Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics

Document de Treball 2018/13, 35 pàg.
Working Paper 2018/13, 35 pag.

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

Document de Treball 2018/06, 35 pàg.
Working Paper 2018/06, 35 pag.

"What drives the spatial wage premium for formal and informal workers? The case of Ecuador"

Alessia Matano, Moisés Obaco and Vicente Royuela

Research Institute of Applied Economics

Regional Quantitative Analysis Research Group

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics
WEBSITE: www.ub-irea.com • CONTACT: irea@ub.edu

AQR
Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group
WEBSITE: www.ub.edu/aqr/ • CONTACT: aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

Abstract

This article investigates the incidence of agglomeration externalities in a typical developing country, Ecuador. In particular, we analyze the role of the informal sector within these relations, since informal employment accounts for a significant part of total employment in the developing countries. Using individual level data and instrumental variable techniques, we investigate the impact of spatial externalities, in terms of population size and local specialization, on the wages of workers in Ecuadorian cities. The results show that spatial externalities matter also for a typical developing country, especially as far as urbanization externalities are concerned. Moreover, analysis of the interaction between spatial externalities and the informal economy shows a general penalization for informal workers in terms of benefits arising from agglomeration externalities. Finally, by investigating the possible channels behind the heterogeneity found in spatial agglomeration gains between formal and informal workers, we show that the advantages from agglomeration for formal workers may well be accounted for by positive sorting and better gains from job changes, while for informal workers they rise from positive learning externalities.

JEL Classification: J31, J46, R23, R12.

Keywords: Agglomeration Externalities, Developing Economies, Informal Employment, Workers' Wages, FUAs, Ecuador.

Alessia Matano AQR-IREA, University of Barcelona (UB). Tel.: +34-934021825; Fax.: +34-934021821. Department of Econometrics, Statistics and Applied Economics, University of Barcelona, Diagonal 690, 08034 Barcelona, Spain. E-mail address: amatano@ub.edu

Moisés Obaco. AQR-IREA, University of Barcelona, Av. Diagonal 690 (08034), Barcelona, Spain. Email: mobacoal7@ub.edu

Vicente Royuela. AQR-IREA, University of Barcelona, Av. Diagonal 690 (08034), Barcelona, Spain. Email: vroyuela@ub.edu

Acknowledgements

A. Matano and V. Royuela express gratitude for the support received from Spain's Ministerio de Ciencia e Innovación through action ECO2016-75805-R. M. Obaco acknowledges the support received by the AGAUR. We thank the participants in the ERSA Conference (Groningen, 2017), Geoffrey J.D. Hewings Regional Economics Workshop (Wien, 2017), XLIII International Conference on Regional Science (Seville, 2017), 8th Slovak Winter Seminar of Regional Science (Terchova, 2018), and University of Groningen Seminar (2018). Our thanks are also due to Jorge de La Roca for useful suggestions.

1. Introduction

With the growth of big cities, analysis of the benefits of agglomeration economies has been pursued in a wide range of studies. From the empirical point of view, an extensive literature has analyzed the extent of agglomeration economies, measured by population size (or density) and industrial specialization at the local level. Their results have shown positive impacts of spatial externalities on productivity and wages (see among others Combes, 2000; Combes et al., 2008; Mion and Naticchioni, 2009).

These studies have generally been carried out on developed countries. Less attention has been paid to the role of agglomeration economies in the developing world. However, the topic is relevant since the growth rate of the world's urban development is being driven by urbanization in the developing world. Exploring the role of agglomeration economies in the developing countries is relevant to assessing the importance of urban economies worldwide, since the urbanization process taking place in the developing world is different from the old urbanization, mainly because of high poverty rates and poor-quality institutions (Glaeser and Henderson, 2017).

A few studies have looked into the importance of agglomeration economies in the developing countries, focusing on those characterized by great geographical extension and large populations, such as China, India, Brazil or Colombia (see Chauvin et al., 2017; Combes et al., 2015 and Duranton, 2016).[1] The findings have shown an important role for spatial externalities in fostering productivity and wages, with impacts higher than detected for the developed world. However, these results have been found for emerging markets economies, while there is no evidence of studies focusing on more typical developing countries. Moreover, the developing countries are generally characterized by a large proportion of informal economy, which leads to a dual labour market (Fields, 1990; La Porta and Shleifer, 2014). Although the theoretical implications of the presence of informal employment on agglomeration externalities are unclear, assuming the non-existence of agglomeration economies for the informal sector is hardly justified (Overman and Venables, 2005). Also, Duranton (2009) points out that formal and informal sectors have strong interconnections, thus suggesting that in both

[1] For a detailed review see Combes and Gobillon (2015).

sectors agglomeration effects generate benefits. Nonetheless, more formal studies on this topic are lacking (Overman and Venables, 2005).

The aim of this paper is twofold. First of all, we analyze the importance of agglomeration economies for a typical developing economy, Ecuador. This country is shaped by a set of characteristics still unexplored in the literature such as average geographical extension and population size[2], low and increasing urbanization rate, weak industrial activity, and widespread informal employment. Our first goal is to understand whether agglomeration externalities exert impacts similar to those detected for the developed world and/or the emerging economies. Secondly, we analyze the importance of the informal sector within these relations, exploring the heterogeneity of spatial externalities between formal and informal workers. Also, we shed light on the channels behind the detected impacts for each category of workers. Our paper contributes to the literature by directly and comprehensively analyzing how spatial externalities affect informal workers in a typical developing country.[3]

We use repeated quarterly cross-section data from the Ecuadorean Labour Surveys (ENEMDU) available from 2005 to 2015, together with historical data from the 1950 and 1990 censuses. We take Functional Urban Areas (FUAs hereafter) as units of analysis, since they represent a suitable economic definition of cities (OECD, 2013).[4] We assess the extent of spatial externalities by analyzing the impact of two measures of agglomeration on individual workers' real wages: population size and sector specialization at the local level.

Two major methodological issues arise in estimating this relationship. First, there could be sorting of workers should the more skilled individuals prove more likely to be located in highly agglomerated areas. This issue has been taken into account in the empirical literature using individual panel data and performing fixed effects

[2] Ecuador is not only in the middle of the world (*en la mitad del mundo*): it has a population size and extension close to the median of the developing countries (the country list comes from the *World Economic Situation and Prospects, WESP, 2014, UN*).

[3] Actually, there are practically no papers analyzing the impact of agglomeration externalities on workers' wages in the informal sector. A few exceptions are Duranton (2016) and Garcia (2016), who focus on Colombia. However, Duranton (2016) does not directly test the impact of agglomeration economies for the informal economy, but infers its impact by comparing estimates for all workers with estimates for formal workers only, finding lower impacts for the latter group, and indicating higher impacts for the informal economy. Garcia (2016), focusing on density, provides a direct test of this relationship showing negative agglomeration externalities for formal workers and positive ones for informal workers. Nonetheless, neither of the articles investigates the channels behind the differences revealed, which is one of the main purposes of the current analysis.

[4] For details, see Section 3 and Obaco et al. (2017).

estimations, which control for unobserved individual heterogeneity (see Combes et al., 2008; Combes et al., 2010; Matano and Naticchioni, 2012; Mion and Naticchioni, 2009, among others). However, for this paper, as for previous studies on developing countries (Chauvin et al., 2017; Combes et al., 2015; Duranton, 2016), the individual panel data that would allow to control for the unobserved individual heterogeneity were not available. Therefore, we are not able to control for individual unobserved heterogeneity in the course of our empirical analysis. Nonetheless, on the one hand, we will do our best from a methodological point of view by introducing into the estimation a broad range of individual controls to capture the individual heterogeneity (as in Duranton, 2016, and Glaeser and Resseger, 2010). On the other hand, we will make use of the information from a panel subsample of the original data that refers to stayer workers (not moving across cities) to get some insight into the sorting of skilled workers in bigger cities or highly specialized areas.

The second methodological concern is the endogeneity due to possible simultaneity in individual choices concerning wages and locations. We will take this issue into account by applying an instrumental variable strategy, using deeply lagged values of our agglomeration measures and a geological attribute to build our instruments as in Combes et al. (2008), Combes et al. (2010), Matano and Naticchioni (2012), and Mion and Naticchioni (2009).

Our results, based on the IV estimation strategy, lead us to the following findings. First, agglomeration externalities increase productivity and wages also in a typical developing country like Ecuador. In particular, we find that urbanization externalities entail a positive impact on individual wages with an elasticity estimate of 3.8%, lower than found for larger developing economies (Chauvin et al., 2017; Combes et al., 2015; Duranton, 2016). As for local industrial specialization, there is also a positive impact, lower than that attributed to urbanization externalities (0.9%). Also, we find evidence of decreasing returns to city size, since on excluding the biggest cities from the estimation, the impact of urbanization externalities increases to 9.1%.

When we consider the informal sector in our analysis, our findings confirm that it matters very much. First of all, there is evidence of a wage penalization for informal workers, since their earnings are significantly lower (-7%) than those of otherwise similar workers employed in the formal labour market, as expected. When considering the interaction with agglomeration externalities, our findings show that informal workers enjoy lower benefits than formal workers. In particular, wage increases for informal workers in an urban agglomeration are 1.6% lower than for formal workers. As for specialization, there are no longer wage benefits for informal workers accruing from working in areas characterized by a high level of local sectoral specialization. On the contrary, workers are slightly penalized. These outcomes might be partially explained on considering that informal workers are mainly represented by low-educated workers, for whom we detect lower agglomeration benefits compared to high-educated workers.[5]

Finally, we analyze the channels behind the detected impacts of spatial externalities on wages for formal and informal workers. In particular, we exploit the information derived from a panel subsample of our original data, and look to the sorting, matching and learning mechanisms to explain the outcomes of the analysis (Combes and Gobillon, 2015; Puga, 2010). As for urbanization externalities, the results for formal workers show that the urban wage premium is driven by the sorting of higher (unobservable) skilled workers and better quality of job match into larger cities. For informal workers, the gains from urbanization externalities are essentially driven by learning effects. As for specialization, for formal workers there is no evidence of sorting of skilled workers into highly specialized areas, while there is evidence of some wage gains from job changes, albeit lower than in highly populated areas. For informal workers, there is evidence of marked negative sorting in highly specialized areas, which helps to account for the wage penalization detected for this category of workers. To conclude, taking into account the fact that formal workers on average have higher skills than informal workers, these findings are in line with the previous literature indicating matching and sorting as channels behind the spatial wage premium for high-skilled workers, while learning externalities are more important for low-skilled workers (Matano and Naticchioni, 2016).

The paper is structured as follows. Section 2 introduces the reference literature on spatial agglomeration in developed and developing countries. Section 3 describes the data and defines the variables used for the empirical analysis. Section 4 shows the methodology, performs the empirical analysis and draws the results, while Section 5 concludes.

[5] This finding contrasts with that of Duranton (2016), who observes higher benefits for informal workers from spatial externalities. However, and in contrast with the present paper, he does not find evidence of higher agglomeration benefits for higher educated workers.

2. Related Literature

The idea of agglomeration economies fostering productivity and wages has been widely investigated in both the theoretical and empirical literature. Two of the factors most frequently analyzed are urban agglomeration and local sectoral specialization.

Marshall (1890) was a pioneer in pointing out the productivity gains that may arise in bigger cities or from the concentration of a specific industry in a given location. The channels were formalized by Duranton and Puga (2004): the learning mechanism, which reflects the idea of knowledge spillovers and face-to-face interactions in agglomerated areas that enhance human capital; the matching mechanism, showing that agglomerated areas offer conditions for a better match between workers and firms; and the sharing mechanism, lying in the advantages generated by sharing indivisible goods such as facilities and risk in new investments, thereby cutting individual and firm costs. These mechanisms are behind both urbanization and specialization positive spillovers, the former being associated with highly dense areas, and therefore across-industries, while the latter occur within specific industries located in the same area.

From an empirical point of view, many works have analyzed the role of spatial externalities in fostering productivity and wages using both aggregated data (see Ciccone and Hall, 1996; Ciccone, 2002; Combes, 2000; among others), and individual level data (see Combes et al., 2008; Combes et al., 2010; De la Roca and Puga, 2017; Glaeser and Marè, 2001; Mion and Naticchioni, 2009). These studies take into account the empirical issues that arise in identification of the role of such externalities. In particular, they address the endogeneity of the relationship by using an instrumental variable strategy, while the role of the sorting of skilled workers in highly agglomerated areas is examined by means of a fixed effect strategy when using individual level panel data. The results have shown that worker sorting captures a large part of the impact imputed to spatial externalities on wages. Further, research has also revealed the relevance of dynamic gains in the biggest cities (De la Roca and Puga, 2017).

However, these studies have in general been carried out on the developed countries.[6] Less attention has been paid to the role of spatial externalities in affecting wages in the developing countries. However, most of the urban population growth is

[6] For an exhaustive review see Combes and Gobillon (2015), Melo et al. (2009) and Rosenthal and Strange (2004).

driven by the growth of cities in the developing countries and, interestingly, not in the largest cities (Royuela and Castells-Quintana, 2015). Thus, in line with Glaeser and Henderson (2017), we consider important to study whether agglomeration economies in the developing countries exert the same impacts on wages and productivity as in the developed world. Few works have addressed this issue. Chauvin et al. (2017) use cross-sectional individual level data and IV estimates to analyze the difference in urbanization economies between similar-sized cities in the US, China, India and Brazil from 1980 to 2010. They show higher elasticities of wages with respect to urban size for China and India (around 8/9% to population size), and lower for Brazil, which is close to the US estimate (5%). Combes et al. (2015) focus on China and find agglomeration impacts on native wages of around 11%.[7] Duranton (2016), taking the case of Colombia, finds an elasticity of wages with respect to population size of 5%, while Ahrend et al. (2017) find a similar magnitude in the case of Mexico (4.2%).[8]

In contrast to the works analyzing developed countries, these studies are not able to control for unobserved individual heterogeneity due to the lack of a longitudinal data structure. Nonetheless, they make use of a wide range of individual level controls in order to take into account individual heterogeneity as far as possible (see Combes and Gobillon, 2015, for a discussion). Furthermore, a major concern in the case of the developing countries lies in the informal economy, which implies the presence of a dual labour market, informal workers representing around half of the labour force (La Porta and Shleifer, 2014; Maloney, 2004; Overman and Venables, 2005). Informal labour is usually represented by low-educated workers (Fields, 1990; Khamis, 2012) and associated with high levels of vulnerability and poverty, low wages and productivity, and lack of legal protection (Bacchetta et al., 2009; Fields, 1990). Consequently, in our view, a thorough analysis of the relevance of spatial externalities in a developing economy should take this dimension into account (Glaeser and Henderson, 2017).

From the theoretical point of view, the literature has not yet reached consensus on the relationship between agglomeration externalities and the informal economy.

[7] They also examine the impact of sectoral specialization at the local level and, by using OLS estimates, find elasticities of wages with respect to specialization of around 6%.

[8] Other related studies are Au and Henderson (2006), who analyze the case of China showing an inverted U relationship between wages and city size, and Lall et al. (2004) who, using firm level manufacturing data, look into the role of spatial externalities for economic productivity in India, disentangling the sources of agglomeration economies between those arising at the firm level, at the industry level and at the regional level.

Overman and Venables (2005) discuss this issue arguing that, on the one hand, the informal economy should decrease the gains from agglomeration economies due to crowding out effects on the formal sector, in line with the spirit of the Harris-Todaro (1970) model. On the other hand, they stress that this crowding out effect might not be fast enough to offset the advantages deriving from agglomeration. Furthermore, they point out that the informal sector itself could contribute to agglomeration economies because of the vitality of this sector, the existence of networks of small firms benefitting from labour market pooling and the role of the informal economy shown in the literature on clusters in developing countries. This spillover effect is also stressed by Duranton (2009), who argues that the formal and informal sectors are closely interconnected, so positive effects could arise within both sectors.

From the empirical point of view, the role of the informal economy within the relation between spatial externalities and wages has generally been neglected. One exception is Duranton (2016), who uses individual level data for Colombia to carry out an indirect test performing estimations of the spatial externalities impact on wages vis-à-vis on the one hand all workers, and on the other hand only workers with a written labour contract. His findings show that spatial externalities have a lower impact when only workers with written contracts are taken into account (around 4% compared to 5% for all workers), thus inferring higher benefits due to agglomeration economies for informal workers.[9]

3. Description of the data and definition of variables of interest

In this paper we focus on Ecuador, a typical developing country characterized by low per-capita income, and low industrialization and urbanization rates as compared with other Latin America countries. The total population size is close to the median of developing countries (around 16 million inhabitants)[10] and Ecuador is included in the group of countries that have gone through rapid urbanization since 1960.

[9] Other related papers are by Garcia (2016) who, like Duranton (2016), uses individual level data to study the case of Colombia finding positive agglomeration benefits (in terms of density) for the informal economy and negative ones for the formal economy. However, he does not analyze the channels behind the detected impact, which is one of the purpose of this paper. Also, Harris (2014), using firm level data for the Nairobi's handicraft industry, finds disadvantages in terms of agglomeration externalities in relation to the informal economy. Further, Bernedo Del Caprio and Patrick (2017) use firm level data for Peru to show lower benefits from agglomeration externalities for the informal economy.

[10] The list of developing economies is from the *World Economic Situation and Prospects 2014*, UN, Statistical Annex.

We use data from the Ecuadorean Labour Surveys (*Encuesta Nacional de Empleo, Desempleo y Subempleo* - ENEMDU) provided by the National Statistics Institute of Ecuador (INEC). These are quarterly surveys (repeated cross-section data)[11] that contain detailed information on Ecuadorean workers: labour income, hours worked, age, gender, ethnicity, education, occupation, previous migrant status and informal employment status. We also have information on the workers' area/city of residence and sector of employment. In order to build the instruments for analysis, we combine these data with data from the population censuses of 1950 and 1990.[12] Finally, we also merge these data with information on a geological variable associated with the city soil (PH, which measures the acidity of the soil) to check the robustness of our IV strategy.

As spatial units of analysis, we use an international harmonized concept of economic cities, namely Functional Urban Areas,[13] as in Ahrend et al. (2017). We identify 28 FUAs in Ecuador (see Figure A1, in the Appendix).[14] The most populated FUAs are the capital city, Quito, and Guayaquil with more than 1.5 million inhabitants each, located in the Andean and Coastal region respectively.[15]

The period of the empirical analysis is from 2005 to 2015 (44 quarters).[16] We restrict the sample to workers located in the FUA areas (males and females) aged between 15 and 64 years old who perform one job.[17] We exclude workers employed in the public sector since their wage is set nationally, and focus on employees and self-employed

[11] In particular, the survey is designed to sample specific buildings/dwellings (*viviendas*) and to provide information on the families living in the sampled buildings/dwellings. Moreover, 25% of the sample in each quarter is subsequently re-sampled for four non-consecutive quarters. Section 4.3. provides a more detailed description of the panel dimension of the database.

[12] 1950 saw the first population census. The data are from the National Institute of Statistics and Census of Ecuador (INEC).

[13] Throughout the paper we will use "Functional Urban Area" and "city" as synonymous.

[14] In order to define a Functional Urban Area, we implement the following steps. We first identify the centers of the FUAs (urban cores), based on grid cells of high population density and population size. More specifically, these urban cores are defined as areas of continuous grid cells of high population density (500 inhabitants per km^2 as minimum threshold for grid cells) with a total population size (summing up the continuous cells) of at least 25,000 inhabitants. These preferred thresholds are suitable for Ecuador for taking into account the relatively low density of this country. Second, we use a travel-time approach to group all urban cores socio-economically connected (thus belonging to the same FUA) and to define the boundaries of the FUAs. The final FUAs represent both the urban cores and the connected municipalities forming the hinterland. For further details, see Obaco et al. (2017).

[15] Quito and Guayaquil are the largest FUAs. The rest of the Ecuadorian FUAs are medium-sized and small, with a substantial gap in terms of population with respect to the two biggest ones (Cuenca, the third city in the country, has around 500 thousand inhabitants).

[16] This time span also avoids the years of the previous economic crises (the peak of the great recession was in 1999) and dollarization process (introduced as national currency of Ecuador in January 2000).

[17] We drop the 2.8% of workers with more than one job.

workers. We clean the dataset by dropping observations with missing data in our variables of interest as well as observations below (above) the 1st (99th) percentile of the workers' real wage distribution, hours worked and real wages per hour.[18] We end up with a sample of 408,197 observations for the period of analysis, around 9,300 per quarter.

Informality is defined on the basis of the latest methodology implemented by the INEC (following the 2013 ILO guidelines): informal are the workers employed in firms with fewer than 100 employees with no tax identification number (*Registro unico del contribuyente*).[19]

We use two measures of agglomeration effects. The first, total FUA population ($Pop_{c,t}$), proxies urbanization externalities capturing the effect of urban scale on wages, as in Duranton (2016) and Ahrend et al. (2017). This information is provided by the INEC.

The second measure of agglomeration externalities is an index of specialization, which proxies sector specialization at the local level. We compute it as in Combes (2000), Matano and Naticchioni (2012) and Mion and Naticchioni (2009):

$$Spec_{c,s,t} = \frac{empl_{c,s,t} / empl_{c,t}}{empl_{s,t} / empl_{t}}$$

Where c stands for the city, s for the sector, and t for time. The specialization index is the ratio between the share of sectoral employment in the total employment in any city c and the corresponding share at the national level.[20]

---

[18] We deflate wages by using the Ecuadorian national CPI. The base year is 2014. CPI information is provided by the Ecuadorian Central Bank.

[19] The INEC has applied different definitions of informality over time. The previous definition was classifying informal workers as those employed in firms with 10 or fewer employees with no full accounting records or tax identification number. The current definition (INEC, 2015) follows the 2013 ILO guidelines and includes in the formal sector firms having tax identification numbers, even though they may not have full accounting records, in order to take into account cases where there is some legal justification. The classification divides workers into formal, informal, dwelling activities and non-classified workers. Formal and informal workers account for 90% of our sample, while 4% perform dwelling activities and 6% are non-classified workers. In the paper we only consider formal and informal workers. The new methodology was introduced in 2015 and applied to former surveys up to the second quarter of 2007 (included). Therefore, we will perform the analysis on informality from the second quarter of 2007 onwards to allow for conceptual comparability.

[20] Sectors are defined according to the ISIC version 3.1 at two-digit level. We use the ENEMDU surveys to compute the specialization index.

Table 1 presents the average of the real wages per hour (in US dollars) for the workers' categories considered in the empirical analysis, at the beginning, middle and end of the period under analysis, as well as for the whole time span of the analysis (2005-2015) combined with their relative presence in the sample.

Turning to the composition of the sample (Table 1, last column), it may be noted that males constitute around 60% of the labour force. The dominant ethnicity is mestizo (87%). Around 77% of the sample is constituted by low/medium-educated workers (educational levels are up to high school at most), while about 52% of the workers are employed in unskilled occupations, 35% in medium-skilled, and 13% in high-skilled ones. In terms of sector distribution, a relatively high percentage of workers (27%) are employed in the wholesale and retail trade sector and, in general, in the service sector (in total around 64% of workers); manufacturing, mining and agriculture account for just 26% of the total sample of workers (16% of which in manufacturing), construction for 10%. As for the dual labour market, workers in the informal sectors represent around 38% of the total workforce in the sample. Moreover, 44% of workers are self-employed, while 56% are employees. As for firm size, 86% of the workers are employed in firms with fewer than 100 employees. Further, considering the regions, 53% of the workers are located in the Andean region, 43% in the Coastal region, and the remaining 4% in the Amazonian region. As for the area of residence, around 89% of the workers live in an urban area of the FUA.[21] Finally, 31% of the workers declare they have not always lived in the city where they currently reside.

As for wages (Table 1, columns 1 to 4), a general increase in real wages per worked hour throughout the period analyzed may be noted. In addition, there is evidence of marked heterogeneity according to the dimensions considered: males earn around 10% more than females; white workers have a significant premium compared to the other ethnicities (12% above the average); wages increase with the worker's level of education and job skill intensity, as well as firm size and previous migrant status of the worker. In terms of the dual labour market, wages are significantly higher (about 60% more) for formal than for informal workers. Further, wages are lower for the self-employed than for the employees.[22] As for the economic sectors, apart from the mining and quarrying sector, wages are significantly higher in the service sector than in

[21] The interconnected economic areas of the FUAs may include rural areas.

[22] This is due to the fact that the self-employed are mostly informal workers.

manufacturing. In terms of regions, wages are higher in the Amazonian region (due to oil extraction), followed by the Andean region. Finally, and as expected, wages are significantly higher in the urban areas than in the rural ones.

[Table 1 around here]

4. Empirical Analysis

4.1. The impact of spatial externalities on average wages

In this section, we use data on repeated cross-sections for Ecuador from 2005 to 2015 to estimate a pool cross-section model as in Duranton (2016). The specification is:

$$\begin{aligned} \log(W_{i,t}) &= \alpha + \delta * \log(Pop_{c(i),t}) + \theta * \log(Spec_{c(i),s(i),t}) + X'_{i,t} * \beta + \mu * dfirmsize_{i,t} + \\ &+ S_{s(i)} + A_{c(i)} + R_{c(i)} + T_t + \varepsilon_{i,t} \end{aligned} \tag{1}$$

Where $log(W_{i,t})$ is the log of real wage per hour worked by worker *i* at time *t*. $log(Pop_{c(i),t})$ is the log of the total population size of city *c* where the worker *i* resides, while $log(Spec_{c(i),s(i),t})$ is the log of the specialization index for city *c* and sector *s* where the worker is employed. δ and θ are the parameters of interest (elasticities) that capture the extent of agglomeration effects on wages. $X_{i,t}$ is a vector of worker's characteristics including: age, age squared, female dummy, migrant dummy (indicating whether the worker has previously lived somewhere else), five categories for ethnicities, seven categories for education, twenty-seven categories for occupation (ISCO 88 at two-digit level), and job category (self-employed/employee) dummies. We also include a firm size dummy (over or under 100 employees) $dfirmsize_{i,t}$, sector dummies (ISIC 3.1, at two-digit level[23]) $S_{s(i)}$, dummies for the natural regions of Ecuador (Andean, Coastal and Amazonian), $A_{c(i)}$, a rural area dummy within the FUA, $R_{c(i)}$, and period (year*quarter) dummies T_t. $\varepsilon_{i,t}$ is the error term.

The OLS estimation of equation (1) may be affected by two main issues. First, there might be sorting of skilled workers into highly agglomerated areas. Generally, this is

[23] In order to instrument the specialization variable, and due to discrepancies in correspondence between sector classifications over time (ISIC2 for the census 1990 and ISIC3.1 for the ENEMDU data), we chose to group the three ISIC31 sectors 50, 51 and 52 in a single group representing the entire *Wholesale and Retail* sector. This prevented significant loss of observations. Nonetheless, we ran a robustness check by repeating the estimation without this grouping (therefore not using around 17,000 observations), and the results remain consistent.

taken into account by introducing individual fixed effects in the econometric specification. In our data, we cannot directly address this point since we would need a panel dataset to control for individual unobserved heterogeneity. Nonetheless, we do our best by introducing a wide range of worker control variables in line with Duranton (2016) and Glaeser and Resseger (2010). Later, in Section 4.3 we will address this point better by exploiting the panel dimension of a subsample of our data. Second, there might be a matter of endogeneity arising from the possible simultaneity in individual choices concerning wages and locations. We address this point by using an instrumental variables strategy. As instruments, we use cities' historical population (in 1950) for total population size and the degree of specialization in 1990 for the specialization index. The idea is that lagged levels of total population size and specialization are correlated to the current levels of spatial variables, but they are assumed not to influence productivity and wages today. Further, in some estimations we also employ the soil PH as extra instrument to check for the validity of our instruments (Combes et al., 2010; Rosenthal and Strange, 2008).

Table 2 shows the results of the pool estimation of model (1). Columns (1) and (2) show the OLS estimates, in column (1) only time, region and sector dummies being inserted as control variables, while in column (2) we enter the full set of control variables included in equation (1). Column (1) shows that the variables capturing agglomeration effects have significant and positive coefficients, with an elasticity of 6.5% of wages with respect to total population size and of 1.1% with respect to the specialization index. On introducing the entire set of control variables specified in equation (1) the elasticity of wages with respect population drops sharply, from 6.5% to 3.9%, while that of specialization remains below these figures (1.6%). Nonetheless, they are still sizeable effects, indicating that agglomeration externalities are at work also in a typical developing country like Ecuador, even if the magnitude of the impacts is lower than those found for the emerging developing countries in the case of population size (see Chauvin et al., 2017; Combes et al. 2015; Duranton, 2016). The magnitude of the specialization index is consistent with that found in other studies based on a similar definition of sector specialization at the local level (Matano and Naticchioni, 2012; Mion and Naticchioni, 2009).

Columns (3) of Table 2 presents the IV estimates. The results show a non-significant difference with respect to the OLS estimates: the elasticity of wages with respect to population is 3.7%, while the elasticity for specialization is around 1%. This is consistent with previous empirical findings indicating that endogeneity does not appear to be a major concern in the analysis of the agglomeration impacts on wages (Combes et al., 2010, Matano and Naticchioni, 2012, Melo et al. 2009). Column (4) of Table 2 presents the same estimates as column (3) with the addition of the soil PH as an extra instrument to test the validity of the instruments. According to the Hansen test (p-value=0.489), we cannot reject the null of the exogeneity of the instruments. Finally, in column (5) previous estimates are replicated excluding from the sample the two largest cities, Quito and Guayaquil, in order to trace any non-linearity in the relationship between spatial variables and wages. The results show a marked increase in the elasticities of wages with respect to city size, rising from 3.8% to 9.1%. This outcome is indicative of decreasing returns to city size. As for specialization, there is no longer a wage benefit for working in a highly specialized industrial area when excluding the two biggest cities, suggesting a key role of the best cities in exploiting these agglomeration advantages.[24]

[Table 2 around here]

4.2 Agglomeration effects and informality

In this section, we take into account the dimension of the informal economy and directly test the role of informality within the relationship between spatial agglomeration and wages for a developing country. We restrict the analysis sample to a span between the second quarter of 2007 and the last quarter of 2015, because only throughout this period is the definition of informality consistent over time. To this end we estimate the following equation:

$$\begin{aligned}\log(W_{i,t}) = \alpha + \delta * \log(Pop_{c(i),t}) + \theta * \log(Spec_{c(i),s(i),t}) + \gamma * \log(Pop_{c(i),t}) * Inf_{i,t} + \\ + \lambda * \log(Spec_{c(i),s(i),t}) * Inf_{i,t} + \rho * Inf_{i,t} + X'_{i,t} * \beta + \mu * dfirmsize_{i,t} + S_{s(i)} + A_{c(i)} + \\ + R_{c(i)} + T_t + \varepsilon_{i,t}\end{aligned} \tag{2}$$

Where $Inf_{i,t}$ is a dummy that takes on a value of 1 if the worker is employed in the informal sector and 0 otherwise. This variable is interacted with our two measures of

[24] As for the control variables, as expected the wages show a concave relationship with age. Moreover, females and workers residing in rural areas within FUAs are penalized, while wages are higher for those who have previously migrated. Further, wages are higher for employees than for the self-employed and increase with firm size and the workers' educational level and occupation skill intensity (not shown).

agglomeration: $log(Pop_{c(i),t})*Inf_{i,t}$ and $log(Spec_{c(i),t})*Inf_{i,t}$. All other variables are the same as defined in equation (1). The parameter ρ captures the difference in average wages between informal and formal workers, all other factors held constant, while parameters γ and λ capture the differential impact on wages due to spatial externalities for informal workers with respect to formal workers. Thus, the significance of the interaction terms will afford insight into the spatial agglomeration impact on wages in the presence of a dual economy.

Table 3 shows the results of this estimation. Columns (1) and (2) present the estimates by OLS, where again in column (1) the estimate includes as control variables only sector, region and time dummies, while in column (2) the full set of control variables specified in equation 2 are included in the estimation. First of all, observing the coefficient related to the informal dummy it is seen that on average, and in line with the descriptive statistics, informal workers are harshly penalized in terms of wages. In fact, the percentage difference in the expected wage between formal and informal workers is around -6/-7%. When taking into account the heterogeneity in the agglomeration impacts between formal and informal workers, the OLS results show a general penalization for informal workers in gains from spatial externalities in terms of both urbanization externalities and specialization (-3% and -2.5% respectively in column (1)), which is however reduced when further controls are added into the estimation (-1.6% and -0.7% respectively).

Column (3) of Table 3 shows the IV estimates. Compared to the OLS estimates, we observe no differences in the estimates for urbanization externalities, where the elasticity of wages with respect to total population size stands at 3.5% for formal workers again being lower by 1.6% for informal workers. As for specialization, the difference with respect to the OLS estimates is important, since on taking into account the endogeneity of the relationship between wages and spatial variables the benefits arising from working in locally specialized areas for informal workers disappear. Indeed, there is a slight wage penalization (around -1.8%).[25] In the next section we will offer a possible explanation for this. For formal workers the wage elasticity with respect to specialization is 2.4%. Column (4) of Table 3 shows the IV estimates using the soil PH as an extra instrument. The instruments of agglomeration variables are

[25] Testing the sum of the main effect and the interaction with informality for the specialization variable in column (3), we reject the null hypothesis that the total effect is not significant, while we cannot reject the null hypothesis that the total effect is not statistically different from -0.01.

exogenous, as confirmed by the Hansen test p-value (0.295), while the estimates remain unchanged. Finally, column (5) shows the IV estimates of column (4) excluding Quito and Guayaquil. In line with what we observed in Table 2, there is evidence of decreasing returns of agglomeration economies to city size, since the elasticity of wages with respect to total population size is 7.7% for the formal and 5.8% for the informal workers. As far as specialization is concerned, we now find a small wage premium for formal workers, albeit even smaller than for the sample with all cities: when excluding Quito and Guayaquil the wage premium for formal workers stands at 1.1%, while for informal workers there is still a slight negative impact (-1.2%). Interestingly, the parameter for informality remains negative, but is no longer significant, which is indicative of the strong effect of informality in the two largest cities.

[Table 3 around here]

Summing up, these results show that informal workers do not enjoy the same wage premium from agglomeration externalities as the formal workers and that they are severely penalized in the largest cities. These findings are at odds with the finding for Colombia by Duranton (2016), who, with an indirect test, finds higher benefits from urbanization externalities for informal workers. He interprets his results suggesting that the income of informal workers is more directly tied to local housing and transportation costs, which make them better able to reap the benefits of agglomeration externalities. Although the same mechanism may also be at play in Ecuador, in this case (and contrary to the case of Colombia), there is evidence of higher returns to agglomeration externalities for the more educated workers (Table A1 in the Appendix) – a finding in line with Bacolod et al. (2009), Glaeser and Resseger (2010) and Wheeler (2001). Since the more educated workers are more likely to be employed in formal occupations,[26] this can (at least partially) explain the higher returns to spatial externalities for formal workers that we detected.[27] We will now go on to present a more detailed picture of the mechanisms through which agglomeration externalities impact on formal and informal workers' wages.

[26] In particular, in our sample, 86% of the high-educated workers are employed in formal jobs, while only 14% are employed in informal jobs. 54% of the low-educated workers are employed in formal jobs, 46% in informal jobs. Moreover, high-educated workers account for 25% of the sample.

[27] In fact, the higher returns to agglomeration externalities for higher educated workers probably offset more than proportionally – if at play - the mechanism observed by Duranton (2016) for informal workers.

4.3 Analysis of the channels of the spatial wage premium: sorting, matching, learning and informality

In this section, we aim to shed light on the channels behind the impacts found so far for formal and informal workers: we analyze the role of sorting, matching and learning mechanisms in the spatial wage dynamics (Combes and Gobillon, 2015; Puga, 2010). To this end, we make use of the data of a panel subsample of our original dataset. The ENEMDU survey is designed to sample families living in specific buildings/dwellings (*viviendas*), resampling of 25% of them being performed quarterly, following a 2-2-2 panel structure: a building/dwelling may be sampled for two consecutive quarters, then left for the two successive quarters, and subsequently again sampled for two more quarters. This means that if a family has not moved house, it could be interviewed at most four times, covering a total time span of six quarters. We take the opportunity of the survey structure to identify working members of households sampled more than once and analyze their wage dynamics.[28] As in the previous sections, we focus on workers aged between 15-64, with data on formality and informality consistent over time. We delete outliers[29] and extreme observations in terms of real wages, hours worked and real wages per hour worked. With this procedure we arrive at a panel of 70,029 observations for 31,200 workers residing in the same place.

To the best of our knowledge this is the first time that the channels behind the spatial wage premium have been analyzed for a developing country by means of a panel dataset. Nonetheless, we must point out that the results of this analysis are limited to the short-run outcomes of the wage dynamics of stayers. Besides, we suggest taking these results with some caution due to the relatively small size of the sample. Since our aim is to understand the differences in the outcomes between formal and informal workers, we present the results separately for these workers' categories.

First of all, we characterize how sorting distributes across space. We define high/low populated (specialized) areas on the basis of the time-invariant median of population (specialization) in the original database. As proxy for individual skills we use individual fixed effects retrieved from a panel estimation where wages are

[28] More specifically, in order to identify household persons interviewed more than once, we select the persons living in the same building/dwelling –included in the panel subsample-, belonging to the same family, having the same position inside the family, with same sex, ethnicity, birth city and (should he/she have previously migrated) reporting the same city as the last place he/she has been living in.

[29] In particular, we identify outlier workers as those reporting data non-consistent over time regarding age, education, gender and previous migration status.

regressed on the set of observable characteristics used as control variables in the main analysis. Table 4, shows the results of the average skills in high/low-populated areas and high/low-specialized areas, for all workers (Column (1)), and for formal and informal workers (Columns (2) and (3)).

On observing the skills distribution across low- and high-populated areas (Table 4, top panel), we find evidence of positive sorting. In fact, the fixed effect average is 0.047 in high-populated areas, greater than in low-populated areas: -0.095. It is worth noting that this sorting effect essentially holds for formal workers, who see a 0.129 increase in fixed effects from low- to high-populated areas. Informal workers are characterized on average by lower unobservable skills than formal workers. Even though there is also an increase in the average of the fixed effects passing from low- to high-populated areas (+0.035), it is marginal compared to that of formal workers (less than one third of the increase). Hence, the higher wage premium (due to population) detected for formal workers is partially due to the sorting of skilled workers in high-populated areas. For informal workers, any wage premium that may occur due to population is not essentially driven by individual sorting.

As for specialization, Table 4 (bottom panel) shows that on average there is mild evidence, if any, of negative sorting of skilled workers along the specialization dimension, since the difference between high and low specialized areas is marginal (-0.021). On analyzing the different workers' categories separately, it is seen that there is no sorting for formal workers, while the informal ones display a marked reduction in skills level passing from lower to higher specialized areas (-0.095). Therefore, the wage penalization observed in previous estimations for informal workers can be explained by the negative sorting of skills for this category of workers (which might more than offset other positive channels).

Summing up, the picture of skills sorting shows that individuals with higher unobservable skills are generally employed in formal jobs and sort into larger cities, which helps explain part of their urban wage premium.

[Table 4 around here]

We now go on to investigate whether the wage premium in high-populated/-specialized areas in the short run can be attributable to better job-matches between workers and firms or to learning mechanisms generated while remaining employed in the same job category. To analyze this issue, we use the data on the specific type of occupation of the worker, as we cannot know if he or she remains in or changes firm. Therefore, our analysis evaluates the impact on wages of a change in type of occupation, which may well proxy a better match between workers and firms in terms of tasks, while not addressing the impact of job change across firms within the same occupation, which might reflect a better match in terms of alternative dimensions, such as wages, working conditions', etc. The focus of our analysis then lies in detection of a wage premium derived from a specific matching linked to the type of occupation, which may occur across firms and also within a firm. We build a dummy signalling job-change with a value equal to 1 when a worker changes occupation. We consider 2- and 3-digit level occupations defined according to the ISCO88 classification.[30] We then perform a wage regression of the log of the real wage per worked hour on the same set of observable characteristics used in the first part of the analysis (equation 1), now expanded with a job-to-job dummy, also interacted with the informal status of the worker, as follows:

$$\log(W_{i,t}) = \alpha + \delta * JTJ_{i,t} + \theta * (JTJ_{i,t} * Inf_{i,t}) + \gamma * Inf_{i,t} + X^{'}_{i,t} * \beta + \mu * dfirmsize_{i,t} + \\ + S_{s(i)} + A_{c(i)} + R_{c(i)} + T_t + \varepsilon_{i,t} \quad (3)$$

where i denotes the worker, c the city, s the sector and t the time. $Log(W_{i,t})$ is the log of real wage per hour of worker i, at time t. $JTJ_{i,t}$ is the dummy for the change in occupation at time t, while $JTJ_{i,t}*Inf_{i,t}$ is the interaction term between the job change dummy and the informal status of worker i at time t. The other control variables are the same as in equation (1) and (2), except for the occupation dummies, which are now introduced at 2- or 3-digit level according to the relative job-to-job digit level considered. Our parameters of interest are δ and θ, which indicate the elasticities of wages with respect to a job change (compared to no job change) and its differential impact should the worker's status be informal. We run an OLS regression, because for most individuals we have only two observations available, with one lost when the JTJ dummy is computed. Standard errors are clustered at the individual level.

[30] It is worth noting that from 2013 to 2015 the ISCO codification applied was ISCO08. Therefore, using the appropriate correspondence table and the technique applied by the OECD Employment Outlook 2017 illustrated in Annex 3.A4, we map the ISCO08 into the ISCO88 and obtain a homogenous ISCO88 classification for the entire period of the analysis. Nonetheless, we carry out a robustness check to take into account this change in classification over time.

With this specification we can see whether wages react positively to a change in occupation, thus suggesting that the wage increases due to a better quality match as compared with remaining in the same kind of occupation.[31] If, on the contrary, there is a penalization due to job change, it means that workers gain relatively more remaining in the same occupation, thus suggesting the presence of learning effects. We perform this estimation for all workers and for workers in high-populated or high-specialized areas separately in order to see if these mechanisms act differently in such areas, thus contributing to an understanding of the channels behind the spatial wage premium. In addition, we decided to focus on workers not changing their formal or informal status when going through job change. In this way we avoid confusing the impact of an occupation change with that of a change in formality status.[32]

Table 5 shows the results. Columns (1), (2) and (3) present the estimates when defining the occupation change at the 2-digit level, columns (4), (5) and (6) when defining the occupation change at the 3-digits level.[33]

As for job-quality match, this appears to be a channel for the wage increase of formal workers, especially in high-populated areas. In fact, the elasticity of wage to job change is in general 2.9%, and increases to 3.7% in high-populated areas when considering a change at the 2-digit level. At the 3-digit level the wage premiums are similar. Hence, for formal workers, job changes entail a higher wage premium compared to remaining in the same job, especially in high-populated areas. This outcome points to the matching channel as one of the factors behind the urban wage premium for this category of workers. In high-specialized areas, the job change dummy still has a positive impact, but smaller (around 2/2.7%) and close to the average.

[31] As for descriptive statistics of job changes at 2-digits level, it is worth noting that around half of job changes for both formal and informal workers represent an upgrade in the occupational position.

[32] Observations accounting for workers who change status between formal and informal or the other way around represent 16% of the total number of observations. We carried out a robustness check using all sample observations and the results remain robust, although the impacts are lower in magnitude (Table A2 in the Appendix).

[33] As previously mentioned, there was an interruption in the ISCO data classification between 2012 and 2013. We therefore mapped the ISCO08 classification into the ISCO88 classification for the years 2013, 2014 and 2015. Since in this section we estimate the impact of a job change on wages according to ISCO classification, we decided to run a robustness check using the original ISCO classification (ISCO88 up to 2012 and ISCO08 as from 2013) to compute the job change, while at the same time excluding the observations affected during the change period. The results are shown in Table A3 in the Appendix and remain robust to the classification considered.

As for the informal workers, the patterns are different: on average a change in occupation penalizes wages compared to remaining in the same job: the total effect is -1.2% in column (1) and -3.9% in column (4). Moreover, the wage penalty increases with increasing size of the FUAs (total effect equal to -2.9% and -5.1% in column (2) and (5) respectively). This suggests that for informal workers the channels through which spatial externalities exert an effect on wages is through learning mechanisms. In highly specialized areas, the total net effect of a job change for informal workers is not statistically different from 0.

Overall, if we consider that informal workers are generally characterized by lower skills than formal workers, these findings are consistent with those in Matano and Naticchioni (2016), where it is seen that in high density areas skilled workers gain relatively more from job change, while unskilled workers benefit from positive knowledge spillovers.

[Table 5 around here]

To sum up, analysis of the channels through which spatial wage premiums arise shows a difference between formal and informal workers. As for the urban wage premium, the impacts for formal workers are driven by both the sorting of high (unobserved) skilled workers into high-populated areas and by a better quality of job matching. On the other hand, for informal workers the gains from agglomeration are mostly driven by positive knowledge spillovers when they are employed in the same occupation. As for specialization, the channels driving the previous results still seem to lies in job-quality matching for formal workers, while for informal workers there is no significant difference between job-matching effects and learning effects within an occupation. In addition, there is evidence of negative sorting for this category of workers, to the effect that those characterized by lower unobservable skills sort into highly-specialized areas, which help to explain the net negative impact detected for this category of workers.[34]

[34] One last important channel to investigate is the one regarding the experience acquired in the biggest cities as shown in De la Roca and Puga (2017). In this case, it is not possible to correctly evaluate the relevance of this issue. In fact, in our data we have information only on whether someone has migrated in the past combined with information about the last place of residence. Therefore, we are not able to trace the migration pattern of migrants, because we cannot know whether the migrant has lived in a big city and for how long. Nonetheless, we ran the estimation of

5. Conclusions

In this paper we have explored the role of agglomeration economies, in terms of total population size and sector specialization at the local level, on the wages of workers for a typical developing country, Ecuador. We have found positive and significant impacts on wages deriving from spatial agglomeration. In particular, the elasticity of wages with respect to total population size is as great as 3.8%, while with respect to specialization it is 0.9%. Moreover, we have also addressed the role of the informal economy in this relationship, analyzing the heterogeneity of the spatial wage premium across formal and informal workers in order to take into account the interaction between spatial agglomeration and the presence of a dual labour market – a common characteristic of most developing countries. Our findings show that informal workers are penalized, since the benefits accruing from spatial externalities are reduced and, in the case of specialization, even cancelled. This shows the importance of considering the role of the duality of the labour market in addressing the relationship between wages and spatial externalities. Furthermore, we also sought to identify the channels through which these wage gains occur across worker categories. The results show evidence of sorting and higher gains deriving from better job match in highly-populated areas for formal workers, as opposed to learning advantages for informal workers. In highly specialized areas, the wage premium due to matching effects still holds for formal workers, while the learning and matching effects are similar for informal workers. In addition, the latter appear to be negatively sorted into these areas, thus explaining the wage penalization found for this category of workers along the specialization dimension. Taking into account that formal workers are relatively higher skilled and educated than informal workers, these findings are in line with the previous literature showing better gains due to good matches for high-skilled workers and greater advantages due to learning externalities for low-skilled workers (Matano and Naticchioni, 2016).

---

model 5 adding a migrant dummy and the interaction with the biggest cities as migrant's last place of residence, finding higher wages on average for migrant workers, but a non-significant interaction with the big cities dummy. The main findings remain confirmed. These estimations are available upon request. In any case, we would like to stress that it has been shown that this mechanism probably acts as complementary to the other mechanisms detected (see Matano and Naticchioni, 2016); we are therefore quite confident that failing to correctly investigate this aspect should not significantly affect our results. We leave this issue open for further research.

References

Au, C.C., Henderson, V., 2006. Are Chinese Cities Too Small?, Review of Economic Studies 73, 549–76.

Ahrend, R., Farchy, E., Kaplanis, I., Lembcke, A., 2017. What Makes Cities More Productive? Evidence from Five OECD Countries on the Role of Urban Governance, Journal of Regional Science 57(3), 385-410.

Bacchetta, M., Ernst, E., Bustamante, J., 2009. Globalization and Informal Jobs in Developing Countries: a joint study of the ILO and WTO, WTO ISBN: 978-92-870-3691-9, ILO ISBN: 978-92-2-122719-9.

Bacolod, M., Blum, B. S., Strange, W. C., 2009. Skills in the city, Journal of Urban Economics 65 (2), 136-153.

Bernedo Del Caprio, M., Patrick, C., 2017. Agglomeration and Informality: Evidence from Peruvian firms, Andrew Young School of Policy Studies, Research Paper Series, 17-13.

Ciccone, A., 2002. Agglomeration effects in Europe, European Economic Review 46 (2), 213– 227.

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

Chauvin, J. P., Gleaser, E., Ma, Y., Tobio, K., 2017. What is different about urbanization in rich and poor countries? Cities in Brazil, China, India, and United States, Journal of Urban Economics 98, 17-49.

Combes, P.P., 2000. Economic structure and local growth: France, 1984 -1993, Journal of Urban Economics 47 (1), 329–55.

Combes, P.P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: sorting matters!, Journal of Urban Economics 63 (1), 723–742.

Combes, P.P., Duranton, G., Gobillon, L., Roux, S., 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects, in: Edward L. Glaeser (Eds.), Agglomeration Economics, Chicago, Chicago University Press, 15–65.

Combes, P.P., Démurger, S., Li, S., 2015. Migration externalities in Chinese cities, European Economic Review 76 (1), 152-167.

Combes, P. P., Gobillon, L., 2015. The Empirics of Agglomeration Economies, in: J.V. Henderson and J.-F. Thisse (eds.), Handbook of Regional and Urban Economics, Vol. 5 (Amsterdam: Elsevier-North-Holland), 247–348.

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

Duranton, G., 2009. Are Cities Engines of Growth and Prosperity for Developing Countries, in: M. Spence, P. Annez, and R. Buckley, (Eds.), Urbanization and Growth, the Commission Growth and Development, Washington, D.C.

Duranton, G., 2016. Agglomeration effects in Colombia, Journal of Regional Science 56 (2), 210-238.

Duranton, G., Puga D., 2004. Micro-foundations of urban agglomeration economies, in: J.V. Henderson and J.-F. Thisse (eds.), Handbook of Regional and Urban Economics, Vol. 4 (Amsterdam: Elsevier-North-Holland), 2063–2117.

Fields, G., 1990. Labour Market Modelling and the Urban Informal Sector: Theory and Evidence, in: D., Thurnham, Salome, B. and Schwarz, A. (Eds.), The Informal Sector Revisited, Paris, OECD.

Garcia, G.A., 2016. Agglomeration Economies in the Presence of an Informal Sector: the Colombian Case, CEDE Working paper, 15.

Glaeser, E. L., Henderson, J.V., 2017. Urban economics for the developing World: An introduction, Journal of Urban Economics 98 (1), 1-5.

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

Glaeser, E. L., Resseger, M. G., 2010. The complementarity between cities and skills, Journal of Regional Science 50 (1), 211–244.

Harris, J., Todaro, M., 1970. Migration, Unemployment, and Development: A Two Sector Analysis, American Economic Review 107 (1), 126-142.

Harris, J., 2014. The messy reality of agglomeration economies in urban informality: Evidence from Nairobi's handicraft industry, World Development 61, 102–113.

ILO, 2013. Measuring Informality: A statistical manual on the informal sector and informal employment, Book ILO, ISBN: 9789221273882.

INEC, 2015. Actualización metodológica: Empleo en el sector informal y la clasificación de los ocupados según sectores, Nota técnica, Quito, Ecuador.

Khamis, M., 2012. A note on informality in the labour market, Journal of International Development 24 (1), 894-908.

La Porta, R., and Shleifer, A., 2014. Informality and Development, Journal of Economics Perspectives 28 (3), 109-126.

Lall, S.V., Shalizi, Z., and Deichmann, U., 2004. Agglomeration Economies in the Indian Industry, Journal of Development Economics 73, 643-673.

Maloney, W.F., 2004. Informality Revisited, World Development 32(7), 1159-1178.

Marshall, A., 1890. Principles of Economics, Book London: MacMillan.

Matano, A., Naticchioni, P., 2012. Wage distribution and the spatial sorting of workers, Journal of Economic Geography 12 (2), 379-408.

Matano, A., Naticchioni, P., 2016. What drives the urban wage premium? Evidence along the wage distribution, Journal of Regional Science 56 (2), 191–209.

Melo, P. C., Graham, D. J., Noland, R. B., 2009. A meta-analysis of estimates of urban agglomeration economies, Regional Science and Urban Economics 39 (3), 332–342.

Mion, G., Naticchioni, P., 2009. The spatial sorting and matching of skills and firms, Canadian Journal of Economics 42 (1), 28–55.

Obaco, M., Royuela V., Vítores, X., 2017. Computing functional urban areas using a hierarchical travel time approach: An applied case in Ecuador, AQR-IREA working paper (5), 46.

OECD, 2013. Definition of functional urban areas (FUA) for the OECD metropolitan database, Paris: OECD Publishing.

OECD, 2017. Employment Outlook, Annex Chapter 3.

Overman, H., and Venables, A., 2005. Cities in the Developing World, CEP Discussion Paper, n.695, London Perry.

Puga, D., 2010. The magnitude and causes of agglomeration economies, Journal of Regional Science 50 (1), 203-219.

Rosenthal, S.S., and Strange W.C., 2004. Evidence on the nature and sources of agglomeration Economies, in: J.V. Henderson and J.-F. Thisse (Eds.), Handbook of Regional and Urban Economics, Vol. 4 (Amsterdam: Elsevier-North-Holland): 2119–2171.

Rosenthal, S.S., and Strange W.C., 2008. The attenuation of human capital spillovers, Journal of Urban Economics 64, 373-389.

Royuela, V., and Castells-Quintana, D., 2015. International migrations and urbanisation: 1960–2010, International Journal of Global Environmental Issues 13, 150-169.

Wheeler, A., 2001. Search, sorting, and urban agglomeration, Journal of Labor Economics 19 (4), 880–898.

Tables

Table 1: Descriptive Statistics. Real Wage per Hour and Sample Composition by Worker Categories

	Mean Real Wage per Hour (US dollars)				Sample Composition
Worker data	**2005**	**2010**	**2015**	**All**	**%**
Total	1.86	1.96	2.52	2.17	100
Gender					
Male	1.91	2.04	2.62	2.26	60.59
Female	1.77	1.84	2.36	2.04	39.41
Ethnicity					
Indigenous	1.31	1.44	1.81	1.65	4.75
White	2.26	2.29	3.02	2.44	4.03
Mestizo*	1.87	1.98	2.58	2.21	87.13
Black	1.43	1.64	2.14	1.83	2.61
Mulatos**	1.40	1.72	2.29	1.83	1.48
Education level					
None	1.09	1.23	1.51	1.33	1.90
Literacy	1.28	1.32	1.57	1.44	0.29
School	1.39	1.54	1.97	1.69	33.45
High School	1.75	1.86	2.34	2.04	41.77
Technical	2.23	2.50	3.22	2.73	1.03
University	2.90	2.89	3.61	3.17	20.88
Post-Univ. Degree	4.38	4.92	6.34	5.43	0.68
Occupation					
Legislators, Professionals and Technicians	3.23	3.45	4.40	3.76	12.84
Clerks, Service Workers and Skilled Agricultural and Fishery Workers	1.73	1.91	2.32	2.06	35.27
Crafts and Trade Workers, Plant and Machine Operators, Elementary Occupations	1.57	1.67	2.18	1.86	51.88
Job Category (1)					
Dependent workers	1.77	2.01	2.70	2.24	56.45
Self-employed	1.96	1.90	2.28	2.09	43.55
Job Category (2)					
Formal	-	2.38	3.02	2.68	62.10
Informal	-	1.52	1.77	1.66	37.90
Economic Sector					
Agriculture, Fishing	1.43	1.57	1.91	1.71	9.25
Mining and Quarrying	2.89	3.32	4.41	3.83	0.52
Manufacturing	1.79	1.97	2.50	2.16	15.82
Electricity, Gas and Water Supply	2.60	2.60	3.24	3.05	0.19
Construction	1.75	1.90	2.44	2.10	9.82
Wholesale and Retail Trade	1.88	1.94	2.45	2.12	27.49
Hotels and Restaurants	1.76	1.88	2.38	2.10	6.44
Transport, Storage and Communications	2.04	2.01	2.53	2.25	9.03
Financial Intermediation	2.77	3.33	4.34	3.68	1.43
Real Estate, Renting and Business Activities	2.41	2.43	2.95	2.65	5.95
Private Education, Health and Social Work	2.47	2.68	3.55	2.93	5.08
Other Service Activities	1.46	1.60	2.19	1.78	8.98
Firm size					
<100 workers	1.81	1.86	2.33	2.05	86.54
>=100 workers	2.25	2.67	3.56	2.95	13.46
Region					
Andean	2.02	2.11	2.67	2.31	52.86
Coastal	1.68	1.78	2.29	1.99	43.20
Amazonian	2.01	2.11	2.56	2.34	3.93
Area					
Urban	1.90	2.03	2.60	2.23	89.09
Rural	1.32	1.51	2.00	1.71	10.92
Migrant					
Yes	1.87	2.04	2.56	2.24	31.00
No	1.85	1.94	2.49	2.14	69.00
Total Obs.	29,351	37,329	54,208	408,197	

Notes: Occupation is defined according to ISCO88, sector according to ISIC 3.1. *Mestizo is the category for mixed white/indigenous race. **Mulato is the category for mixed white/black race.

Table 2: OLS and IV regressions of wages on the spatial variables. Dependent variable: log of real worker's wage per hour.

	(1)	(2)	(3)	(4)	(5)
	OLS	**OLS**	**IV**	**IV**	**IV**
Log(Pop)	0.065***	0.039***	0.037***	0.038***	0.091***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
Log(Specialization)	0.011***	0.016***	0.009***	0.009***	-0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
Female dummy		-0.157***	-0.157***	-0.157***	-0.172***
		(0.002)	(0.002)	(0.002)	(0.003)
Age		0.034***	0.034***	0.034***	0.035***
		(0.000)	(0.000)	(0.000)	(0.001)
Age squared		-3.63e-04***	-3.63e-04***	-3.63e-04***	-3.81e-04***
		(6.19e-06)	(6.19e-06)	(6.19e-06)	(7.56e-06)
Migrant dummy		0.037***	0.037***	0.037***	0.035***
		(0.002)	(0.002)	(0.002)	(0.003)
Firm size		0.194***	0.195***	0.195***	0.212***
		(0.002)	(0.002)	(0.002)	(0.003)
Rural dummy		-0.053***	-0.053***	-0.053***	-0.059***
		(0.003)	(0.003)	(0.003)	(0.004)
Ethnicity dummies	no	yes	yes	yes	yes
Education dummies	no	yes	yes	yes	yes
Job category dummies	no	yes	yes	yes	yes
Occupation dummies	no	yes	yes	yes	yes
Sector dummies	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes
Constant	-0.622***	-0.665***	-0.646***	-0.647***	-1.458***
	(0.015)	(0.008)	(0.008)	(0.008)	(0.108)
Observations	408,197	408,197	408,197	408,197	275,138
R-squared	0.13	0.29	0.29	0.29	0.28
Weak identification test (F-value)			48,134	32,029	19,441
Over identification test (p-value)				0.489	0.439

Standard errors clustered at worker level in parentheses *** p<0.01, ** p<0.05, * p<0.1. Occupation is codified according to ISCO 88, two-digit level, sector according to ISIC 3.1, two-digit level. Total population size is instrumented using total population in 1950, while specialization is instrumented using the specialization level in the 1990. Soil PH is used as an extra instrument in column (4) and (5). The two largest cities, Guayaquil and Quito, are excluded from the regression in Column (5).

Table 3: OLS and IV regressions of wages on the spatial variables and interaction with informality. Dependent variable: log of real worker's wage per hour.

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	IV	IV	IV
Log(Pop)	0.058***	0.036***	0.035***	0.035***	0.077***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Log(Pop)*Informal dummy	-0.030***	-0.016***	-0.016***	-0.016***	-0.019***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)
Log(Spec)	0.016***	0.016***	0.024***	0.024***	0.011***
	(0.003)	(0.002)	(0.004)	(0.004)	(0.005)
Log(Spec)*Informal dummy	-0.025***	-0.007*	-0.042***	-0.042***	-0.023***
	(0.005)	(0.004)	(0.006)	(0.006)	(0.007)
Informal dummy	-0.057***	-0.070***	-0.067***	-0.066***	-0.030
	(0.025)	(0.024)	(0.026)	(0.026)	(0.066)
Female dummy		-0.150***	-0.150***	-0.150***	-0.165***
		(0.003)	(0.003)	(0.003)	(0.003)
Age		0.033***	0.033***	0.033***	0.034***
		(0.000)	(0.000)	(0.000)	(0.001)
Age squared		-3.54e-04***	-3.54e-04***	-3.54e-04***	-3.74e-04***
		(7.13e-06)	(7.13e-06)	(7.13e-06)	(8.70e-06)
Migrant dummy		0.033***	0.033***	0.033***	0.031***
		(0.002)	(0.002)	(0.002)	(0.003)
Firm size		0.151***	0.149***	0.149***	0.157***
		(0.003)	(0.003)	(0.003)	(0.004)
Rural dummy		-0.041***	-0.040***	-0.040***	-0.044***
		(0.004)	(0.004)	(0.004)	(0.004)
Ethnicity dummies	no	yes	yes	yes	yes
Education dummies	no	yes	yes	yes	yes
Job category dummies	no	yes	yes	yes	yes
Occupation dummies	no	yes	yes	yes	yes
Sector dummies	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes
Constant	-0.022	-0.220*	-0.201*	-0.203*	-1.070***
	(0.018)	(0.117)	(0.117)	(0.117)	(0.172)
Observations	304,631	304,631	304,631	304,631	208,618
R-squared	0.21	0.32	0.32	0.32	0.30
Weak identification test (F-value)			14,250	10,954	9,296
Over identification test (p-value)				0.295	0.976

Standard errors clustered at worker level in parentheses *** p<0.01, ** p<0.05, * p<0.1. Occupation is codified according to ISCO 88, two-digit level, sector according to ISIC 3.1, two-digit level. Total population size is instrumented using total population in 1950, while specialization is instrumented using the specialization level in the 1990. Soil PH is used as an extra instrument in column (4) and (5). The two largest cities, Guayaquil and Quito, are excluded from the regression in Column (5).

Table 4: Mean Fixed Effects across Space

	All	Formal	Informal
Low Population	-0.095	0.014	-0.251
High Population	0.047	0.143	-0.216
Low Specialization	0.011	0.108	-0.184
High Specialization	-0.010	0.104	-0.279
N. Obs.	70,029	47,976	22,053

Table 5: Job-change impacts on wages by formal/informal workers and differently populated/specialized area

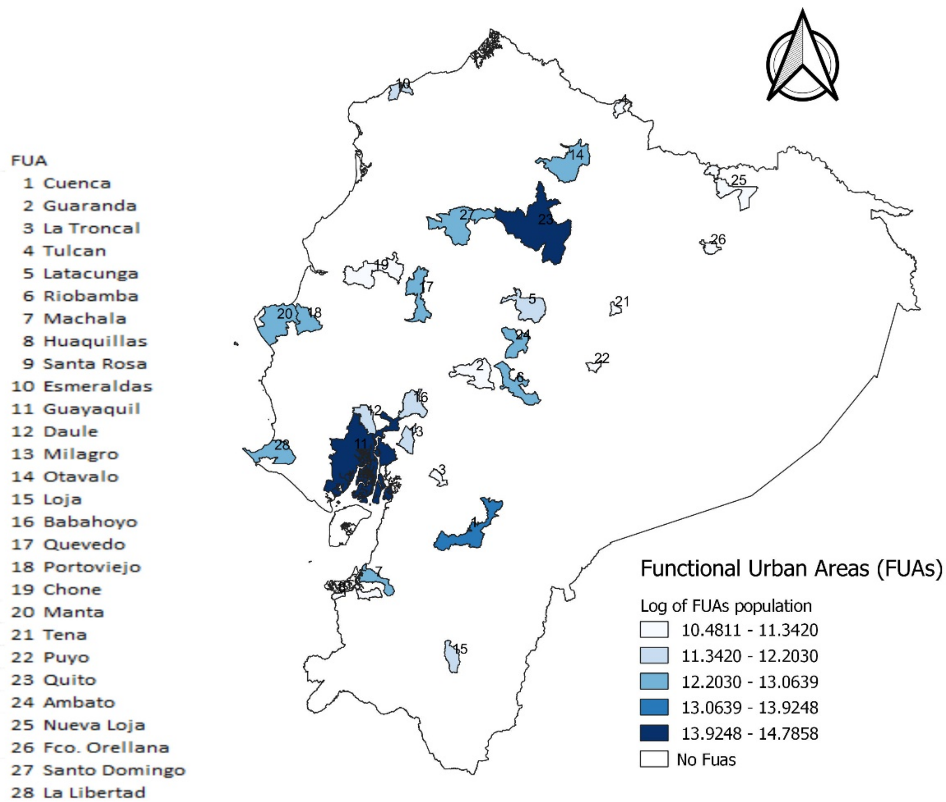
	(1)	(2)	(3)	(4)	(5)	(6)
	All	High-Pop	High-Spec	All	High-Pop	High-Spec
Job change	0.029***	0.037***	0.027***	0.029***	0.035***	0.020**
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)
Job change*Informal	-0.041**	-0.066***	-0.009	-0.068***	-0.086***	-0.030
	(0.016)	(0.022)	(0.023)	(0.015)	(0.020)	(0.022)
Informal	-0.296***	-0.289***	-0.363***	-0.291***	-0.288***	-0.359***
	(0.012)	(0.016)	(0.018)	(0.013)	(0.017)	(0.019)
Female dummy	-0.152***	-0.141***	-0.136***	-0.157***	-0.146***	-0.143***
	(0.008)	(0.009)	(0.010)	(0.008)	(0.010)	(0.011)
Age	0.036***	0.039***	0.037***	0.036***	0.038***	0.036***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age squared	-3.98e-04***	-4.24e-04***	-4.09e-04***	-3.90e-04***	-4.15e-04***	-3.98e-04***
	(2.30e-05)	(2.78e-05)	(3.12e-05)	(2.29e-05)	(2.77e-05)	(3.12e-05)
Migrant dummy	0.043***	0.034***	0.034***	0.042***	0.033***	0.035***
	(0.008)	(0.009)	(0.011)	(0.008)	(0.009)	(0.011)
Firm size	0.161***	0.160***	0.147***	0.162***	0.162***	0.148***
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)
Rural dummy	-0.088***	-0.131***	-0.087***	-0.081***	-0.120***	-0.078***
	(0.016)	(0.023)	(0.023)	(0.016)	(0.023)	(0.023)
Ethnicity dummies	yes	yes	yes	yes	yes	yes
Education dummies	yes	yes	yes	yes	yes	yes
Job category dummies	yes	yes	yes	yes	yes	yes
Occupation dummies	yes	yes	yes	yes	yes	yes
Sector dummies	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes	yes
Constant	0.252	0.175	0.076	0.403*	0.241	0.439***
	(0.167)	(0.183)	(0.200)	(0.227)	(0.240)	(0.112)
Observations	32,659	22,599	17,509	32,659	22,599	17,509
R-squared	0.37	0.39	0.40	0.38	0.40	0.41

Notes: Standard errors clustered at worker level in parentheses *** p<0.01, ** p<0.05, * p<0.1. Occupation is coded according to ISCO88 2-digit level from columns (1) to (3) and ISCO88 3-digit level from columns (4) to (6). The sector is coded according to ISIC 3.1 2-digit level. From column (1) to (3) job change is considered at 2-digit level, from column (4) to (6) job change is considered at 3-digit level.

Appendix

Figures

Figure A1. FUAs in Ecuador



Tables

Table A1: OLS and IV regressions of wages on the spatial variables and interaction with high education. Dependent variable: log of real worker's wage per hour.

	(1)	(2)	(3)	(4)	(5)
	OLS	**OLS**	**IV**	**IV**	**IV**
Log(Pop)	0.048***	0.029***	0.027***	0.027***	0.069***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Log(Pop)*High-education dummy	0.022***	0.015***	0.015***	0.015***	0.013**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)
Log(Spec)	0.018***	0.008***	-0.004	-0.004	-0.008
	(0.003)	(0.002)	(0.004)	(0.004)	(0.005)
Log(Spec)*High-education dummy	0.023***	0.010***	0.048***	0.047***	0.030***
	(0.006)	(0.005)	(0.007)	(0.007)	(0.009)
High-education dummy	0.134***	-0.004	-0.004	-0.003	0.025
	(0.030)	(0.027)	(0.029)	(0.029)	(0.076)
Informal dummy		-0.287***	-0.287***	-0.287***	-0.280***
		(0.003)	(0.003)	(0.003)	(0.004)
Female dummy		-0.152***	-0.152***	-0.152***	-0.166***
		(0.003)	(0.003)	(0.003)	(0.003)
Age		0.034***	0.034***	0.034***	0.035***
		(0.000)	(0.000)	(0.000)	(0.001)
Age squared		-3.81e-04***	-3.81e-04***	-3.81e-04***	-4.01e-04***
		(7.14e-06)	(7.14e-06)	(7.14e-06)	(8.71e-06)
Migrant dummy		0.027***	0.027***	0.027***	0.026***
		(0.002)	(0.002)	(0.002)	(0.003)
Firm size		0.156***	0.157***	0.157***	0.163***
		(0.003)	(0.003)	(0.003)	(0.004)
Rural dummy		-0.058***	-0.058***	-0.058***	-0.063***
		(0.004)	(0.004)	(0.004)	(0.004)
Ethnicity dummies	no	yes	yes	yes	yes
Job category dummies	no	yes	yes	yes	yes
Occupation dummies	no	yes	yes	yes	yes
Sector dummies	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes
Time dummies	yes	yes	yes	yes	yes
Constant	-0.181***	0.066	0.097	0.095	-0.748***
	(0.017)	(0.116)	(0.115)	(0.115)	(0.170)
Observations	304,631	304,631	304,631	304,631	208,618
R-squared	0.19	0.31	0.31	0.31	0.29
Weak identification test (F-value)			4,714	3,770	9,233
Over identification test (p-value)				0.263	0.925

Standard errors clustered at worker level in parentheses *** p<0.01, ** p<0.05, * p<0.1. High-educated workers are those with education level higher than high school (i.e. technical, university and post-university education). Occupation is codified according to ISCO 88, two-digit level, sector according to ISIC 3.1, two-digit level. Total population size is instrumented using total population in 1950, while specialization is instrumented using the specialization level in the 1990. Soil PH is used as an extra instrument in column (4) and (5). The two largest cities, Guayaquil and Quito, are excluded from the regression in Column (5).

Table A2: Job change impacts on wages by formal/informal workers and size area, including workers who change status between formal and informal when having job change.

	(1)	(2)	(3)	(4)	(5)	(6)
	All	High-Pop	High-Spec	All	High-Pop	High-Spec
Job change	0.021***	0.028***	0.020**	0.020***	0.026***	0.013
	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)
Job change*Informal	-0.034**	-0.053***	-0.012	-0.057***	-0.061***	-0.020
	(0.014)	(0.018)	(0.020)	(0.014)	(0.018)	(0.020)
Informal	-0.253***	-0.250***	-0.300***	-0.245***	-0.252***	-0.296***
	(0.010)	(0.013)	(0.015)	(0.011)	(0.014)	(0.016)
Observations	38,829	26,272	20,505	38,829	26,272	20,505
R-squared	0.34	0.36	0.37	0.35	0.37	0.38

Notes: Standard errors clustered at worker level in parentheses *** p<0.01, ** p<0.05, * p<0.1. The other control variables are a female dummy, age, age squared, firm size, and dummies for education, ethnicity, occupation (ISCO88 2-digit level, in columns (1) to (3) and ISCO88 3-digit level in columns (4) to (6)), sector (ISIC 3.1 2-digit level), migrant status, rural area, macro-areas and time. From column (1) to (3) job change is considered at 2-digit level, from column (4) to (6) job change is considered at 3-digit level.

Table A3: Job change impacts on wages by formal/informal workers and size area. Original ISCO codification and excluding individuals during the change in ISCO versions.

	(1)	(2)	(3)	(4)	(5)	(6)
	All	High-Pop	High-Spec	All	High-Pop	High-Spec
Job change	0.033***	0.040***	0.029***	0.031***	0.036***	0.019*
	(0.008)	(0.009)	(0.010)	(0.007)	(0.008)	(0.010)
Job change*Informal	-0.063***	-0.088***	-0.030	-0.082***	-0.097***	-0.040*
	(0.018)	(0.023)	(0.025)	(0.017)	(0.022)	(0.024)
Informal	-0.293***	-0.286***	-0.357***	-0.286***	-0.284***	-0.353***
	(0.013)	(0.017)	(0.020)	(0.014)	(0.018)	(0.021)
Observations	28,341	19,512	15,102	28,341	19,512	15,102
R-squared	0.37	0.39	0.40	0.38	0.40	0.41

Notes: Standard errors clustered at worker level in parentheses *** p<0.01, ** p<0.05, * p<0.1. The other control variables are a female dummy, age, age squared, firm size, and dummies for education, ethnicity, occupation (ISCO88 2-digit level, in columns (1) to (3) and ISCO88 3-digit level in columns (4) to (6)), sector (ISIC 3.1 2-digit level), migrant status, rural area, macro-areas and time. From column (1) to (3) job change is considered at 2-digit level, from column (4) to (6) job change is considered at 3-digit level.