatially constrained clustering algorithms, Version 0.9.9. RiSE-group (Research in Spatial Economics). EAFIT University. Retrieved from <http://code.google.com/p/clusterpy/>

- Duque, J., Royuela, V., & Noreña, M. (2013). A stepwise procedure to determine a suitable scale for the spatial delineation of urban slums. In E. Fernandez & F. RubieraMorollón (Eds.), *Defining the spatial scale in modern regional analysis. Advances in spatial science* (pp. 237–254). Springer-Verlag.
- Duque, J., Patiño, J., Ruiz, L., & Pardo-Pacual, J. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landscape and Urban Planning*, 135, 11–21.
- Elhorst, J. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, 5(1), 9–28.
- Field, E. (2005). Property rights and investment in urban slums. *Journal of the European Economic Association*, 3(2–3), 279–290.
- Fingleton, B., & Le Gallo, J. (2010). Estimating spatial models with endogenous variables, a spatial lag, and spatially dependent disturbances: Finite simple properties. *Papers in Regional Science*, 87(3), 319–339.
- Fotheringham, A., & Wong, D. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, 23(7), 1025–1044.
- Galiani, S., Gertler, P., Undurraga, R., Cooper, R., Martínez, S., & Ross, A. (2017). Shelter from the storm: Upgrading housing infrastructure in Latin American slums. *Journal of Urban Economics*, 98, 187–213.
- Gallego, S., Muñoz, E., & García, G. (2018). Análisis espacial de la informalidad laboral intraurbana en Medellín. *Sociedad y Economía*, 35, 9–31.
- García, G. (2019). Agglomeration economies in the presence of an informal sector: The Colombian case. *Revue d'Économie Régionale and Urbaine*, 2, 355–388.
- Gobillon, L., Selod, H., & Zenou, Y. (2007). The mechanisms of spatial mismatch. *Urban Studies*, 44(12), 2401–2427.
- Halleck Vega, S., & Elhorst, J. (2015). The SLX model. *Journal of Regional Science*, 55(3), 339–363.
- ILO. (1993). *Resolution concerning statistics of employment in the informal sector*. Fifteenth International Conference of Labour Statisticians. International Labour Organization (ILO). Accessed Jun 2023
- ILO. (2018). *Women and men in the informal economy: A statistical picture* (3rd ed.). ILO.
- Kain, J. F. (1968). Housing segregation, Negro employment, and metropolitan decentralization. *Quarterly Journal of Economics*, 82(2), 175–197.
- Kelejian, H., & Prucha, I. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, 17, 99–121.
- Kelejian, H., & Prucha, I. (2004). Estimation of simultaneous systems of spatially interrelated cross-sectional equations. *Journal of Econometrics*, 118(1–2), 27–50.
- Kelejian, H., & Robinson, D. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors and an application to a county expenditure model. *Papers in Regional Science*, 72, 297–312.
- Le Gallo, J., & Fingleton, B. (2021). Endogeneity in spatial models. In M. M. Fischer & P. Nijkamp (Eds.), *Handbook of Regional Science* (2nd ed., pp. 2237–2265). Springer-Verlag.
- LeSage, J., & Pace, R. (2009). *Introduction to spatial econometrics*. Taylor and Francis CRC Press.
- Liu, S. (2017). Agglomeration, urban wage premiums, and college majors. *Journal of Regional Science*, 57(4), 611–630.
- López, F. A., Mínguez, R., & Mur, J. (2020). ML versus IV estimates of spatial SUR models: Evidence from the case of Airbnb in Madrid urban area. *The Annals of Regional Science*, 64(2), 313–347.
- Mahabir, R., Crooks, A., Croitoru, A., & Agouris, P. (2016). The study of slums as social and physical constructs: Challenges and emerging research opportunities. *Regional Studies, Regional Science*, 3(1), 399–419.
- Maloney, W. (2004). Informality revisited. *World Development*, 32(7), 1159–1178.
- Martínez-Jiménez, E., Le Gallo, J., Pérez-Campuzano, E., & Aguilar, A. (2022). The effect of land Price in the peri-urban fringe of Mexico City: Environmental amenities for informal land parcel purchasers. *Urban Studies*, 59(1), 222–241.
- Marx, B., Stoker, T., & Suri, T. (2013). The economics of slums in the developing world. *The Journal of Economic Perspectives*, 27(4), 187–210.
- Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1–2), 175–212.
- Mur, J., López, F., & Herrera, M. (2010). Testing for spatial effect in seemingly unrelated regressions. *Spatial Economic Analysis*, 5(4), 399–440.

- OECD/ILO. (2019). *The Key Indicators of Informality based on Individuals and their Households*. OECD Publishing.
- Openshaw, S., & Taylor, P. (1981). The modifiable areal unit problem. In N. Wrigley & R. J. Bennett (Eds.), *Quantitative geography* (pp. 60–70). Routledge.
- Paelinck, J. H. (2000). On aggregation in spatial econometric modeling. *Journal of Geographical Systems*, 2(2), 157–165.
- Perry, G. Maloney, W., Arias, O., Fajnzylber, P., Mason, A., & Saavedra-Chanduvi, J. (2007). *Informality: Exit and Exclusion*. World Bank Latin American and Caribbean Studies. World Bank.
- Posada, H. (2018). Informal housing, spatial structure, and city compactness. *Journal of Regional Science*, 58(4), 822–836.
- Posada, H. M., & Moreno-Monroy, A. I. (2017). Informality, city structure and rural–urban migration in Latin America. *The Annals of Regional Science*, 59(2), 345–369.
- Posada, H., & Moreno-Monroy, A. (2019). Informal employment and informal housing in Latin America: Facts, theory, and ideas for a unified framework. *Revue d'Économie Régionale and Urbaine*, 2, 213–237.
- Rains, E., & Krishna, A. (2020). Precarious gains: Social mobility and volatility in urban slums. *World Development*, 132, 105001.
- Rallet, A. (2000). De la globalisation à la proximités géographique. In J. P. Gilly, & A. Torre (Ed.), *Dynamiques de proximités*. L'Harmattan, Paris.
- Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357.
- Roy, D., Lees, M. H., Palavalli, B., Pfeffer, K., & Sloot, M. P. (2014). The emergence of slums: A contemporary view on simulation models. *Environmental Modelling and Software*, 59, 76–90.
- Suárez, M., Murata, M., & Delgado, J. (2016). Why do the poor travel less? Urban structure, commuting and economic informality in Mexico City. *Urban Studies*, 53(12), 2548–2566.
- UN-Habitat. (2003). *The challenge of slums: global report on human settlements*. UN-Habitat.
- UN-Habitat. (2006). *State of the world's cities 2006/7*. Human Settlements Programme.
- UN-Habitat. (2010). *State of the Cities 2010-11 - cities for All: Bridging the urban divide*. UN-Habitat.
- UN-Habitat. (2011). *For a better urban future: a mission for the 21st century*. United Nations Human Settlements Programme.
- UN-Habitat. (2012). *State of the world's cities 2012/13: Prosperity of cities*. UN-Habitat.
- UN-Habitat. (2013). *Planning and design for sustainable urban mobility: global report on human settlements 2013*. Routledge.
- UN-Habitat. (2016a). *World cities report 2016: urbanization and development—emerging futures*. UNHabitat.
- UN-Habitat. (2016b). *Slum Almanac 2015/2016: tackling improvement in the Lives of slum dwellers*. UNHabitat.
- van Ham, M., & Manley, D. (2015). Occupational mobility and living in deprived neighborhoods: Housing tenure differences in “neighborhood effects.” *Applied Spatial Analysis and Policy*, 8, 309–324.
- Vargas, J., Brassiolo, P., Sanguinetti, P., Daude, C., Goytia, C., Álvarez, F., Estrada, R., & Fajardo, G. (2017). *RED 2017. Urban growth and access to opportunities: a challenge for Latin America*. CAF.
- Villarraga, H., Sabater, A., & Módenes, J. (2014). Modeling the spatial nature of household residential mobility within municipalities in Colombia. *Applied Spatial Analysis and Policy*, 7, 203–223.
- Weeks, J., Hill, A., Stow, D., Getis, A., & Fugate, D. (2007). Can we spot a neighborhood from the air? Defining neighborhood structure in Accra, Ghana. *Geojournal*, 69(1–2), 9–22.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.