EURASIAN GEOGRAPHY AND ECONOMICS
https://doi.org/10.1080/15387216.2022.2112608

Routledge
Taylor & Francis Group

Urban wage premium and informal agglomeration in Vietnam

Thi Bich Tran[a], Hai Anh La[b], Nguyet Hoang Minh Le[a] and Khoa Dang Nguyen[a]

[a]Faculty of Statistics, The National Economics University, Hanoi, Vietnam; [b]National Centre for Social and Economic Modelling, University of Canberra, Canberra, Australia

ABSTRACT

Despite the ongoing debate on the interaction between the informal sector and agglomeration economies in cities, little is known about whether informal agglomeration generates the wage premium in cities in developing countries. This paper provides empirical evidence on the impacts of urbanization and localization, separated from formal and informal sources, on the earnings of informal workers and how the concentration of informal workers affects the wages of formal employees in cities in Vietnam. Using data from the 2013-2020 Vietnam Labor Force Surveys, this paper shows higher wages in bigger cities. In addition to workers' skills and local non-human endowment, agglomeration generates the wage premium in cities. Nevertheless, the impacts are different between small and large cities. Our results show that while informal workers benefit from urbanization and localization in large cities, their existence is over-concentrated in small cities and crowds out the formal sector in large cities of Vietnam. Thus, urban development policies should focus on the over-concentration of informal sector workers in cities.

ARTICLE HISTORY
Received 31 January 2020
Accepted 9 August 2022

KEYWORDS
Agglomeration; informality; wage disparities; Vietnam

Introduction

Wages are higher in bigger cities and urban areas than in smaller towns and rural places. This city and urban wage premium is one of the factors which attract workers, including informal laborers, to migrate to urban places and concentrate more in bigger cities. The last two decades have seen an increasing interest in the urban economic literature on explaining the benefits of the concentration of economic activities and workers in clusters in cities and urban places. These benefits are known as agglomeration economies or agglomeration externalities that come from various advantages of locating nearby (Glaeser and Maré 2001; Wheaton and Lewis 2002; Combes, Duranton, and Gobillon 2008; Puga 2010).

CONTACT Thi Bich Tran bichtt@neu.edu.vn Faculty of Statistics, The National Economics University, Room 814, Building A1, 207, Giai Phong Street, Dong Tam, Hai Ba Trung, Hanoi, Vietnam

Studies investigating the impacts of agglomeration on urban earnings often quantify the wage premium into three different sources. These include worker skills or human endowment (Glaeser and Maré 2001; Combes, Duranton, and Gobillon 2008; De la Roca and Puga 2017), local non-human endowments across cities (Combes, Duranton, and Gobillon 2008; Duranton 2016), and agglomeration economies (Glaeser and Maré 2001; Combes, Duranton, and Gobillon 2008; De la Roca and Puga 2017). The decomposition of sources of the urban wage premium aims to isolate agglomeration impacts to evaluate whether cities are productive by generating positive agglomeration effects on firms and workers. Substantial empirical evidence for the modern or formal sector in developed nations (Glaeser and Maré 2001; Combes, Duranton, and Gobillon 2008; De la Roca and Puga 2017) and developing countries (Barufi, Haddad, and Nijkamp 2016) show that cities make workers more productive.

Under the presence of informality, agglomeration externalities in cities are still under debate. On the one hand, Duranton (2008) states that both formal and informal sectors generate positive agglomeration effects. On the other hand, Overman and Venables (2010) argue that the concentration of the informal sector in cities drives up urban costs and crowds out the formal sector. However, it is not sufficiently fast enough for these negative impacts to offset the positive effects of increasing city size. Their argument is strongly supported by Moreno-Monroy (2012).

Recent empirical evidence in South American developing countries is often inconclusive. For example, Duranton (2016) and García (2019) show that cities have stronger effects on the wages of informal workers than their formal counterparts in Colombia. However, Matano, Obaco, and Royuela (2020) point out that agglomeration economies benefit only formal employment and informal workers working in the formal sector, not those employed in the informal sector in Ecuador. Though providing insights into local scale externalities in the informal sector, none of these studies unveils agglomeration effects from informal sources.

With more than 60% of the world's employed population earning their livelihoods in the informal economy (Bonnet, Vanek, and Chen 2019) and cities continuing to cater to a large proportion of informal employment in developing countries (Moreno-Monroy 2012; Ghani and Kanbur 2013), quantifying sources of the urban wage premium of informal workers is crucial for development policies. If the earnings premium mainly comes from the ability bias and local endowments (including geographical features and location infrastructure), cities are not productive and urban development policies should focus on labor mobility and concern about the over-concentration of informality in cities (Overman and Venables 2010). On the other hand, if the wage premium is generated from agglomeration externalities (*i.e.* the geographical concentration of economic activities), cities still provide a safety net for informal workers (Moreno-Monroy 2012).

This study investigates whether cities generate positive agglomeration effects, separated by formal and informal sources, on the earnings of the informal workers in one of the East Asian developing countries - Vietnam. It also examines if the existence of informal workers in Vietnamese cities affects the wages of formal workers. The study of Vietnam is interesting because it is one of the typical cases of developing countries that are predominant by the informal sector and fast urbanization process. The country's informal workers account for nearly 60% of non-agricultural employment, of which approximately 40% are living in urban places (General Statistics Office of Vietnam 2018) that occupy only 10% of the country's territory.[1]

The remainder of the paper is organized as follows. Section 2 discusses sources of urban wage disparities and the interaction between the informal sector and agglomeration in cities. Section 3 describes data used to investigate the urban wage premium of informal workers and the effects of the concentration of informal employment on the wages of formal workers in cities. The distribution of informal employment and spatial wage disparities are presented in Section 4. Empirical models and the measurement of variables are shown in Section 5. Section 6 discusses the identification strategy, while Section 7 provides empirical results. Concluding remarks are presented in Section 8.

Urban wages and agglomeration economies

Sources of urban wage disparities

The last two decades have seen an increasing interest in the urban economic literature explaining why wages are higher in bigger cities. Reviews from the literature show three possibilities explaining this spatial wage disparity across cities. Firstly, the empirical analysis in advanced countries reveals that higher earnings in bigger cities result from a higher human endowment or workers' skills (Glaeser and Maré 2001; Combes, Duranton, and Gobillon 2008). Within this earning premium, De la Roca and Puga (2017) decompose it into two parts: one results from the spatial sorting of more productive workers, and the other comes from the dynamic benefits of learning by working in bigger cities. Their findings reveal that higher wages result from more valuable experiences accumulated in bigger cities rather than unobserved ability.

The second set of explanations proposes that wage differences across areas come from the local non-human endowment. The benefits of local endowment include geographical features and infrastructures such as airports and closeness to a navigable river or deep-sea harbor. These are assumed to benefit the productivity and wages of laborers (Combes, Duranton, and Gobillon 2008).

The third strand of explanations comes from the argument that firms and employees are more productive in large and dense urban areas than in other locations as locating nearby generates benefits (see Puga (2010) for a review).

These benefits are known as local scale externalities or agglomeration economies, resulting from the interaction between firms and individuals (Barufi, Haddad, and Nijkamp 2016). There are two types of agglomerations: localization and urbanization economies.

Localization has resulted from the concentration of similar economic activities or the concentration of firms and workers in the same industry. This localization is known as the Marshall, Arrow, Romer (MAR) economies in the dynamic context (Henderson 2003). The first advantage of localization is that knowledge and technologies are easily transferred between firms in the same industry through spatial proximity (Jaffe, Trajtenberg, and Henderson 1993), industrial linkages, and subcontracting opportunities (Lall, Shalizi, and Deichmann 2004). The second is that labor market search and matching are improved with a scale that enhances productivity (Kim 1989; He, Wei, and Pan 2007).

Urbanization comes from cities' size or the diversity of industrial activities. The former represents the benefit of greater market access (Glaeser and Maré 2001) when a firm locates in a dense area. In addition, the availability of a large labor pool in dense and diverse urban places allows for better matching between employers and employees and business partners (Puga 2010). The latter – so-called Chinitz-Jacobs urbanization in the dynamic context (Henderson 2003) - implies the benefits of knowledge transmission across sectors (Jacobs 1969; Porter 1990), such as the exchange of complementary knowledge (Barufi, Haddad, and Nijkamp 2016). It also reflects the benefit of better access to supporting services (Puga 2010).

As firms benefit from productivity advantages in bigger cities, they are willing to pay higher wages to their workers. Previous studies investigating spatial wage disparities of the formal sector in advanced countries show that wages are higher in bigger cities, and both urbanization and localization positively affect wages. For instance, Combes, Duranton, and Gobillon (2008) reveal that returns to agglomeration in France are mainly driven by urbanization. Similarly, Rosenthal and Strange (2008) and De la Roca and Puga (2017) point out that urbanization contributes to wage disparities in the US and Spain. Localization also generates the urban earnings premium in the US. Evidence of agglomeration effects on wage disparities is also found in the formal sector in developing countries. For example, Barufi, Haddad, and Nijkamp (2016) find that urbanization generates positive effects on wages of formal workers in Brazil while localization impacts depend on industrial sectors. Following Punzo, Castellano, and Bruno (2022), we create Table 1 that summaries sources of the urban earning premium and their expected relations with workers' earnings.

The interactions between the informal sector and agglomeration in cities

Investigations on urban earnings premium have assumed that cities generate benefits through agglomeration or geographical concentration of economic

Table 1. Sources of the urban wage premium.

Groups	Variables	Expectation	References
Workers' skills	- *Age; Age squared*	*U_shape effects*: Earnings increase with increasing age to a certain level of age. Beyond that level, the relationship becomes negative	• Combes, Duranton, and Gobillon (2008); • Duranton (2016); • García (2019); • Håkansson and Isacsson (2019); • Matano et al. (2020);
	Sex	Female are expected to earn lower than their counterparts.	• Duranton (2016); • García (2019); • Matano et al. (2020).
	Education	*Positive*: Higher education levels often accompany with higher earnings	• Glaeser and Maré (2001); • Barufi, Haddad, and Nijkamp (2016); • García (2019); • Håkansson and Isacsson (2019); • Matano et al. (2020).
	Experience	*Positive or no effect*: There was no clear pattern of experience effects on incomes	• Glaeser and Maré (2001); • De la Roca and Puga (2017).
	Job category	Earnings are different across job categories	• Barufi, Haddad, and Nijkamp (2016); • De la Roca and Puga (2017); • Matano et al. (2020).
Firms' characteristics	*Firm size and firm ownership*	Earnings are different across firm size and firm ownership	• Håkansson and Isacsson (2019).
Local non-human endownment	*Sea; Mountain; Lake; Heritage*	*Mixed effects*: Being closer to the sea promotes income; other endowments have ambiguous effects.	• Combes, Duranton, and Gobillon (2008) • Duranton (2016);
	Regional dummy	*Insignificant*: Earnings are not different among regions	• Duranton (2016); • Håkansson and Isacsson (2019); • Matano et al. (2020).
Agglomeration			

(*Continued*)

Table 1. (Continued).

Groups	Variables	Expectation	References
Urbanization	*Population and pop density; Population squared*	*Positive*: Earnings are usually higher in more populous areas. *U_shape effects*: Pop has non-linear effects on informal workers' earnings	• Barufi, Haddad, and Nijkamp (2016); • De la Roca and Puga (2017); • Matano et al. (2020). • Duranton (2016);
	Region employment density	*Positive*: Area employment has positive effects on earnings of formal and informal sector workers.	• Combes, Duranton, and Gobillon (2008); • García (2019). • Rosenthal and Strange (2008);
	Diversity	*Mixed effects*: There was no clear pattern of the diversity effects on wages	• Combes, Duranton, and Gobillon (2008); • Barufi, Haddad, and Nijkamp (2016).
Localization	*Specialization*	*Mixed effects*: Formal workers usually benefit from specialization but not their informal counterparts.	• Matano et al. (2020); • Barufi, Haddad, and Nijkamp (2016).
	Own industry employment density	*Mixed effects*: Own industry employment has either positive or insignificant effects on productivity.	• Lall, Shalizi, and Deichmann (2004)
Market access	*Density of population in neighboring areas*	*Positive*: Market potential promotes earnings	• Combes, Duranton, and Gobillon (2008); • Duranton (2016).

activities. However, studies express concerns about the over-concentration of firms and workers due to higher incomes in urban places. When the concentration is over-sized, costs might arise, and agglomeration economies generate negative impacts (known as congestion) on productivity and wages of the urban laborers (Arnott 1979; Fujita 1989). If it is the case, cities are considered not productive (Overman and Venables 2005, 2010). In this setting, the literature raises concerns that the existence of informality may reduce the benefits of city scale, especially in the formal sector, because informality pushes up costs in cities (Overman and Venables 2005, 2010; Moreno-Monroy 2012).

These concerns attract substantial explanations for the interaction between the informal sector and agglomeration in cities. As Mukim (2015) indicates, the benefits of agglomeration including greater access to specialized inputs, labor-market pooling, and knowledge and technology spillover (Ellison, Glaeser, and Kerr 2010) apply equally to both formal and informal firms. These are the reasons why informal workers tend to migrate to dense cities. Informal workers not only concentrate in cities; empirical evidence also shows they tend to co-agglomerate with formal workers (Mukim 2015; Tran 2015; Tran and La 2018). Nevertheless, the opinions on the co-agglomeration of formal and informal sectors are divided. Some argue that agglomeration can be generated in informal enterprises when these firms participate in the value chain through backward and forward linkages or sub-contracts with formal enterprises (Overman and Venables 2005; Moreno-Monroy 2012). Others point out the possibility of isolated informal firms in cities due to their inability (Moreno-Monroy 2012). Consequently, agglomerations are not generated on the earnings of informal workers. Even worst, the concentration of the informal sector in cities drives up urban costs and crowds out the formal sector (Overman and Venables 2010; Moreno-Monroy 2012).

Empirical evidence on returns to agglomeration under the presence of informality in South American developing countries reveals that agglomeration matter also. However, the results are inconclusive. For instance, Duranton (2016) finds that urbanization, measured by city population, has stronger effects on wages of informal workers (defined as those without written labor contracts) than formal laborers in Colombia. Similarly, García (2019) shows that the effects of urbanization, measured by employment density, are more significant for informal sector workers than formal sector workers in Colombia. By separating agglomeration into urbanization (indicated by population density) and localization (measured by the specialization of economic activities), Matano, Obaco, and Royuela (2020) reveals that while both urbanization and localization generate the urban wage premium for formal workers, they have positive impacts on only informal workers working in the formal sector, not those who work in the informal sector.

In this study, we separate localization into informal and formal sources and evaluate the impacts of each type of localization on the earnings of informal

workers. We also evaluate whether informal agglomeration affects the wages of formal employment. Our results provide further empirical evidence to the debate in the literature on the interaction between the informal sector and agglomeration in cities.

Data

This paper employs quarterly cross-sectional data from the Vietnam Labor Force Survey (LFS) from 2013 to 2020 to examine whether cities generate positive agglomeration impacts on the earnings of informal workers and whether informal agglomeration affects the wages of formal workers. This survey has been carried out annually by the General Statistics Office (GSO) of Vietnam since 2007 and provides comprehensive information on informal employment in Vietnam. However, as surveys before 2013 provide data from only one point in time, these data are excluded from the analysis.

The GSO applied the 17th International Conference of Labor Statisticians definition to measure informal workers as those working in the informal sector and those working in the formal sector but not eligible for social insurance. The informal sector includes all unincorporated enterprises that do not register and engage in non-agricultural activities. We limit our sample to wage earners because the information on the income of household businesses is not available. Although this limitation excludes business owners in the informal sector, [2] our sample is expected to be more homogenous.

The LFS applies the multi-stage sampling strategy. In the first stage, small areas (primary sampling units – PSUs- which contain around 100 households, classified into urban and rural strata for each province/city) are selected using the method of probability proportional to size. In the second stage, households in selected PSUs are sampled using the systematic random sampling method. This sampling strategy allows this survey to be representative at the provincial level combined with urban and rural areas. The survey was implemented four rounds per year.

In this study, we define the employment location at the district level. The hierarchical administrative division of Vietnam starts from provinces under the management of the central government. Each province has smaller administrative units called districts, and the districts break down into communes. The classification of provinces is based on their historical features and development speeds. With this classification, Vietnam has five centrally managed cities. Hanoi in the North and Ho Chi Minh in the South are the two largest provinces considered metropolitan areas. Three other centrally managed provinces include Hai Phong in the North, Da Nang in the Center, and Can Tho in the South. These three provinces are endowed with favorable navigable rivers and deep-sea harbors. It is worth noting that although these three central provinces have developed along with the history of Vietnam, their development speeds

may be less than some emerging industrial provinces, including Quang Ninh and Bac Ninh in the North and Dong Nai and Binh Duong in the South.

As the literature points out, the choice of geographical units makes no difference in estimating agglomeration effects. For instance, Briant, Combes, and Lafourcade (2010) compared the results of several exercises in spatial economics with different choices of geographical units using French data. They conclude that the shape of units makes no difference in estimating agglomeration effects. Similarly, Howard, Newman, and Tarp (2016) found consistent results when analyzing the driving forces of agglomeration in the formal sector in Vietnam with different uses of the geographical units, including provinces, districts, and communes.

To investigate whether wages are higher in bigger cities, we select only the data set in the urban setting. In the LFS, urban and rural areas are defined at the grass-root level of the administrative division, *i.e*. the commune level. In Vietnam, it is worth noting that most districts in five centrally managed cities are urban ones. Other provinces have one urban district that serves as the provincial capital and one to four other districts considered commercial centers and classified as urban areas. The rest of the districts in the province is rural areas. However, each district has one commune that serves as the district capital, and one to two other communes are considered urban ones. The definition of rural and urban areas at the commune level follows Glaeser and Maré (2001) suggestions that small towns, which are urban places, should not be ignored when analyzing the impacts of agglomerations. Although we define the employment location at the district level in our paper, we can separate urban and rural communes in each district. All urban communes in each district (*so-called* cities) are kept in our sample.

In this paper, two different cities are distinguished: small towns with urban populations of up to 30 thousand people and the rest are large cities. Data on districts' urban populations are taken from the statistical yearbook of Vietnam from 2013 to 2020. Table 2 displays the distribution of cities across provinces. As shown in this Table, large cities are often located in metropolitan areas and centrally managed cities with more favorable local endowments. On the other hand, small cities are more pronounced in the four emerging industrial provinces and other provinces of the country (Table 2).

Since the paper focuses on the disparity of earnings across cities and whether cities generate agglomeration effects on wages of workers, our final sample includes all waged workers aged from 15 years old[3] in non-agricultural activities. The annual average sample includes 469,161 formal workers and 260,079 informal workers, representing above 6.2 million formal workers and 3.2 million informal workers in the urban population setting, respectively (Table 3). The population of formal and informal workers are employed in around 156,946 and 1,864,325 formal and informal firms in urban Vietnam.[4] As shown in Table 3, formal and informal workers concentrate more in large cities.

Table 2. District distribution across provinces.

	2013		2016		2020	
Province types	Small	Large	Small	Large	Small	Large
Metropolitan cities	19	33	19	34	17	36
	(36.5)	(63.5)	(35.9)	(64.1)	(32.1)	(67.9)
Centrally managed cities	12	18	12	18	11	18
	(40.0)	(60.0)	(40.0)	(60.0)	(37.9)	(62.1)
Industrial provinces	24	15	24	15	20	17
	(61.5)	(38.5)	(61.5)	(38.5)	(54.1)	(45.9)
Others	455	84	449	115	445	126
	(84.4)	(15.6)	(79.6)	(20.4)	(77.9)	(22.1)
Total	510	150	504	182	493	197
	(77.3)	(22.7)	(73.5)	(26.5)	(71.5)	(28.5)
	660		686		690	

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.
Notes: Figures in parentheses are percentages
Small cities have urban populations up to 30 thousand people

Table 3. Socio-demographic characteristics of workers in urban areas.

	Small cities - Formal	Small cities - Informal	Large cities - Formal	Large cities - Informal	All – Formal	All – Informal
Number of workers (000 people)	922.3	581.2	5365.0	2635.8	6287.3	3217.0
Hourly wage (000 VND)	38.1	22.3	43.9	27.0	43.1	26.1
Characteristics						
Age(years)	37.1	37.3	35.9	37.0	36.0	37.1
Male(%)	47.4	68.1	50.7	62.8	50.2	63.8
Married(%)	82.3	67.8	74.2	61.8	75.4	62.9
Education(%)						
- *Not finished primary school*	0.8	11.9	1.6	9.6	1.5	10.0
- *Primary school*	3.7	25.1	6.3	22.6	5.9	23.1
- *Lower secondary school*	8.5	32.2	13.6	28.5	12.8	29.1
- *Higher secondary school*	13.0	20.3	19.3	24.1	18.4	23.4
- *College*	27.1	7.3	15.7	8.1	17.3	8.0
- *Bachelor or above*	46.9	3.2	43.5	7.1	44.0	6.4
Experience(%)						
- *Below one year of experience*	3.5	13.5	4.2	15.5	4.1	15.1
- *From one years of experience*	96.5	86.5	95.8	84.5	95.9	84.9
Ownership(%)						
- *Household business*	0.8	72.6	0.8	64.4	0.8	65.9
- *Private firm*	17.7	19.0	42.4	28.1	38.9	26.5
- *State-owned firm*	70.9	7.1	38.7	5.3	43.3	5.6
- *Foreign-invested firm*	10.7	1.2	18.0	2.2	17.0	2.0
Observations	120,267.0	69,426.0	348,894.0	190,653.0	469,161.0	260,079.0

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.
Notes: Population weights are used. The sample excludes non-wage earners and those working in the agricultural sector.
Hourly wage is measured at the 2020 value.

Table 3 reveals that wages and the education level of formal workers are much higher than those of their informal counterparts in small and large cities. Regarding workers' experience, we use a cut-point of one year to distinguish

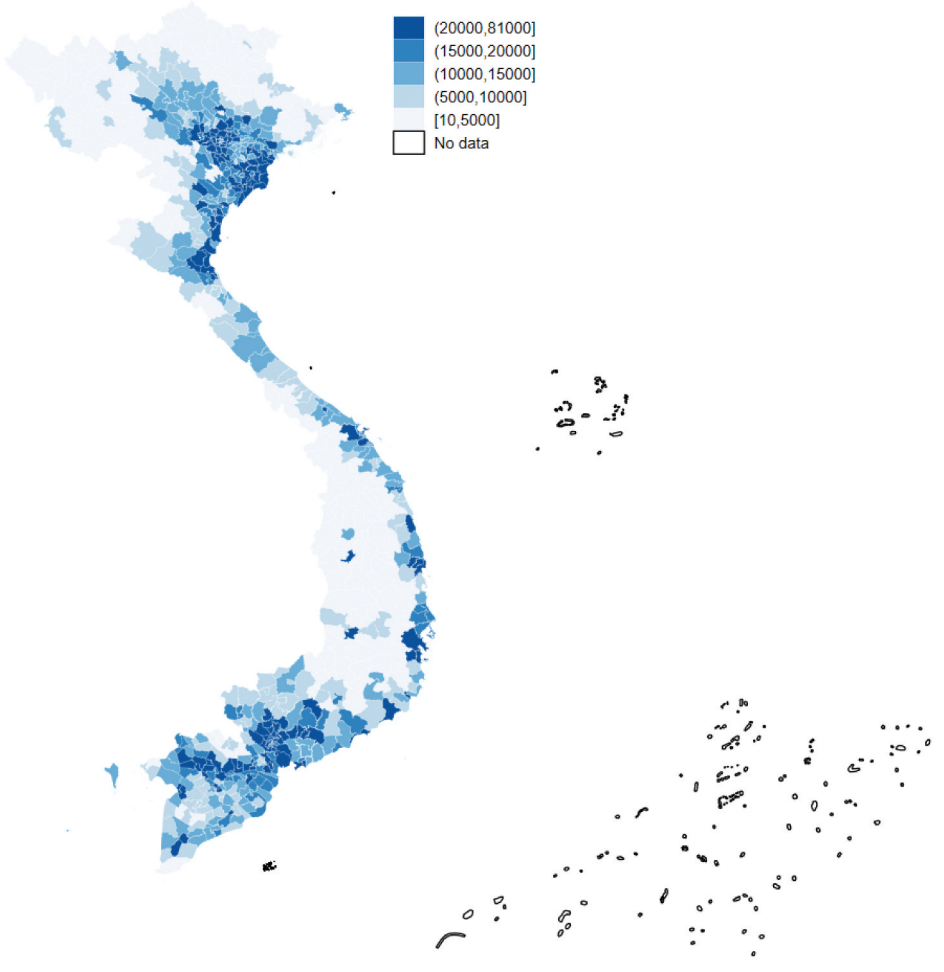

Figure 1. Density of informal workers across provinces of Vietnam. Source: Authors' calculation using the 2013-2020 Labour Force Survey.

between season and long-term workers. According to Table 3, informal workers concentrate in the household business sector. A few of them work in state-owned firms and foreign-invested enterprises.

Informal employment and wage disparities in urban areas of Vietnam

Although urban areas account for only 10% of the country's territory, 40% of the informal employment is concentrated in urban places of Vietnam (General Statistics Office of Vietnam 2018). The previous report from GSO shows that the density of informal workers in urban areas is approximately 1.5 times higher than in rural places from 2007 to 2013 (General Statistics Office of Vietnam 2014). Moreover, informal workers tend to cluster in the most highly developed provinces, as shown in Figure 1. Moreover, the co-agglomeration of informal

and formal enterprises is much higher in urban than in rural districts in almost all the provinces of Vietnam (Tran and La 2018).

In Figure 1, the darker the color, the more concentration of informal employment in these locations. As can be seen from this figure, informal workers concentrate in the country's most developed regions, including the Red River Delta (in the North), the Mekong River Delta (in the South), and coastal areas. Within these areas, informal employment concentrates in Metropolitan areas, including Hanoi in the North and Ho Chi Minh City in the South; three other centrally-managed cities, including Hai Phong in the North; Da Nang in the Center; and Can Tho in the South; and the four emerging industrial provinces (Quang Ninh and Bac Ninh in the North and Dong Nai and Binh Duong in the South).

The disproportion of informal employment in urban areas of Vietnam matches perfectly with the Harris and Todaro model framework, which shows that workers migrate from rural to urban areas because they expect to earn higher in urban places without worrying about the real level of their income. They, however, often end up their job-searching route to the low-income sectors in the destination. Using data from the 2013-2020 LFS, we calculate the mean wage gaps between formal and informal workers for urban and rural areas separately. We then draw a scatterplot of the wage differences in urban and rural areas in Figure 2. In this figure, the 45-degree line reflects the same

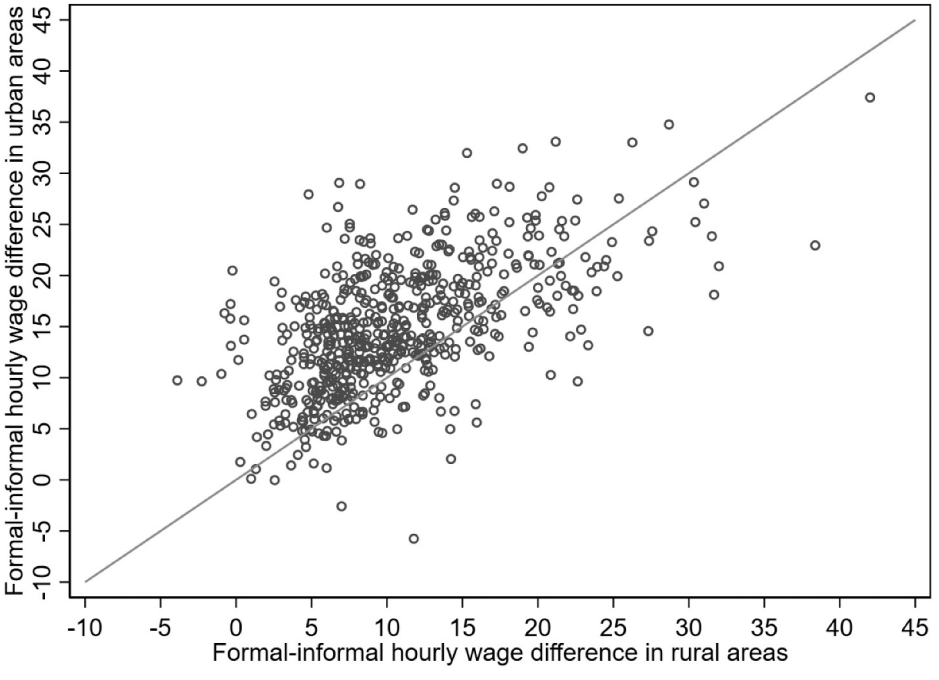

Figure 2. Nominal wage gaps between formal and informal workers in urban and rural areas. Source: Authors' calculation using the 2013-2020 Labour Force Survey.

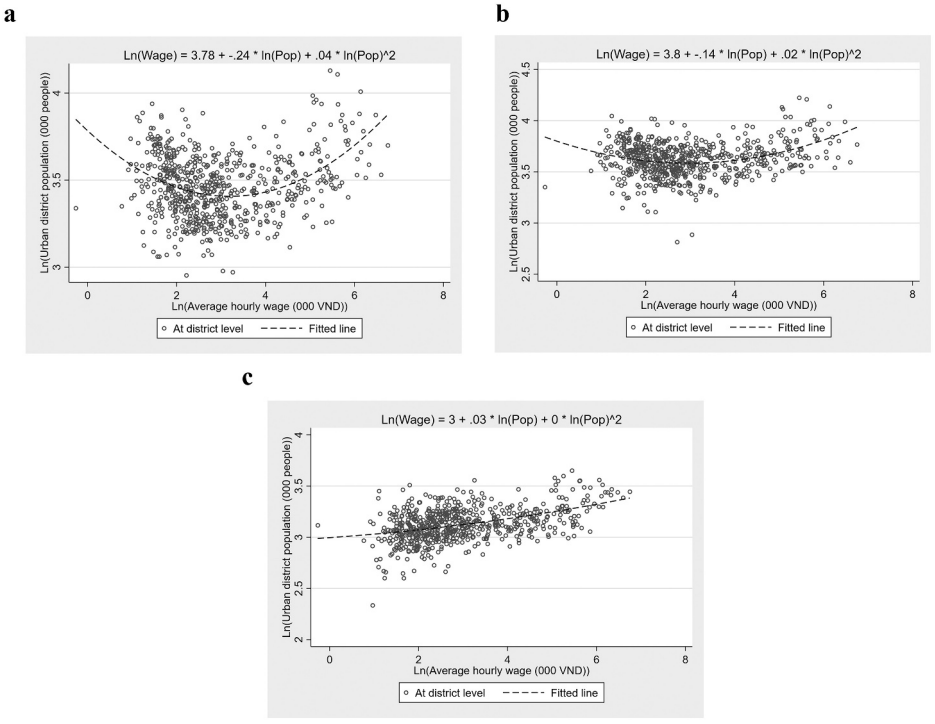

Figure 3. (a) Relationship between mean wages and population across urban districts. (b) Relationship between mean wages of formal workers and population across urban districts. (c) Relationship between mean wages of informal workers and population across urban districts. Source: Authors' calculation using the 2013-2020 Labour Force Survey.

level of formal-informal wage gaps in urban and rural places. The points above the line reveal higher wage gaps between formal and informal workers in urban than rural areas. As Figure 2 shows, formal-informal wage differences in urban areas are higher than in rural areas.

However, Nguyen and Minh (2016) show that urban-in migrants receive substantially lower wages than non-migrants and often end up their job-searching route in the informal sector in the major cities in Vietnam. After controlling for observed demographic characteristics, the wage gap disappears. They conclude that the main difference in observed wages between migrants and non-migrants is explained by differences in age and education between migrants and non-migrants.

To explore whether earnings are higher in bigger cities, we regress the logarithm of aggregated average earnings on the logarithm of the population at the district level and present the results in Figure 3(a). We also run separate functions of the mean wages on the district's population for informal and formal workers and display estimated results in Figure 3(b,c). Interestingly, the shape of the wages of formal workers dominates the earning pattern in urban districts. The wage pattern of formal workers (Figure 3(b)) follows the U-shape, which is

similar to that of the earnings pattern of the entire employment (Figure 3(a)). The wage pattern in Figure 3(a,b) indicates that although wages of employment, in general, and formal workers, in particular, are higher in bigger cities, the growth rate of salaries when the city is below a certain threshold diminishes when the city size increases. It then accelerates when the city size goes beyond this threshold level. Nevertheless, this pattern is unclear for informal workers (Figure 3(c)).

Wages are higher in bigger cities. The findings imply that rural migrants continue to come and search for jobs in major cities and urban places in Vietnam. Regarding policy making and urban planning, it is important to know the sources of higher wages in urban places. If higher wages mostly come from worker skills and local non-human endowments, policy designs should focus on investments to create jobs in rural areas. On the other hand, if the economy of scale generates benefits in urban places and cities still provide a safety net for the poor, it is not to worry about the flow of urban-in migrants. The following sections investigate whether cities generate positive agglomeration impacts on workers' wages. Confounding factors, including workers' skills and local endowment, which may mislead the impact of scale economies, will be isolated to unveil agglomeration's "*de-facto*" effects.

Model and variable measurement

A unified framework, which regresses workers' skills, local endowment, and agglomeration variables on wages in one step, imposes a stringent data requirement (Combes, Duranton, and Gobillon 2008). We, therefore, follow Combes, Duranton, and Gobillon (2008) to apply the two-step estimation strategy to decompose earning differences across areas into differences in individual skills and differentials in labor productivity caused by local endowments and agglomeration economies. In the first step, wages are regressed on a set of individuals' observable characteristics and the area-sector-time fixed effects. This estimation is reflected in the following equation:

$$\log(\omega_{i,t}) = \alpha + \beta X_{i,t} + AST_{a(i),k(i),t} + \varepsilon_{i,t} \quad (1)$$

where $X_{i,t}$ is a set of workers' observable characteristics, $AST_{a(i),k(i),t}$ is the area-sector-time dummies, and $\varepsilon_{i,t}$ is the error term that is assumed to be *i.i.d* across workers. Wages in (1), $\omega_{i,t}$, are measured as hourly earnings of worker *i* at time *t*. Observable individual skills include gender, age, marital status, education, and experience. We also control for types of firms where workers are employed, including household businesses, private firms, state-owned enterprises, and foreign-invested companies. The inclusion of individual skills and firm characteristics is motivated by the literature shown in Table 1. Descriptive statistics of these variables by formal and informal workers across cities are presented in Table 3.

From the first step, $\widehat{AST}_{a(i),k(i),t}$, the estimated fixed effect for each area-sector at time t, is retrieved. $\widehat{AST}_{a(i),k(i),t}$ is the urban earnings premium left after excluding the wage disparity resulting from workers' skills. To evaluate whether this earnings premium comes from local non-human endowment or agglomeration externalities, we calculate the weighted average of these estimated fixed effects over time, using the weight as the number of observations in each time and area-sector cell. In the second step, these weighted time-average fixed effect, $\overline{\widehat{AST}_{a,k}}$, is run over a set of variables representing agglomeration economies and local endowment. The estimated equation is written below:

$$\overline{\widehat{AST}_{a,k}} = \delta + \theta Urban_a + \tau Loc_{a,k} + C_a + \vartheta_{a,k} \tag{2}$$

where $Urban_a$ and $Loc_{a,k}$ reflect urbanization and localization economies, and C_a represents differences in local non-human endowments. In this paper, we use the distance to the coast to reflect the natural advantages of each area.

As discussed previously, one source of productivity gains that induce firms to pay higher for their workers emanates from agglomeration externalities. This paper investigates the impacts of urbanization and localization on workers' earnings. In the urban economics literature, urbanization arises from either city size or industries' diversity. The former can be measured in terms of population or employment, reflecting the benefits of access to markets and the availability of a large labor pool that a firm could have if located in dense urban areas. The latter measures the industry mix or the diversity of industrial activities. This measurement reflects the benefits of technology spillover and knowledge transfer across industries. In the paper, we measure urbanization as the logarithm of the district's urban population. This is because the majority of informal firms in Vietnam are weak. Their clustering is not driven by technology transfer (Tran and La 2018) but by the benefits of market access and labor matching in dense areas.

Lall, Shalizi, and Deichmann (2004) summarize three ways of measuring localization. These include own industry employment in the region, own industry establishment, and an index of concentration that reflects the disproportionately high concentration of the industry in the region compared to the nation. As in the case of urbanization, we measure localization as the size of the district's industry employment to capture better the benefit of labor matching in the informal sector. Formal localization and informal localization are separated by formal and informal employment of each industry in the district.

Identification strategy

Studies on spatial wage disparities encounter two problems: endogeneity and omitted ability. First, the urban economics literature identifies agglomeration as

an endogenous variable because higher output per worker might not be a consequence of higher employment density but its cause (Moomaw 1981; Brulhart 1998). If a location has conditions favoring greater productivity and thus wages, it will attract more firms and workers and become larger as a result. As such, there will be a correlation across geographical units in the error term in estimated equations.

A common way to deal with the endogeneity of agglomeration in the literature is to use the IV method (Håkansson and Isacsson 2019). Ciccone and Hall (1996) are the first persons who use instruments such as historical population and characteristics of locations that are assumed to be exogenous for the employment density, such as the distance from the location to a specific seaboard, when analyzing agglomeration effects on spatial disparities of productivity in the US. This paper uses the district population taken from the Population Census in 1989, the district proportion of the urban population in 1989, and the market potential calculated using the 1989 population as IVs.[5]

As Ciccone and Hall (1996) indicated, the choice of instruments rests on the hypothesis that the population agglomeration in the past is not related to employees' current productivity and wages. This hypothesis is more likely to hold for very long lags (Combes, Duranton, and Gobillon 2008). Vietnam's district population in 1989 meets this hypothesis because it nearly reflects the original agglomerations of Vietnam, thanks to governmental migration policies that restricted urban-in migration before the mid-1990s. Since its reunification in 1975, Vietnam's most visible population policy has been population redistribution and rural resettlement to reduce population pressure in the densely populated provinces and urban centers (Decision 95-CP on 27/3/1980). The mechanism of the centrally planned economy allowed the government to regulate the migration flow. However, due to a lack of resources, this migration plan was not successful, and around 50% of migrants came back to their original places soon after they arrived in the new resettlement areas (Dang, Goldstein, and McNally 1997). Until 1997, regulations on industrial zones created a second wave of massive spontaneous migration (General Statistics Office of Vietnam 2011b, 2011a). Thus, the district population before 1990 reflects the preferred location of workers a long time in the past and is not the result of high earnings today.

As indicated in the literature, better access to the market induces firms to agglomerate as firms are more profitable when located near a large mass of consumers (Overman and Venables 2010). To calculate the market potential index long time in the past, we apply the original accessibility indicator for tradable goods proposed by Hansen (1959) with reasonable adjustments to match the setting of the informal sector in Vietnam. Market access, as proposed by Hansen, is written as follows:

Table 4. Descriptive statistics of IVs and time-average for agglomeration variables.

Variables	Mean	SD	Min	Max
Urban population (000 pp)	49.4	95.9	0.8	854.2
Localization (000 pp)	15.5	31.6	0.4	301.5
- Informal localization (000 pp)	5.3	9.2	0.0	80.8
- Formal localization (000 pp)	10.2	23.3	0.1	239.4
Population in 1989 (000 pp)	92.9	59.4	5.9	337.1
Urbanization rate (%) in 1989	18.3	27.2	0.0	100.0
Market access in 1989 (000 pp)	107.4	135.3	0.0	882.2
Distance to sea (km)	76.2	83.5	0.0	438.6
Observations	646.0			

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.
Notes: Non-wage earners and those working in the agricultural sector are excluded.

$$MA_a = \sum\nolimits_m \frac{S_m}{d_{a-m}^b} \tag{3}$$

where MA_a is the accessibility indicator estimated for district MA_a, a is a size indicator of destination a measured by the population of that district, a is the distance from district MA_a to its destination a, and m is a decay factor describing how increasing distance reduces the expected level of interaction. In the original model proposed by Hansen, MA_a is calculated by connecting region MA_a to every region in the studied area. In this paper, we limit neighboring districts to a 50 km radius. This is because transportation means in Vietnam during the 1980s relied heavily upon non-machinery vehicles. We use Euclidean distance to measure the distance between two points. In addition, we follow Mukim (2015) to set b equal to 1 in this study. The description of IVs and their instrumented agglomeration variables are presented in Table 4.

Second, higher earnings in bigger cities may result from higher-skilled workers (including their observed and unobserved ability) working in these areas. The best identification strategy to deal with this issue is to use worker-fixed effects with panel data as studies in developed countries (Glaeser and Maré 2001; Combes, Duranton, and Gobillon 2008). However, panel data, especially for informal workers, are not available in developing countries. Thus, we follow the literature that investigates returns to agglomeration of the informal sector in developing countries (Duranton 2016; García 2019; Matano, Obaco, and Royuela 2020) to add observable skills of workers. Moreover, we allowed for intra-correlation within clusters. As suggested by Cameron and Miller (2015), when observations are grouped into clusters (e.g. more productive workers located in some clusters), model errors for individuals in the same region may be correlated, and failure to control for within-cluster error correlation can lead to very misleadingly small standard errors, large *t*-statistics and low p-values. Controlling observable worker characteristics and adjusting standard errors in cross-sectional regression can attenuate endogeneity and selection bias problems when estimating the effects of urbanization on workers' wages. In this paper, standard errors are clustered at the district or city level.

Empirical results

Section 4 reveals the U-shape correlation between the average wage of workers and the urban population when analyzing at the district level. As the analytical framework in Section 5 shows, wage differences across areas can result from differences in individual skills, local endowments, and agglomeration economies. This section investigates the impacts of urbanization and localization on workers' wages in Vietnam, controlling for individual skills and local non-human endowment.

Agglomeration effects on wages

Baseline results

We start our analysis by estimating Equations (1) and (2) using the OLS method. The results are documented in Table 5. Column (1) in Table 5 represents results from the first step estimation. As seen from this column, the education level is the main driving factor of wage differentials amongst workers. In addition, workers in state-owned firms and foreign-invested enterprises earn much higher than their counterparts working in the household business sector (Column 1, Table 5).

Columns (2) and (3) show OLS estimates in the second step. To compare agglomeration effects under the presence of informality in Vietnam with those in other developing countries, we introduce only linear effects of urbanization and calculate localization for the entire economy (overall localization) in column (2). Estimated results from this column reveal that the elasticity of wages with respect to city population is 5.5%. Using the same measure of urbanization, Duranton (2016) reports similar linear urbanization effects on workers' wages in Colombia. Finally, in column (3), we add the quadratic term of urbanization. The significant coefficient of the squared logarithm of the district's urban population confirms the non-linear correlation between districts' urban population and wages shown in Figure 3(a).

The elasticity of wages with respect to sector employment (localization) is about 1%. Impacts of local non-human endowment (proxied by the distance to the coast) are positive and consistent in all columns in Table 5. The result indicates that the closer the city's sea is, the lower the earnings premium is. This result is somewhat strange. However, it is worth noting that the OLS estimates may be biased because of the endogeneity issue as described in Section 6.

IVs estimation

Figure 3(a) suggests the quadratic effects of the district's urban population on wages, and the OLS estimates in Table 5 confirm this non-linear correlation.[6] We,

Table 5. Agglomeration effects on wages of all workers in urban areas using OLS regressions.

	(1) 1st step	(2) 2nd step	(3) 2nd step
Ln(Urban population)		0.055***	0.025***
		(0.002)	(0.009)
Ln(Urban population) squared			0.004***
			(0.001)
Ln(Localization)		0.011***	0.010***
		(0.002)	(0.002)
Ln(Distance to sea (km))		0.011***	0.010***
		(0.001)	(0.001)
Age	0.048***		
	(0.001)		
Squared age	−0.001***		
	(0.000)		
Male	0.129***		
	(0.001)		
Married	0.046***		
	(0.002)		
Education			
- *Primary school*	0.072***		
	(0.003)		
- *Lower secondary school*	0.117***		
	(0.003)		
- *Higher secondary school*	0.210***		
	(0.003)		
- *College*	0.389***		
	(0.004)		
- *Bachelor or above*	0.619***		
	(0.004)		
Experience			
- *From one years of experience*	*0.170***		
	(0.003)		
Firm ownership			
- *Private firm*	0.109***		
	(0.003)		
- *State-owned firm*	0.169***		
	(0.003)		
- *Foreign-invested firm*	0.222***		
	(0.004)		
Constant	1.603***	−0.352***	−0.296***
	(0.012)	(0.012)	(0.021)
Observations	729,240	16,650	16,650
F-statistics	6342.2	413.6	313.1
Prob > F	0.000	0.000	0.000

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.
Notes: Standard errors in parentheses.
Categories of 'Not finished primary school', 'Below one year of experience' and 'Household business' are ref. groups.
Robust standard errors are applied. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Observations in Columns 2&3 are at the area-sector level.

therefore, add both linear and quadratic terms of the logarithm of districts' urban population in the IV estimation in the second step and report only the results from this step in Table 6. In this Table, we instrument three endogenous agglomeration variables, including the logarithm of the district's urban population, the squared logarithm of the district's urban population, and localization by three IVs. The IVs include the district population in 1989, the share of districts'

Table 6. Agglomeration effects on wages of all workers in urban areas using IV regressions.

	(1) 2nd step	(2) 2nd w/o individual characteristics in 1st step	(3) 2nd step	(4) 2nd step w/o individual characteristics in 1st step	(5) 2nd step	(6) 2nd step w/o individual characteristics in 1st step
Ln(Urban population)	−0.854***	−1.559***	−0.568**	−1.314***	−0.953***	−1.643***
	(0.222)	(0.325)	(0.277)	(0.389)	(0.231)	(0.343)
Ln(Urban population) squared	0.119***	0.219***	0.079**	0.184***	0.130***	0.228***
	(0.030)	(0.044)	(0.038)	(0.054)	(0.031)	(0.046)
Ln(Localization)	0.028*	0.024				
	(0.015)	(0.022)				
Ln(Formal localization)			0.028**	0.024		
			(0.013)	(0.020)		
Ln(Informal localization)					0.051*	0.043
					(0.029)	(0.043)
Ln(Distance to sea (km))	−0.016**	−0.022**	−0.011	−0.018*	−0.016**	−0.022**
	(0.007)	(0.010)	(0.008)	(0.011)	(0.007)	(0.011)
Constant	1.129***	2.147***	0.719	1.796***	1.261***	2.260***
	(0.400)	(0.581)	(0.487)	(0.685)	(0.404)	(0.594)
Observations	16,650	16,650	16,650	16,650	16,650	16,650
IV tests						
Endogeneity						
- Wooldridge 95 robust score test (*p*-value)	0.000	0.000	0.000	0.000	0.000	0.000
- Wooldridge 95 robust regression-based test (*p*-value)	0.000	0.000	0.000	0.000	0.000	0.000
Weak instrument						
Adj. Partial R2						
- Ln(Population (000 people)	0.55	0.55	0.55	0.55	0.55	0.55
- Ln(Population (000 people)) squared	0.57	0.57	0.57	0.57	0.57	0.57
- Ln(Localization)	0.28	0.28	0.16	0.16	0.10	0.10
F-test first stage						
- Ln(Population (000 people)	10,151	10,151	10,151	10,151	10,151	10,151
- Ln(Population (000 people)) squared	8323	8323	8323	8323	8323	8323
- Ln(Localization)	2129	2129				
- Ln(Formal localization)			1512	1512		
- Ln(Informal localization)					584	584

Source:Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.
Notes: Standard errors in parentheses.
Ln(Urban population), Ln(Urban population) squared, and the logarithm of localisation are endogenous variables. IVs include ln(Population (000 people) in 1989), the proportion of urban population in 1989, ln(market access (000 people) in 1989).
Robust 2SLS IV regression is applied and standard errors are clustered by cities. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Observations are at the area-sector level.

urban population in 1989, and the market potential calculated using the 1989 population. As we have the exact identification, we implement only the endogeneity and weak instrument tests. The endogeneity test at the end of Table 6 shows that the OLS estimator is biased and inconsistent. Furthermore, the partial R^2 and *F-tests* from the first stage regressions reveal all IVs are relevant (Table 6). Hence, the validity of the instruments is vigorously tested and justified.

In addition to urbanization and local endowment, Table 6 estimates the effects of overall localization (columns 1&2) and each type of localization calculated from formal (columns 3&4) and informal (columns 5&6) sources on the urban earnings premium. For each specification, we compare estimates of agglomeration and local endowment with and without controlling workers' observable characteristics in the first stage.

The significance of both linear and quadratic terms of urbanization in all specifications in Table 6 confirms the correlation pattern in Figure 3(a–c), implying that urbanization effects on wages become stronger when cities go beyond a certain threshold. Heterogeneous effects of agglomeration across cities are investigated later in Section 7.2. Impacts of urbanization on wages are higher in specifications that do not control individual characteristics in the first step only when the derivate of wages on urbanization become positive, at the cutoff population level of 36 thousand people. In particular, at the mean level of urban population documented in Table 4, the elasticity of wages with respect to city population estimated in column (1), which controls for individual characteristics in the first step, reduces about half when being compared with column (2) that do not control individual characteristics in the first step (from 15% to 7%).[7] This result reflects evidence of sorting on workers' skills. Duranton (2016) and Matano, Obaco, and Royuela (2020) also find that sorting matters in some South American developing countries.

It should be noted that localization benefits workers only in specifications that control for individual characteristics. Results from Table 6 show that not only the overall localization (column 1) but also each type of localization, including formal (column 3) and informal (column 5) localization, generate the earnings premium in cities. Compared to the OLS estimator in Table 5, the coefficient of the variable "*distance to the sea*" is now negative, implying the closer the city is to the sea, the higher the wage premium is. The result helps explain why the informality density is higher in coastal provinces, as shown in Figure 1.

Agglomeration and informality

Our first interest is to examine whether cities generate agglomeration effects on the earnings of informal workers in Vietnam. To retrieve the fixed effects for all informal workers, informal sector workers, and informal employment in the formal sector, we follow Matano, Obaco, and Royuela (2020) to add an

interaction term between the area-sector-time fixed effects and the dummy for each type of informal worker in (1) in the first step. Table 7 documents the estimated results from IVs second step for all informal workers (columns 1-3), informal sector workers (columns 4-6), and informal employment in the formal sector (columns 7-9).

As shown in Table 7, informal workers benefit from urbanization but not localization (columns 1-3). Interestingly, neither urbanization nor localization generates the earnings premium for informal sector workers (columns 4-6). On the other hand, informal workers in the formal sector benefit from positive urbanization effects (column 9) when cities go beyond a certain threshold and formal localization (column 8). Specifically, the elasticity of formal localization on wages of informal employees in the formal sector is 5% (column 8, Table 7). Matano, Obaco, and Royuela (2020) also find that spatial externalities generate positive effects on wages of informal employment in the formal sector but have no impact on the earnings premium of the informal sector workers in Ecuador.

We now turn on to the second interest, which explores the impacts of informal localization on the wages of formal employees. Similar to the case of informal workers, we interact the area-sector-time fixed effects with the dummy of formal workers in the first step to estimate the fixed effects for formal workers. Estimated results of IVs second step for formal workers are presented in Table 8. As in the case of informal sector workers, formal workers benefit from neither agglomeration nor local endowment (Table 8). These results may come from the heterogeneity of agglomeration impacts explored below.

Heterogeneity of agglomeration effects

Previous results suggest different impacts of agglomeration across cities. We, therefore, split the sample into two sub-samples by city size to carry out the analysis of agglomeration effects on wages of workers. The first sub-sample is in small cities with urban populations up to 30 thousand people, and the second sub-sample is in large cities having urban populations greater than 30 thousand people. Results from the IVs second step for small and large cities are presented in Tables 9 and 10, respectively.

In small cities, results from Table 9 show that formal workers benefit from both urbanization and localization, including informal localization (columns 4-6). On the other hand, only informal workers in the formal sector enjoy benefits from urbanization (columns 13-15). Furthermore, localization generates negative impacts on the wages of informal workers (columns 7-9). The results imply that the concentration of the informal sector is already over-sized in small cities. Thus, costs arising from this over-concentration of workers cancel out the benefits of locating nearby (Overman and Venables 2005, 2010; Moreno-Monroy 2012).

Table 7. Agglomeration effects on wages of informal workers in urban areas using IV regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All informal workers			Informal workers in the informal sectors			Informal workers in the formal sectors		
Ln(Urban population)	−1.372**	1.532	−1.068**	−1.383	−0.093	−0.713	−3.095	0.170	−1.822**
	(0.620)	(1.793)	(0.420)	(1.656)	(0.255)	(0.574)	(2.459)	(0.528)	(0.823)
Ln(Urban pop) squared	0.193**	−0.214	0.151***	0.196	0.013	0.101	0.428	−0.024	0.252**
	(0.085)	(0.251)	(0.057)	(0.231)	(0.036)	(0.078)	(0.337)	(0.073)	(0.111)
Ln(Localization)	−0.048			−0.114			−0.134		
	(0.033)			(0.164)			(0.139)		
Ln(Formal localization)		0.094			0.019			0.049**	
		(0.075)			(0.021)			(0.022)	
Ln(Informal localization)			−0.035			−0.051			−0.066
			(0.021)			(0.058)			(0.045)
Ln(Distance to sea (km))	−0.035*	0.042	−0.028**	−0.036	−0.002	−0.019	−0.086	0.009	−0.050*
	(0.019)	(0.046)	(0.013)	(0.046)	(0.006)	(0.017)	(0.075)	(0.014)	(0.027)
Constant	2.360**	−2.898	1.775**	2.801	0.022	1.338	5.739	−0.559	3.215**
	(1.187)	(3.118)	(0.783)	(3.647)	(0.444)	(1.263)	(4.871)	(0.926)	(1.601)
Observations	13,814	13,814	13,814	7467	7467	7467	12,454	12,454	12,454
IV tests									
Endogeneity									
- Robust score test (*p*-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
- Robust regression-based test (*p*-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weak instrument									
Adj. Partial R2									
- Ln(U. Pop. (000 people)	0.55	0.55	0.55	0.54	0.54	0.54	0.55	0.55	0.55
- Ln(U.Pop. (000 people))^2	0.57	0.57	0.57	0.55	0.55	0.55	0.57	0.57	0.57
- Ln(Localization)	0.29			0.32			0.29		
- Ln(Formal localization)		0.21			0.29			0.21	
- Ln(Informal localization)			0.25			0.25			0.24
F-test first stage									
- Ln(U. Pop. (000 people)	8724	8724	8724	4763	4763	4763	8035	8035	8035
- Ln(U.Pop. (000 people))^2	7328	7328	7328	4208	4208	4208	6860	6860	6860
- Ln(Localization)	2061			1407			1841		
- Ln(Formal localization)		1543			1176			1347	
- Ln(Informal localization)			1565			913			1407

Source:Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.

Notes: Standard errors in parentheses

Ln(Urban population), Ln(Urban population) squared, and the logarithm of localisation are endogenous variables. IVs include ln(Population (000 people) in 1989), the proportion of urban population in 1989, ln(market access (000 people) in 1989).

Robust 2SLS IV regression is applied and standard errors are clustered by cities. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Observations are at the area-sector level.

The story of scale externalities is different in large cities. Urbanization benefits informal workers but not their formal counterparts (Table 10). Furthermore, overall localization and formal localization positively affect the wage premium of all types of workers, as shown in Table 10. It is interesting that while informal localization brings benefits to informal workers in all sectors (column 12 and column 15, Table 10), it does not generate the wage premium for formal workers (column 6, Table 10). Our results show that while informal workers benefit from agglomeration externalities, their existence crowds out the formal sector (Overman and Venables 2005, 2010; Moreno-Monroy 2012), at least in the case of large cities in Vietnam.

Table 8. Agglomeration effects on wages of formal workers in urban areas using IV regressions.

	(1)	(2)	(3)
Ln(Urban population)	−4.294	−4.404	−2.393*
	(4.015)	(4.423)	(1.385)
Ln(Urban population) squared	0.581	0.595	0.315
	(0.545)	(0.597)	(0.197)
Ln(Localization)	−0.057		
	(0.170)		
Ln(Formal localization)		−0.054	
		(0.163)	
Ln(Informal localization)			0.069
			(0.124)
Ln(Distance to sea (km))	−0.118	−0.120	−0.056
	(0.124)	(0.132)	(0.047)
Constant	7.455	7.607	3.808
	(7.669)	(8.286)	(2.541)
Observations	12,452	12,452	12,452
IV tests			
Endogeneity			
- Wooldridge 95 robust score test (*p*-value)	0.000	0.000	0.000
- Wooldridge 95 robust regression-based test (*p*-value)	0.000	0.000	0.000
Weak instrument			
Adj. Partial R2			
- Ln(Population (000 people)	0.56	0.56	0.56
- Ln(Population (000 people)) squared	0.57	0.57	0.57
- Ln(Localization)	0.26		
- Ln(Informal localization)		0.24	
- Ln(Formal localization)			0.16
F-test first stage			
- Ln(Population (000 people)	8325	8325	8325
- Ln(Population (000 people)) squared	7203	7203	7203
- Ln(Localization)	1582		
- Ln(Formal localization)		1355	
- Ln(Informal localization)			833

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.

Notes: Standard errors in parentheses.

Ln(Urban population), Ln(Urban population) squared, and the logarithm of localization are endogenous variables. IVs include ln(Population (000 people) in 1989), the proportion of urban population in 1989, ln(market access (000 people) in 1989).

Robust 2SLS IV regression is applied and standard errors are clustered by cities. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Observations are at the area-sector level.

Table 9. Agglomeration effects on wages of workers in urban areas in small cities using IV regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	All workers			Formal workers			Informal workers			Informal workers in the informal sector			Informal workers in the formal sector		
Ln(Urban popion)	−1.566**	−1.534***	−1.531***	−1.218**	−1.360**	−0.696	−1.083	−1.495***	−1.014	−0.771	−1.406*	−0.641	−1.175*	−1.387***	−1.130*
	(0.634)	(0.565)	(0.557)	(0.596)	(0.655)	(0.453)	(0.683)	(0.514)	(0.718)	(1.139)	(0.834)	(1.215)	(0.620)	(0.492)	(0.652)
Ln(U.Pop)^2	0.339**	0.333***	0.329***	0.255**	0.288**	0.139	0.245*	0.329***	0.232	0.182	0.312*	0.156	0.261**	0.303***	0.253*
	(0.135)	(0.122)	(0.115)	(0.127)	(0.140)	(0.097)	(0.145)	(0.110)	(0.151)	(0.238)	(0.176)	(0.253)	(0.132)	(0.106)	(0.137)
Ln(Localization)	0.014			0.081**			−0.070*			−0.079			−0.043		
	(0.036)			(0.034)			(0.039)			(0.054)			(0.036)		
Ln(Formal localization)		0.011			0.094**			−0.050*			−0.063			−0.029	
		(0.028)			(0.041)			(0.029)			(0.044)			(0.025)	
Ln(Informal localization)			0.017			0.059**			−0.074*			−0.089			−0.047
			(0.043)			(0.025)			(0.041)			(0.060)			(0.039)
Ln(Distance to sea (km))	−0.002	−0.002	−0.001	0.005*	0.005	0.005*	−0.005*	−0.002	−0.007**	−0.003	−0.002	−0.003	−0.004	−0.002	−0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)
Constant	1.511***	1.511***	1.493***	0.815	0.920	0.572	1.327**	1.536***	1.238*	1.035	1.439	0.922	1.309**	1.391***	1.252**
	(0.551)	(0.553)	(0.518)	(0.525)	(0.564)	(0.474)	(0.604)	(0.533)	(0.644)	(1.081)	(0.908)	(1.141)	(0.553)	(0.510)	(0.588)
Observations	10,348	10,348	10,348	6921	6921	6921	8036	8036	8036	3835	3835	3835	6953	6953	6953

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.

Notes: Standard errors in parentheses.

Ln(Urban population), Ln(Urban population) squared, and the logarithm of localization are endogenous variables. IVs include ln(Population (000 people) in 1989, the proportion of urban population in 1989, ln(market access (000 people) in 1989).

Robust 2SLS IV regression is applied and standard errors are clustered by cities. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

IV tests show all IVs are valid.

Observations are at the area-sector level.

Table 10. Agglomeration effects on wages of workers in urban areas in large cities using IV regressions.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	All workers			Formal workers			Informal workers			Informal workers in the informal sector			Informal workers in the formal sector		
Ln(Urban population)	−0.333**	−0.956***	−0.213	−0.225	−0.199	−4.624*	−0.504***	−1.246***	0.435	−0.671***	−1.460***	0.126	−0.451***	−1.238***	0.619
	(0.153)	(0.118)	(1.316)	(0.197)	(0.171)	(2.472)	(0.150)	(0.140)	(0.326)	(0.244)	(0.205)	(0.541)	(0.165)	(0.147)	(0.382)
Ln(U.Pop)^2	0.023	0.098***	−0.078	0.008	0.012	0.426**	0.046***	0.129***	−0.059	0.061**	0.149***	−0.029	0.040**	0.128***	−0.081*
	(0.018)	(0.013)	(0.242)	(0.024)	(0.020)	(0.215)	(0.018)	(0.014)	(0.039)	(0.029)	(0.020)	(0.065)	(0.020)	(0.015)	(0.045)
Ln(Localization)	0.193***			0.254***			0.138***			0.123***			0.151***		
	(0.026)			(0.038)			(0.024)			(0.041)			(0.026)		
Ln(Formal localization)		0.077***			0.180***			0.058***			0.045***			0.062***	
		(0.010)			(0.023)			(0.011)			(0.016)			(0.011)	
Ln(Informal localization)			1.388			0.713			0.249***			0.230***			0.285***
			(1.570)			(0.463)			(0.051)			(0.089)			(0.061)
Ln(Distance to sea (km))	0.012***	−0.001	0.108	0.016***	0.009***	0.015	0.009***	−0.003	0.029***	−0.001	−0.012***	0.014	0.013***	0.001	0.036***
	(0.003)	(0.002)	(0.121)	(0.004)	(0.003)	(0.016)	(0.003)	(0.003)	(0.006)	(0.004)	(0.004)	(0.009)	(0.003)	(0.003)	(0.007)
Constant	−0.246	1.859***	−4.447	−0.829	−0.442	8.735*	0.349	2.589***	−2.241**	0.929	3.279***	−1.372	0.140	2.540***	−2.843***
	(0.420)	(0.270)	(7.630)	(0.561)	(0.447)	(4.676)	(0.410)	(0.321)	(0.934)	(0.687)	(0.482)	(1.595)	(0.452)	(0.339)	(1.102)
Observations	6302	6302	6302	5531	5531	5531	5778	5778	5778	3632	3632	3632	5501	5501	5501

Source: Authors calculations from the 2013-2020 Vietnam Labor Force Surveys.

Notes: Standard errors in parentheses.

Ln(Urban population), Ln(Urban population) squared, and the logarithm of localization are endogenous variables. IVs include ln(Population (000 people) in 1989), the proportion of urban population in 1989, ln(market access (000 people) in 1989).

Robust 2SLS IV regression is applied and standard errors are clustered by cities. * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

IV tests show all IVs are valid.

Observations are at the area-sector level.

Conclusion

While workers, including informal employees, continue to migrate to dense urban areas and bigger cities, quantifying sources of the urban wage premium and isolating agglomeration impacts is essential for development policies. This study uses data from the 2013-2020 Vietnam Labor Force Survey to investigate whether cities generate positive agglomeration impacts on the earnings of informal workers and whether informal agglomeration has effects on the wages of formal workers in urban areas of Vietnam.

The results, analyzed at the aggregated district level, reveal that wages are higher in bigger cities though the pattern of wage growth relative to population is slightly different between the earnings of formal and informal workers. To evaluate whether cities, especially bigger cities, are productive by generating positive agglomeration impacts on wages, we regress the logarithm of workers' earnings on agglomeration, including urbanization, informal and formal localization, controlling for individual skills and local non-human endowments. When analyzing at the individual level, our results still show that wages are higher in bigger cities. In addition to workers' skills and local non-human endowment, agglomeration generates the wage premium in cities. Nevertheless, the scale externalities are different depending on the city size.

Our results show that while urbanization economies generate the wage premium for formal workers in small cities, their informal counterparts take advantage of urbanization externalities in large cities in Vietnam. Both localization and urbanization do not generate the earnings premium for informal workers in small cities, indicating the over-concentration of informal employment in these cities. In large cities, while informal workers benefit from the positive effects of formal and informal localization, their existence does not generate the earnings premium for formal workers.

Our results provide evidence to the literature that the informal sector is over-concentrated in small cities, and the existence of informality crowds out the formal sector in large cities of Vietnam. Thus, development policies and urban planning need to consider the over-concentration of informal sector workers in cities in Vietnam.

Notes

1. Data on the territory is taken from Portal (2018).
2. This is not a problem because only around 4% of our sample are informal business owners in urban areas.
3. The Labor Code of Vietnam requires the minimum age of entering the labor force is 15. We do not impose the upper bound for age because there is no age limit for the informal worker.
4. The number of firms is calculated from the Vietnam Establishment Census in 2018.

5. Vietnam carried out its first population census in 1979 since the country's union. However, given its limited source after a long time in the war, the data was not stored at the district level. Therefore, the 1989 population census and the intercensal population survey in 1993 are the longest lags of population agglomeration at the district level.
6. We carry out the Ramsey RESET test to check for quadratic effects of the urban population. The p_value is equal to 0.000, which allows rejecting the null hypothesis that the linear combination of exploratory variables is appropriate.
7. The elasticity of wages with respect to the city population estimated in the model that controls for individual characteristics in the first step and the one without individual control in the first step varies from −0.918 and −1.676 at the minimum urban population value to 0.753 and 1.398 at the maximum value of the population.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is funded by the National Economics University, Hanoi, Vietnam.

ORCID

Hai Anh La http://orcid.org/0000-0003-1063-3891

References

Arnott, R. 1979. "Optimal City Size in a Spatial Economy." *Journal of Urban Economics* 6 (1): 65–89. doi:10.1016/0094-1190(79)90017-2.

Barufi, A. M. B., E. A. Haddad, and P. Nijkamp 2016. "Industrial Scope of Agglomeration Economies in Brazil." *The Annals of Regional Science* 56 (3): 707–755. doi:10.1007/s00168-016-0768-3.

Bonnet, F., J. Vanek, and M. Chen 2019. Women and Men in the Informal Economy: A Statistical Brief. Geneva: International Labor Office.

Briant, A., P.-P. Combes, and M. Lafourcade 2010. "Dots to Boxes: do the size and shape of spatial units jeopardize economic geography estimations?." *Journal of Urban Economics* 67 (3): 287–302. doi:10.1016/j.jue.2009.09.014.

Brulhart, M. 1998. "Economic Geography, Industry Location and Trade: the evidence." *The World Economy* 21 (6): 775–801. doi:10.1111/1467-9701.00163.

Cameron, A. C., and D. L. Miller 2015. "A Practitioner's Guide to Cluster-Robust Inference." *The Journal of Human Resources* 50 (2): 317–372. doi:10.3368/jhr.50.2.317.

Ciccone, A., and R. E. Hall 1996. "Productivity and the Density of Economic Activity." *The American Economic Review* 86 (1): 54–70.

Combes, P.-P., G. Duranton, and L. Gobillon 2008. "Spatial Wage Disparities: sorting matters!." *Journal of Urban Economics* 63 (2): 723–742. doi:10.1016/j.jue.2007.04.004.

Dang, A., S. Goldstein, and J. McNally 1997. "Internal Migration and Development in Vietnam." *The International Migration Review* 31 (2): 312–337. doi:10.1177/019791839703100203.

De la Roca, J. D. L., and D. Puga 2017. "Learning by Working in Big Cities." *The Review of Economic Studies* 84 (1): 106–142. doi:10.1093/restud/rdw031.

Duranton, G. 2008. "From Cities to Productivity and Growth in Developing Countries." *Canadian Journal of Economics/revue Canadienne d'économique* 41 (3): 689–736. doi:10.1111/j.1540-5982.2008.00482.x.

Duranton, G. 2016. "Agglomeration Effects in Colombia." *Journal of Regional Science* 56 (2): 210–238. doi:10.1111/jors.12239.

Ellison, G., E. L. Glaeser, and W. R. Kerr 2010. "What Causes Industry Agglomeration? Evidence from Co-Agglomeration Patterns." *The American Economic Review* 100 (3): 1195–1213. doi:10.1257/aer.100.3.1195.

Fujita, M. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge Books Online: Cambridge University Press. doi:10.1017/CBO9780511625862.

García, G. A. 2019. "Agglomeration Economies in the Presence of an Informal Sector: the Colombian case." *Revue dEconomie Regionale Urbaine* 2: 355–388.

General Statistics Office of Vietnam. 2011a. "The 2014 Vietnam Intercensal Population and Housing Survey: Migration and Urbanization in Vietnam."

General Statistics Office of Vietnam. 2011b. "Vietnam Population and Housing Census in 2009/migration and Urbanization in Vietnam: Patterns, Trends, and Differentials."

General Statistics Office of Vietnam. 2014. "*Informal Employment in the Period of 2007-2013*." Hanoi, Vietnam: GSO report.

General Statistics Office of Vietnam. 2018. "*Bao cao lao dong phi chinh thuc 2016* [The 2016 Report on Informal Employment in Vietnam]." ILO Working Paper.

Ghani, E., and R. Kanbur 2013. "Urbanisation and (In) Formalisation." The World Bank.

Glaeser, E. L., and D. C. Maré 2001. "Cities and Skills." *Journal of Labor Economics* 19 (2): 316–342. doi:10.1086/319563.

Håkansson, J., and G. Isacsson 2019. "The Spatial Extent of Agglomeration Economies Across the Wage Earnings Distribution." *Journal of Regional Science* 59 (2): 281–301. doi:10.1111/jors.12411.

Hansen, W. G. 1959. "How Accessibility Shapes Land Use." *Journal of the American Institute of Planners* 25 (2): 73–76. doi:10.1080/01944365908978307.

He, C., Y. D. Wei, and F. Pan 2007. "Geographical Concentration of Manufacturing Industries in China: the importance of spatial and industrial scales." *Eurasian Geography and Economics* 48 (5): 603–625. doi:10.2747/1538-7216.48.5.603.

Henderson, J. V. 2003. "Marshall's Scale Economies." *Journal of Urban Economics* 53 (1): 1–28. doi:10.1016/S0094-1190(02)00505-3.

Howard, E., C. Newman, and F. Tarp 2016. "Measuring Industry Co-Agglomeration and Identifying the Driving Forces." *Journal of Economic Geography* 16 (5): 1055–1078. doi:10.1093/jeg/lbv037.

Jacobs, J. 1969. *The Economy of Cities*. New York: Random House.

Jaffe, A. B., M. Trajtenberg, and R. Henderson 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics* 108 (3): 577–598. doi:10.2307/2118401.

Kim, S. 1989. "Labor Specialisation and the Extent of the Market." *The Journal of Political Economy* 97 (3): 692–705. doi:10.1086/261622.

Lall, S., Z. Shalizi, and U. Deichmann 2004. "Agglomeration Economies and Productivity in Indian Industry." *Journal of Development Economics* 73 (2): 643–673. doi:10.1016/j.jdeveco.2003.04.006.

Matano, A., M. Obaco, and V. Royuela 2020. "What Drives the Spatial Wage Premium in Formal and Informal Labor Markets? the Case of Ecuador." *Journal of Regional Science* 60 (4): 823–847. doi:10.1111/jors.12486.

Moomaw, R. L. 1981. "Productivity and City Size: a critique of the evidence." *The Quarterly Journal of Economics* 96 (4): 675–688. doi:10.2307/1880747.

Moreno-Monroy, A. 2012. "Critical Commentary. Informality in Space: understanding agglomeration economies during economic development." *Urban Studies* 49 (10): 2019–2030. doi:10.1177/0042098012448554.

Mukim, M. 2015. "Coagglomeration of Formal and Informal Industry: evidence from india." *Journal of Economic Geography* 15 (2): 329–351. doi:10.1093/jeg/lbu020.

Nguyen, C. V., and T. P. Minh 2016. "Are Migrants in Large Cities Underpaid? Evidence from Vietnam." *IZA Journal of Migration* 5 (1): 20. doi:10.1186/s40176-016-0068-9.

Overman, H. G., and A. J. Venables. 2005. "Cities in the Developing World." Centre for Economic Performance, London School of Economics.

Overman, H. G., and A. J. Venables. 2010. "Evolving City Systems." WIDER Working Paper.

Portal, H. E. (2018). *He thong do thi Vietnam voi su phat trien dat nuoc* [Cities during the development in Vietnam]. Accessed 29 July 2020. https://hanam.gov.vn/sxd/Pages/he-thong-do-thi-viet-nam-voi-su-phat-trien-dat-nuoc.aspx

Porter, M. E. 1990. "The Competitive Advantage of Nations." *Harvard Business Review* 68 (2): 73–93.

Puga, D. 2010. "The Magnitude and Causes of Agglomeration Economies." *Journal of Regional Science* 50 (1): 203–219. doi:10.1111/j.1467-9787.2009.00657.x.

Punzo, G., R. Castellano, and E. Bruno 2022. "Using Geographically Weighted Regressions to Explore Spatial Heterogeneity of Land Use Influencing Factors in Campania (Southern Italy)." *Land Use Policy*, 112, 105853.

Rosenthal, S. S., and W. C. Strange 2008. "The Attenuation of Human Capital Spillovers." *Journal of Urban Economics* 64 (2): 373–389. doi:10.1016/j.jue.2008.02.006.

Tran, T.B. 2015. "*Sản xuất công nghiệp phi chính thức ở Việt Nam: tập trung theo ngành hay da dạng hoá? Tạp chí Kinh tế & Phát triển*, Số 223 [The urban informal manufacturing sector in Vietnam: localisation or urbanisation?." *Journal of Economics and Development* 223, 21–29.

Tran, T. B., and H. A. La 2018. "Agglomeration Effects: productivity of the informal sector in vietnam." *The Journal of Development Studies* 54 (2): 292–311. doi:10.1080/00220388.2017.1283013.

Wheaton, W. C., and M. J. Lewis 2002. "Urban Wages and Labor Market Agglomeration." *Journal of Urban Economics* 51 (3): 542–562. doi:10.1006/juec.2001.2257.