UNIVERSITAT DE BARCELONA

Three Essays on Agglomeration Economies in Ecuador

Moisés Lenyn Obaco Álvarez

PhD in Economics

Three Essays on Agglomeration Economies in Ecuador

Moisés Lenyn Obaco Álvarez

2018

PhD in Economics | Moisés Lenyn Obaco Álvarez

UNIVERSITAT DE BARCELONA

PhD in Economics

Thesis title:

Three Essays on Agglomeration Economies in Ecuador

PhD student:

Moisés Lenyn Obaco Álvarez

Advisors:

Vicente Royuela

Alessia Matano

Date:

July 2018

UNIVERSITAT DE BARCELONA

Preface and acknowledgments:

This PhD thesis has been the fruit of studying at the University of Barcelona since the year 2012, which also includes my master's in Economics at the same university. Thus, I can kindly say that I consider Barcelona as my second home.

With this PhD thesis, a solid professional research experience has been gained, and not only an academic degree is obtained, but also the personal experience to be a better person. Thus, I do not feel able to express in this short writing how important and rewarding is for me to accomplish a goal that I fixed in my mind after finishing my bachelor's Degree in economics at the University of Guayaquil (Ecuador), back in 2012.

Furthermore, the outcome of this PhD thesis does not only reflect my effort. It is also the effort of a lot of people that supports me and believe in my capacity to do it. I will try to express my acknowledgments to all of them.

I want to express all my acknowledgements in chronological order. As I mentioned before, this experience started in 2012 when I left my natal city, Guayaquil-Ecuador. Hence, I want to thank my family in Ecuador: my mother, Fúlvia Margarita Álvarez Romero, my father, Ignacio Obaco Guaycha, my sister, Mercy Geovanna Obaco Álvarez, and my other two brothers. I want also to thank my family in Spain: my sister, María Carmita Obaco Álvarez, and my nephews and nice. In addition, the University of Guayaquil and the Econ. Marina Mero Figueroa for all the support they gave me.

From my Master, I really want to say thanks to all professors, whose background was really important to accomplish this PhD thesis. Also, I would like to mention great friends from my master, especially Daniel Daví Arderius known as the engineer.

From my PhD period, I want to thank the UBeconomics team, and the Department of Econometrics, Statistics and Applied Economics. Special attention to my research group AQR (Regional Quantitative Research Group) to which I am always going to belong to. I must mention that my time as a member of AQR is one of the best experiences that a student can have. Among all the member of AQR, there are my supervisors: Vicente Royuela and Alessia Matano. Both are excellent advisors and extraordinary persons, who showed me that doing a PhD is more than obtaining a degree, it contributes to become a better person. ¡¡Mil gracias a los dos!! Moreover, I want to say thank to my officemate, Hoon Choi, who always was an amazing friend during my PhD period.

I acknowledge the support of the Generalitat de Catalunya through the FI-scholarship for three years of funding. I am very grateful for this. Also, I would like to cite and say thank for all suggestions of all the people in workshops, seminars, and congresses in which these essays have been presented. The essay one, named as chapter two: Functional Urban Areas co-authored with Vicente Royuela and Xavier Vítores, has been presented in the XLI International Conference in Regional Science (2015), the AQR Lunch Seminar (2015), 56th ERSA Congress (2016), the Spatial Clustering Workshop (2016), the 10th STATA Users Meeting (2016), the UB Economics Workshop (2016), and the II Workshop in Economics of the University of Guayaquil (2017). The essay two, named as Agglomeration Effects and Informal Employment co-authored with Alessia Matano and Vicente Royuela, has been presented in XLII International Conference on Regional Science (2016), 57th ERSA Congress (2017), 10th Geoffrey J.D. Hewings Regional Economics Workshop / the 10th Summer Conference in Regional Science (2017), XLIII International Conference on Regional Science (2017) and in Lunch seminar of the Faculty of Spatial Sciences in the Groningen University (2018). The essay three named as Agglomeration Effects and Slums is also coauthored with Alessia Matano and Vicente Royuela.

Finally, thanks to ERSA (European Regional Science Association) and AECR (*Asociación Española de Ciencia Regional*) for allowing me to improve and participate in every event that they did. I also want to thank the referees and the commission for their time and work.

Content

List of Tables

Chapter two

Chapter Three

Chapter Four

Appendix of Chapter Four

List of Figures

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter One: Introduction

This PhD develops an empirical research of agglomeration economies in Ecuador. As any PhD thesis it has the goal to push knowledge forward. We pursue this task by taking into account both a conceptual perspective, deeply studying new dimensions within agglomeration economies, and a practical point of view, analyzing a new case study in the literature: Ecuador.

To present a clear idea of the goal of this PhD thesis and structure of the essays, this introductory chapter summarizes the main ideas of the theoretical framework of reference. Ecuador is also presented to the reader, within an international context.

The structure of this chapter is the following. First, a short introduction on agglomeration economies and the main empirical findings are presented. Next, the general context and the motivation of the essays are introduced. Finally, a short presentation of Ecuador is mentioned, so the readers can be familiar with the case study. Finally, the research objectives and research limitations for this thesis are introduced.

1.1 Agglomeration economies: context and motivation

The notion of agglomeration economies relies on the idea that urban concentration, in terms of both population and firms, generates positive externalities that enhances productivity (Rosenthal & Strange, 2004 p.1). The initial conceptualization of the idea of agglomeration effects is attributed to Marshall (1890). He labels the effects of the concentration in cities as external economies, which are independent of firms and workers, because they work at an aggregated city level. In simple terms, cities might present increasing returns to scale by being spatially concentrated, although their components (both firms and workers) might present constant returns to scale. He details a description of the sources of these advantages, such as information spillovers, local non-traded inputs, and local skilled labor pool. All these Marshall's considerations take into account firms of the same industry that are spatially concentrated as clusters that enjoy increasing returns to scale.

McCann (2001) mentions three kinds of agglomeration economies that might be present in cities: internal returns to scale, economies of localization and economies of urbanization (urbanization externalities). The first ones are internal to the firm and result from a large scale of production. Economies of localization refer to the agglomeration economies which accrue to a group of firms within the same industry located in the same place, such as clusters. Finally, urbanization economies are those economies of agglomeration across different sectors which accrue to firms belonging to different industries again set in the same location (Jacobs 1969).

Duranton and Puga (2004) formalize the micro-foundations of agglomeration economies that justify the mechanisms described by Marshall. These are the learning, matching and sharing channels. The learning mechanism reflects the idea that knowledge spillovers and face-to-face interactions in agglomerated areas enhance human capital. The matching mechanism points out that agglomerated areas offer the possibility to achieve a better match between workers and firms. And the sharing mechanism refers to the advantages generated by sharing indivisible goods, including facilities and risk in new investments, what in turn decreases individuals' and firms' costs. These mechanisms are behind both urbanization and specialization positive spillovers, being the former associated with highly dense areas, and therefore across-industries, while the latter takes place within specific industries located in the same area. In fact, as proposed by McCann (2001), it is hard to separate urbanization and localization economies once the micro-foundations of agglomeration are considered. Also, as Puga (2010) argues, the literature is still far away of distinguishing among the channels behind agglomeration effects.

Many works have provided empirical evidence on the benefits of agglomeration economies (Holmes, 2010). In particular, the literature has shown a positive association between city size and labor market outcomes such as wages and productivity, better quality of life and larger number of amenities (for a review, See Combes and Gobillon, 2015; Roshental and Strange, 2004). Although the literature has paid special attention to the benefits of agglomeration externalities, there is also a growing literature devoted to study the costs associated with living in urban areas, usually labelled as disamenities (Koster and Rouwendal, 2013; Kahn, 2010; Suedekum, 2006).

Despite the variety of approaches, the most studied topic within the agglomeration literature is, by far, the relationship between city size and productivity, often proxied by using worker's wages (Holmes, 2010). The hypothesis is that larger cities tend to offer higher wages, as workers are more productive thanks to agglomeration effects.

Most empirical studies on agglomeration economies have focused on the developed world such as USA (Glaeser and Resseger; 2010; Rosenthal and Strange, 2008; Glaeser and Maré, 2001; Ciccone and Hall, 1996), United Kingdom (D'Costa and Overman, 2014), Norway (Carlsen et al., 2016), France (Combes, 2008; Combes, 2000), Spain (De la Roca and Puga, 2017), Italy (Matano and Naticchioni, 2012; Mion and Naticchioni, 2009), Europe (Camagni et al., 2015; Foster and Stehrer, 2009; Ciccone, 2002), OECD countries (Ahrend et al., 2017), and Korea (Henderson et al., 2001). The overall effect of agglomeration economies in these developed countries ranges between 2 and 5 per cent.

Nevertheless, there are some studies on the developing countries: Brazil and India (Chauvin et al., 2017), China (Combes et al., 2015; Shanzi, 2010; Au and Henderson, 2006), Mexico (Ahrend et al., 2017) and Colombia (Duranton, 2016; Garcia, 2017).[1] The results for this group of countries report a larger size of agglomeration economies, which ranges from 5 to 13 per cent. Consequently, the magnitude of agglomeration effects in developing economies seem to be relatively higher than the one found for the developed economies (Combes and Gobillon, 2015).

However, the studied sample of developing countries is hardly representative of a typical developing economy. In fact, the set of countries analyzed is characterized by large population size and geographical extension. Most of them are labelled as emerging markets due to their active economic performance. Therefore, attention is needed for analyzing the case of more typical developing economies characterized by small population size, geographical extension and lower economic activity (Glaeser and Henderson, 2017, Combes and Gobillon, 2015, Duranton, 2014). Moreover, Glaeser and Henderson (2017) stress that urbanization process in developing

[1] Melo et al. (2009) present a meta-analysis of the studies on agglomeration effects. The share of studies of agglomeration economies in the developing world account just for 7% of the total papers.

countries is different from the one in the developed economies. The urbanization in developing world is characterized by extreme poverty and poor quality of their institutions. Consequently, some attention needs to be paid to both issues. In this thesis we analyze labor market institutions by looking at the role of informality in the labor market on the agglomeration economies. Besides, we take into account poverty by studying the creation of slums in the urban system of a developing economy. We expect that the investigation of these two dimensions can help to explain the lack of association between urban expansion and economic growth in developing countries that has been found in the literature (Chen et al., 2014; Castells-Quintana, 2016). Next, we briefly introduce both the concept of informality and slums.

Informal employment is the results of a lack of adequate labor market institutions. The concept itself is hard to define, and there is not a standard concept and measurement. Moreover, the informal sector is generally associated to lack of legal protection and lower productivity and it is mostly composed by unskilled workers (Bacchetta et al., 2009; Fields, 1990). In this thesis, and according to the 2013 ILO guidelines, informal workers are those workers employed in firms with fewer than 100 employees with no tax identification number (*Registro único de contribuyentes*).[2]

The developing world contains two-thirds of the global population and is the main driver of the growth of the world's urban rate. Nevertheless, this process, that can be associated to economic growth and development, is not exempt of problems. According to the UN-Habitat (2015), 30 per cent of the urban population in the developing world live in slums. A slum household is defined as a group of individuals living under the same roof lacking one or more of the following conditions: access to improved water, access to improved sanitation, sufficient living area, durability of housing and security of tenure. Consequently, poverty in general and slums in particular are important concerns about the quality of life in cities of the developing world. Glaeser (2011) argues that slums might give opportunities to poor people to obtain benefits from agglomeration over time, as they have the opportunity to work in larger cities, while Castells-Quintana (2016) shows that the low

[2] The changes in the concept of informal employment in Ecuador are available at http://www.ecuadorencifras.gob.ec/institucional/home/

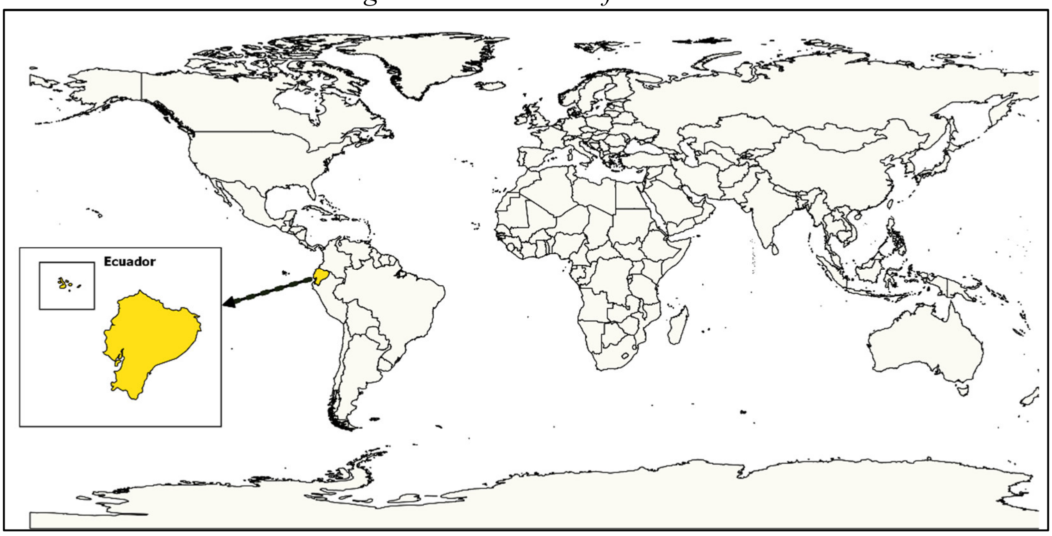
coverage of basic needs is likely to handicap the benefits of agglomeration economies.

1.2 Presentation of the case study: Ecuador

This thesis focuses on city-specific economies of agglomeration in developing economies. Latin America is the most urbanized region of the developing world characterized also by a lack of basic infrastructures (Jaitman, 2015). Within this continent, Ecuador is the chosen case study. It is one of the smallest countries in Latin America with a specific set of characteristics that makes it an interesting case study to expand the spatial knowledge on agglomeration effects.

Ecuador is a small open developing economy (UN, World Economic Situation and Prospects 2014, Statistical Annex). It is in South America. It borders to the north with Colombia, the East and South with Perú, and to the West with the Pacific Ocean. Ecuador has a geographical extension of 283,561 km^2 and it is characterized by four natural regions: Coastal region, Andean Highlands region, Amazon region and the Galapagos Islands. These geographical characteristics of Ecuador make it the fourth smallest country of the continent and one of the ten countries with most biodiversity per squared kilometer of the world. Figure 1-1 shows the location of Ecuador in the world map.

Figure 1-1. Location of Ecuador



In terms of population, Ecuador has a total population of around 17 million of inhabitants in the year 2018. Therefore, it is a country with a geographical extension similar to Italy or Great Britain, but much less densely populated: with a similar extension, Ecuador has around 25% of population of these two countries. Besides, the population is not evenly distributed across regions. Two regions, the Coast and the Andean Highlands, concentrate most of Ecuadorian population, while the Amazon and Galapagos islands are less populated. The population is also characterized by different ethnicities, with the predominance of *mestizos,* which are a mixed race between white and native indigenous. There also ethnic minorities such as white, black, and *mulatos*.

The political division of Ecuador is defined in three levels: provinces (25), *cantones* (224), and *parroquias*/parishes (1,024).[3] The last one is the closest definition to municipality. Ecuadorean authorities consider urban areas as the main cantonal header of each province. The rural population is composed by the remaining population outside of main cantonal header.[4] Thus, there is not a minimum of population size to be an urban area. Figure 1-2 shows the natural regions and the political divisions of Ecuador.

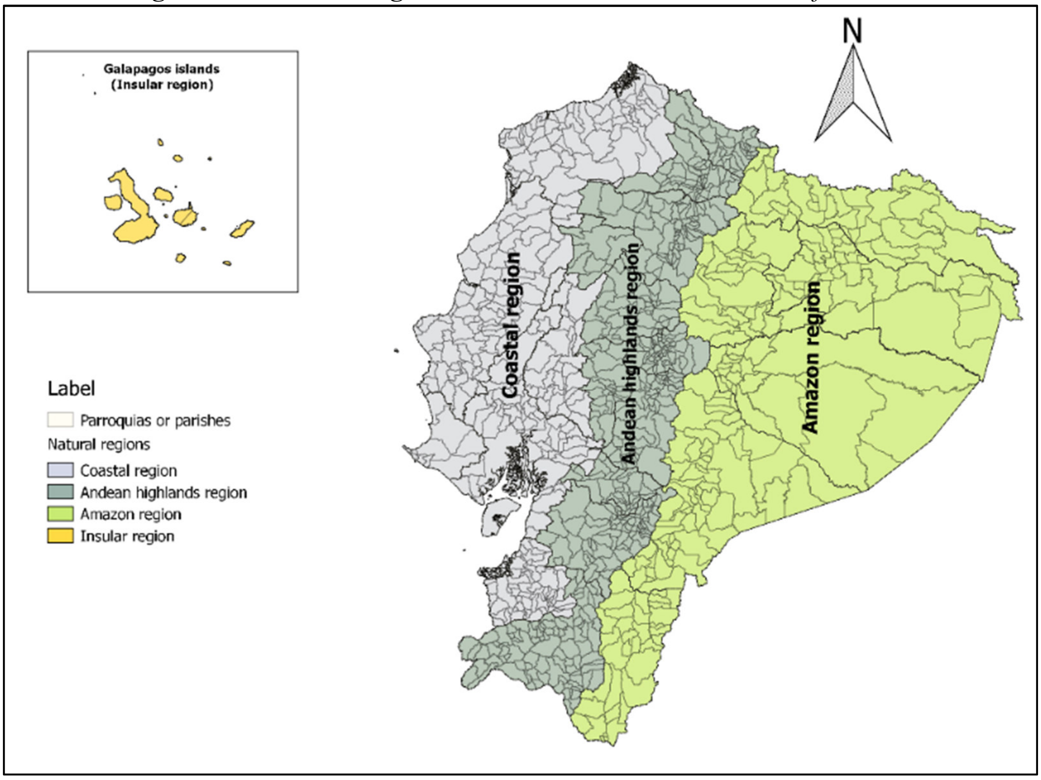
In terms of urbanization, Ecuador has faced a rapid urbanization since the 1960s. Nowadays, the urbanization rate of Ecuador is around 65%, still below the average of Latin America (75%). The rural-urban migration is one of main driver of these changes. Two cities concentrate most of the urban population, Guayaquil located in the Coastal region and Quito located in the Andean Highland region. According to the last Ecuadorean population census (year 2010), these two cities contain around 27% of the total population, and the 35% of the total urban population. These two cities can be considered as metropolitan cities.[5]

In terms of human development, Ecuador presents an HDI (Human Development Index) of 0.74, ranking 87th in 2015, being labelled as a High Human Development country, (ranked between Thailand and China). Life

[3] Numbers of administrative divisions according to the census of population 2010.
[4] See http://www.siise.gob.ec/siiseweb/PageWebs/glosario/ficglo_areare.htm
[5] According the census of 2010, the four most populated parishes are Guayaquil with 2,291,158 inhabitants, Quito with 1,619,146, Cuenca with 331,888 and Santo Domingo with 305,632 inhabitants.

expectancy at birth in Ecuador is around 76 years, and the mean (expected) years of schooling are 8.3 (14).[6]

Figure 1-2. Natural regions and administrative divisions of Ecuador



Ecuador is a small open economy, with an estimated GDP of 97.8 billion US dollars in 2016, ranked between the 60/65 in the GDP world list, and close to the GDP of Puerto Rico or Morocco. Its economic structure is based on agricultural products and oil extraction. The latter is mainly localized in the Amazon region. Oil extraction revenues belongs to the government. Total exports weigh around 30% of the GDP, and the oil exports weigh around 12% of the GDP.[7] The economy is that of a typical developing country, characterized by low industrial activity and an unskilled labor force.

---

[6] See http://hdr.undp.org/en/composite/HDI, 2016 last year available.
[7] Information gathered from the Ecuadorean Central Bank (BCE) available at https://www.bce.fin.ec/index.php/component/k2/item/763

An interesting characteristic of Ecuador is that it is one of the few countries which is completely dollarized.[8] Thus, Ecuador is highly exposed to changes in the international markets, a problem that sums to the usual political conflicts that characterize government changes in Latin American countries. As any other developing country, Ecuador also presents high rates of income inequality, e.g. a Gini index of 0.45 in 2016.[9]

As for the presence of informality within the country, the informal sector of the economy accounts for around 30 to 50% of the local labor force of the cities. Informal workers are low-educated, with no legal protection and high vulnerability. Also, the informality status is quite persistent over time.

As for poverty, Ecuador does not provide any national information on slums, but according the UN-Habitat (2015), around 35% of the Ecuadorean population live in slums in the year 2014.[10] This slums rate is above the average of Latin America countries, which is around 21%. Thus, the analysis of the link between urbanization and slums is an important concern for the development and good work of cities.

---

[8] Dollarization was implemented since the year 2000 after a severe economic crisis, which took place in 1999. This crisis also boosted international migration that account around more than 1 million of Ecuadoreans.

[9] See https://data.worldbank.org/indicator/SI.POV.GINI, 2016 last year available.

[10] See https://unhabitat.org/download-data/, 2014 last year available.

1.3 Thesis objectives

1.3.1 Global thesis objectives

Following all the considerations presented above, the main objective of this thesis is to study agglomeration economies in Ecuador. A country with different characteristics with respect to those already analyzed (Combes and Gobillon, 2015; Glaeser and Henderson, 2017). The thesis focuses on two main aspects, the effect of city size on wages and on quality of life. Furthermore, the role of informal employment is also studied.

1.3.2 Specific research objectives

We now present the specific research objectives of this thesis, which are analyzed at each specific chapter.

- Chapter two introduces a new approach to identify FUAs that allows to overcome the problem of lack of administrative data. This new approach needs to be simple and suitable for the identification of economic cities. Thus, it should also give similar results with respect to those obtained in previous literature. The use of satellite data and the application of a varying travel time approach is proposed in this chapter.

- Chapter three studies the association between agglomeration effects and wages. This chapter focuses on the extent of agglomeration economies in Ecuador, by analyzing the impacts of urban size and sectoral specialization at the local level on the wages of Ecuadorian workers. This chapter also studies the relationship between agglomeration effects and informal employment. The channels behind agglomeration effects between the informal and formal employment are also investigated.

- Chapter four analyzes the relationship between agglomeration effects and slums in Ecuador. This chapter provides evidence on the role of cities to overcome problems associated to un-planned urban growth. A slum severity index is built to address this relationship, using the information on several variables that cover the lack of basic infrastructure in individual households.

1.3.3 Research limits

This thesis is not exempt of limits. In fact, before starting the analysis of agglomeration effects in Ecuador, an initial and important point is to build urban areas following an economic definition. We assume that administrative boundaries might not be the best spatial units. The Functional Urban Areas (FUAs) are conceived as economic cities and preferred over the administrative boundaries as units of analysis (OECD, 2013; 2012; Veneri, 2017; 2016; Ahrend et al., 2017). However, they are not identified in Ecuador yet.

As for most developing countries, the lack of data is a huge barrier for the economic analysis. Similarly, Ecuador also shows some important research barriers. Ecuador does not account for standard data to identify FUAs. The standard data to identify FUAs is population density and commuting data.[11] Thus, we devote the chapter two to the identification of the FUAs, by means of adapting the standard methodology. The lack of administrative data is replaced by satellite data sets. Satellite data sets give information of the population density and road network system of the country. In addition, a new script written in STATA is introduced in this thesis. The script is used to connect urban areas based on the travel time distance between pairs of observations in a hierarchical procedure. However, it can be used for different purposes in which connecting observations in a hierarchical procedure is needed.[12]

---

[11] Any other economic concept of city that could be applied such as metropolitan areas or Local Labor Market Areas also use commuting data for their identification.

[12] All the commands are available to be downloaded at https://sites.google.com/view/moisesobaco/stata-codes.

In addition, most analysis of agglomeration economies in developed economies have been done using longitudinal data of workers.[13] However, for developing economies such data are generally unavailable. Likewise, Ecuador does not have information of workers in a longitudinal data sets. Thus, we rely on a pool model as main specification to explore our individual data. However, we try to provide robust evidence of the relationship analyzed by both using a wide set of individual controls in the analysis, and by building a subsample of longitudinal data that will be used to identify the channels behind the agglomeration impacts on wages. Moreover, instrumental variable technique will be used to treat the likely endogeneity in the analysis of agglomeration effects. We propose the use of historical population and geological indicators.

This thesis uses labor market surveys named as ENEMDU (*Encuesta Nacional de Empleo, Desempleo and Subempleo*) that are quarterly cross-sectional surveys that contain household information of workers and physical structure of the houses. The ENEMDU surveys are used in Chapters three and four. However, the ENEMDU surveys present some limitations. The codifications -e.g. the ISIC (International Standard Industrial Classification) versions- or the informality concept have changed several times in the survey. Thus, a first important step is to standardize the concepts used for the sample of study. In addition, for the longitudinal analysis presented in chapter three, the ENEMDU surveys do not follow the households, but only the houses. Thus, a unique identification code by household is not available. The longitudinal analysis is build using the identification code of the houses available in the surveys and comparing the household's structure over time to build a consistent panel of workers. Nonetheless, a large number of observations are lost when the panel structure of stayers is built.

We also use the population censuses of Ecuador.[14] However, the population censuses of Ecuador do not have all the specific characteristics to define slum as it has been proposed by UN-Habitat (2003). Thus, we approximate the slum definition by covering four of the five characteristics to define slums. Besides, we go beyond the dichotomy of slums by means of exploring the severity of the slums in the case study.

---

[13] The methodology for the identification of agglomeration economies is presented in Combes et al. (2011).

[14] All the information used in this thesis is available at http://www.ecuadorencifras.gob.ec/banco-de-informacion/.

References

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.

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

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.

Carlsen, F., Rattsø, J. & Stokke, H. E. (2016). Education, experience, and urban wage premium. Journal of Regional Science and Urban Economics, 60 (1), 30–49.

Castells-Quintana, D. (2016). Malthus living in a slum: Urban concentration, infrastructure and economic growth. Journal of Urban Economics, 98, 158-173.

Chauvin, J. P., Glaeser, 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.

Chen, M., Zhang, H., Liu, W. & Zhanf, W. (2014). The global pattern of urbanization and economic growth: Evidence from the last three decades. Plos One, Open access (9).

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.

Camagni, R. Capello, R. & Caragliu, A. (2015). Static vs. Dynamic agglomeration economies. Spatial context and structural evolution behind urban growth. Papers in Regional Science, 95 (1), 133–159.

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

Combes, P. P., Démuger, S. & Li, S. (2015). Migration externalities in Chinese cities. European Economic Review, 76, 152 – 167.

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

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. (2011). The identification of agglomeration economies. Journal of Urban Economics, 11 (2), 253 –266.

D’Costa, S. & Overman, H. G. (2014). The urban wage growth premium: Sorting or learning?. Journal of Regional Science and Urban Economics, 48 (1), 168–179.

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

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 Handbook of Regional and Urban Economics, 4 ed. J.V. Henderson and J.-F. Thisse (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. (Ed.), The Informal Sector Revisited. Paris, OECD.

Foster, N. & Stehrer, R. (2009). Sectoral Productivity, Density and Agglomeration in the Wider Europe. Spatial Economic Analysis, 4(4), 427-446.

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

Glaeser, E. (2011). Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. Penguin Books. ISBN-13: 978-0143120544.

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.

Henderson, V. (2002). Urban primacy, external costs and quality of life. Resource and Energy Economics, 24, 95 – 196.

Henderson, V., Lee, T. & Lee Y. J. (2001). Scale externalities in Korea, Journal of Urban Economics, 49, 479-504.

Holmes, T. J. (2010). Structural, experimentalist, and descriptive approaches to empirical work in regional economics. Journal of Regional Science, 50, 5–22.

Jacobs, J. (1969). The Economy of Cities. Random House, New York.

Jaitman, L. (2015). Urban infrastructure in Latin America and the Caribbean: public policy priorities. Latin America Review, 24-13.

Kahn, M. E. (2010). New evidence on trend of the cost of urban agglomeration. Agglomeration Economics; Chapter of the book: 339 – 354. ISBN: 0-226-29789-6.

Koster, H. & Rouwendal, J. (2013). Agglomeration, commuting costs, and the internal structure of cities. Regional Science and Urban Economics, 43, 352 –366.

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.

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

McCann (2001). Urban and Regional Economics. Oxford University Press.

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.

OECD. (2012). Redefining “Urban”. A new way to measure metropolitan areas. Paris: OECD publishing.

OECD. (2013). Definition of Functional Urban Areas: for the OECD metropolitan database. Paris: OECD publishing.

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

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

Rosenthal, S.S. & W.C. Strange. (2008). The Attenuation of Human Capital Spillovers. Journal of Urban Economics, 64, 373–389.

Shanzi, K. (2010). Agglomeration, productivity, and spatial spillovers across Chinese cities. Annals of Regional Science, 45, 157-179.

Suedekum, J. (2006). Agglomeration and Regional Costs of Living. Journal of Regional Science, 46 (3), 529–543.

United Nations. (2014). World Economic Situation and Prospects, WESP. United Nations Publication Sales No. E.14.II.C.2, ISBN: 978-92-1-109168-7, 198.

UN-Habitat. (2003). The challenge of slums. United Nations Human settlements program, Earthscan Publications on behalf of UN-Habitat.

UN-Habitat. (2015). Slum Almanac 2015 - 2016, UNON, Publishing Services Section, Nairobi. ISO 14001:2004-certified.

Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter?, Computers, Environment and Urban Systems, 56, 86-94.

Veneri, P. (2017). Urban spatial structure in OECD cities: Is urban population decentralising or clustering?. Papers in Regional Science, Forthcoming https://doi.org/10.1111/pirs.12300.

Chapter two: Functional Urban Areas

2.1 Introduction

The physical and functional expansion of urban areas beyond their administrative boundaries results in the creation of integrated cities. Identifying the right dimensions for urban zones is important, as they are the zones with the highest concentrations of population and economic activity. Cities are the engines of regions and countries.

Several approaches are used to define cities, usually resulting in alternative results. Ferreira et al. (2010) systematize a set of methodologies used in the literature with the aim of delineating urban agglomerations. They divide the methodologies into two groups. First, they describe methodologies based on morphology, demography, economic, and social structures. From a morphological point of view, cities are high-density agglomerations with contiguous urban building, such as the Urban Morphological Zones created by the European Environment Agency, which are defined areas composed of continuously built-up areas with a maximum spacing of 200m (Guerois et al., 2012; Bretagnolle et al., 2010 and Milego, 2007).

The second methodological approach Ferreira et al. (2010) describe is based on functional delimitation. This methodology considers commuting patterns between locations and is, by far, the most popular means of defining cities or Functional Urban Areas (FUAs). This kind of analysis is the basis of the first systematic approach to defining local labor markets, which was developed in the US in the 1940s to identify zones in which workers can change jobs without changing their residence. Fox and Kumar (1965) propose a method to create local areas based on commuting data, merging spatial areas hierarchically according to workers' daily travels. Coombes et al. (1986), among others, systematize this procedure by developing algorithms that are widely used in many countries and regions (see Casado-Díaz and Coombes, 2011 for a review). Commuting data is the basis of the Eurostat definition of Larger Urban Zones, building on the methodology established by the OECD in collaboration with the European Commission (OECD, 2012; Djikstraa and Poelman, 2012), with a view to facilitating the construction of a commuting zone around a core city.

The functional delimitation approach is linked to the economic definition of a city as a delimited spatial market. From an economic point of view, a city is a dense area that can be considered an independent market in which supply and demand for goods and production factors are traded and an equilibrium price exists. In the classical economics literature, travel distance to the city determines residential location (Alonso, 1964; Muth, 1969). Many alternative definitions can be used. In terms of price, we ask for price uniformity, which can be applied to city specific elements, such as prices or housing rents or workers' wages. In terms of transportation costs, a city ends at a location where no one accepts to be located, as it is too far away from the city center due to excessive transportation costs relative to cheaper rents. Consequently, the cost of rent, the size of the city, and the commuting flows, depend on the density of the core.

In the joint initiative of the European Commission and the OECD, the functional approach is used to define cities as FUAs. This initiative increases international comparability and helps in the collection of statistical data. The methodology identifies 1,251 FUAs of different sizes in 30 OECD countries and Colombia, which gave as a further result the OECD metropolitan dataset, which considers close to 300 cities with populations of 500,000 or more. Economists prefer the use of FUAs as units of measurement in economic analyses (OECD, 2016; Veneri, 2016, 2017; Schmidheiny & Suedekum, 2015).

The OECD's method uses population density to identify urban cores and commuting flows to identify policentricity and urban hinterlands. The latter data is available in most (if not all) developed countries, but this is usually not the case in developing countries. Consequently, additional work is needed to generalize the use of the methodology to the rest of the world. Our work takes this opportunity and applies the OECD definition of integrated cities to Colombia and Ecuador—neighboring countries in Latin America. We consider the main driver of commuting flows: transportation costs.

We use GIS data such as LandScan, Google Maps, and Open Street Maps. LandScan stores information about the density of a country in grid cells of 1 km^2, which allows us to identify urban areas. Google Maps and Open Street Maps provide information on road network systems connecting urban areas. Contrary to other studies defining cities using satellite data from a morphological point of view (as in Van de Voorde et al., 2011), we use travel time information to proxy accessibility, which can be a reasonable substitute for commuting information. In addition, we incorporate spatial heterogeneity into the travel time threshold by assuming varying boundaries depending on the size of the urban core, in line with the expectations of the theoretical economic models.

We build FUAs for Colombia and Ecuador, neighboring countries with and without commuting data respectively. We use Colombia to calibrate our approach for Ecuador and propose alternative methods to test the sensitivity of the proposed methodology by generating commuting flows using methods such as the radiation model, the gravity model, and internal migration. Our results are satisfactory, and the proposed method is flexible enough to build FUAs in environments with little available data. Our work defines a travel time threshold for defining an urban area and is innovative in two ways: we calibrate travel time thresholds using commuting data and we propose using a differentiated threshold for every city depending on its size and extension.

The rest of the chapter is structured as follows. First, it presents the background of the study. Second, the methodology is introduced. Next, the case study and the data sources are presented. Then, the results and several robustness checks are provided. Finally, the chapter concludes by summarizing the main outcomes of our work.

2.2 Functional Urban Areas

Administrative regions are “the expression of a political will: their limits are fixed according to the tasks allocated to the territorial communities, according to the sizes of population necessary to carry out these tasks efficiently and economically, and according to historical, cultural and other factors” (Eurostat, 1999, p.7). Although they are not spatially random units, administrative regions are not the best spatial units for socioeconomic analyses. One way to overcome the problems associated with administrative units is to identify and modify political divisions in order to shape them into an existing socioeconomic relationship (Cörvers et al., 2009; Frey & Speare, 1992; Karlsson & Olsson, 2006). In this line, a FUA can be understood as the harmonized economic definition of “city”: a functional economic unit (OECD, 2012). FUAs are preferable to political definitions in analyzing, designing, and considering urban policies, although this creates tensions and causes planning problems, since several local governments are responsible for planning, which calls for cooperation among agents within an integrated space.

Cities are not only large and dense areas, but they are also integrated environments. Urban agglomerations are the result of urbanization processes, including the transformation of land cover and land use to re-categorize non-developed areas as developed (Pham et al., 2011; Weber, 2000). The final extension of every area is defined in terms of socioeconomic flows among spatial units, the most common being daily interactions in the labor market (Feria et al., 2015; Klapka & Tonev, 2013; Casado-Díaz & Coombes, 2011; Flórez-Revuelta et al., 2008; Smart, 1974). The process of clustering spatial units according to similar characteristics or attributes is generally considered a regionalization procedure (Kim et al., 2015; Kim et al., 2013; Duque et al., 2007). Kim et al. (2016) identify three types of regionalization: districts, coverage, and incomplete coverage. Metropolitan areas are usually associated with incomplete coverage, as they are based on centers of spatial concentration that are not exhaustive in space. We find different approaches to defining integrated areas as spatial clusters. See Davoudi (2008) for a critical review and Adams et al. (1999) and Tong & Plane (2014) for applications.

The OECD follows a three-step process to identify FUAs. First, urban cores are identified according to density measures. All areas above the minimum threshold of population density are then characterized as potential urban cores. Thresholds vary for every country; the OECD applies a threshold of 1,500 inhabitants per km^2, which is lowered to 1,000 inhabitants per km^2 for the US and Canada. The OECD often refers to satellite imagery to assess land cover in this identification step. Today, all of this information is available and easy to gather for most countries in the world (some recent examples of its use are Gisbert & Marti, 2014; OECD, 2012; Weng, 2012; Ferreira et al., 2010; Herold et al., 2003). The quality of such data depends of the quality of satellite images and the further recognition of density.

In this first step, a second condition must be fulfilled: areas need to have a minimum population to be considered an urban core. These minimum thresholds are established by the OECD at 50,000 inhabitants for Europe, US, Chile, and Canada and 100,000 for Japan, Korea, and Mexico where cities are, on average, larger. In addition, as geographic areas usually do not coincide with administrative areas, the method assumes that a municipality is part of an urban core if the majority (at least 50%) of its population lives within the urban cluster.

The second identification step connects the urban areas found in the first step. These urban areas may not be contiguous, but they may belong to the same integrated space. In this way, FUAs account for polycentric urban structures. Two non-contiguous areas are associated if they show some degree of accessibility. The OECD uses labor commuting data and posits that two urban cores are integrated and belong to the same FUA if at least 15% of the population of any of the cores commutes to work in the other core.

The third and final step of the methodology defines the hinterland or worker catchment area—the area of influence of the urban cores—considering accessibility according to labor commuting. The OECD defines this hinterland as all municipalities with at least 15% of employed residents working in a certain urban core.

In the developing world, data scarcity is a significant barrier to identify these spatial relationships. Carrying out any kind of analysis related to urban policies, planning, or socioeconomics is extremely difficult. Hence, the developing world is excluded from most applied socioeconomic analyses. Coombes (2004) proposes alternative approaches to the use of commuting data to integrate urban systems, such as internal migration flows, concentration indexes, or cluster analysis. Internal migration requires a broad range of data, and it presents some problems, the most significant being that migration not only takes place within urban areas, which can be interpreted as a substitute for commuting, but also between them. Royuela & Vargas (2009) use both commuting and migration flows to define Housing Market Areas and conclude that the former is preferred. Concentration indexes require detailed information that is generally unavailable. Finally, cluster analyses do not consider integration links, which makes them a poor proxy.

To overcome the lack of commuting data, the gravity approach is a common option in territorial studies, including migration and trade (Ahlfeldt & Wendland, 2016; Cohen et al., 2008; Wang & Guldmann, 1996). The simplest expression of the gravity model derives flows from limited data, including masses of population and distance between units. Recently, Simini et al., (2012) and Masucci et al., (2013) have used the radiation model to estimate flows such as commuting or migration. Such models appeared first in physics to study the travel process of energetic particles or waves through a vacuum. The model is parameter free, which makes it suitable for predicting flows when there is no data for setting parameters in gravitational models.

Some authors have performed the task of identifying FUAs in developing countries. Commuting data is available in a few recent cases. Duranton (2015) uses the commuting census of 2005 to define local labor markets in Colombia, and Sanchez-Serra (2016) uses the OECD methodology to identify FUAs in Colombia, again with labor market flows. Rodrigues da Silva, et al. (2014) use cluster analysis and the road supply index in the Brazilian region of Bahia to identify functional regions. Gajovic (2013) uses artificial neural networks, isochrones, and cluster analysis in Serbia. Apart from the cases using commuting information, other methods are either highly dependent on data (such as the self-organizing maps that Gajovic proposes), do not report good approximations for urban centers (K-means clustering), or are case specific, using city-specific clusters based on population density and not on accessibility (Rodrigues da Silva et al., 2014).

As Arsanjani et al. (2014) propose, new techniques for FUA identification should be easy to apply, require little data, and be able to predict urban boundaries precisely. In our view, the OECD methodology deserves the attention of researchers such that it can be expanded with few data requirements. Some exercises have been developed for the case of China based on the concept of accessibility, such as the OECD's (2015) use of road network availability and gradient density to identify FUAs. However, this work is based on limited steps to connect urban cores (contiguity) and to define hinterlands, mostly based on density rather than accessibility.

We propose using the concept of accessibility expressed in terms of travel time on the road network system. This allows us to measure and define proximity between urban cores and the extension of the worker catchment areas. This alternative has been considered in other multinational contexts, such as in the ESPON Project "Study on Urban Functions" (ESPON, 2005), where isochrones were fixed at 45 minutes to determine the boundaries. Travel time has also been considered in coverage analysis, where the main purpose is to identify the spatial extent of the functional form. This approach usually involves covering the total demand for private or public services, such as emerging systems, fire stations, police stations, market areas, etc. (Toregas et al., 1971).

Our work connects previous experiences and links them to the standardized procedure based on the OECD definition of FUA. We present a technique that can be easily calibrated and for which data is available for most regions in the world.

2.3 Methodology

We follow the OECD's methodology in three steps, described as follows:

1. Identifying urban cores: Our first step is identical to the OECD's (2012) procedure. We identify high population density areas using satellite data reporting grid cells, which are classified in terms of inhabitants per km^2. An area is categorized as high density if it is beyond a minimum threshold. We identify clusters of contiguous grid cells of high population density according to the majority rule: if at least five out to the eight cells surrounding a cell belong to the same high-density cluster, the lower-density cell will be added. This procedure is repeated until no more cells are merged. The resulting high-density area is required to have a minimum population size to be considered an urban core. Finally, an administrative unit, e.g. a municipality, is included as part of an urban core if at least 50% of its population lives within the urban cluster.

2. Connecting non-contiguous urban cores that belong to the same functional area: two non-contiguous urban cores belong to the same FUA if they are connected, allowing for poly-centricity in FUAs. This step requires the estimation of travel time between urban cores to infer if they are close enough to have socioeconomic interactions. Next, we introduce the assumption that urban cores follow a hierarchical pattern in space, with some areas having a superior role to others. Then, a clustering algorithm sorts urban cores using the hierarchical variable of population size. Starting with the largest urban core, we test iteratively if any urban core is within a time threshold t, defined as the travel time from centroid to centroid of each urban core. The travel time can be fixed for all urban cores or vary as a function of the area of each urban core. For the latter, we propose using a generic expression such as $Tc_i = \alpha_1 * {Au_i}^{\beta_1}$, where Tc_i is the time in minutes from the urban core and Au_i is the geographic area of the urban core. Parameters α_1 and β_1

will vary according to every analyzed case (country), which calls for some calibration. In a perfect circle $ru_i = \sqrt{1/2\pi} * Au_i^{-1/2}$, where ru_i represents the radius of the urban core. Parameters α_1 and β_1 will capture aspects such as average speed, geography, etc. If two urban cores are within such threshold, they are hierarchycally clustered. This procedure is repeated until there are no possible additional merges.

3. Identifying the hinterlands or fringe: The worker catchment area uses a new threshold, defined as travel time from the centroid of each urban core to surrounding political divisions that are not covered by urban cores. Again, such threshold can be fixed (e.g. 60 minutes) or can be proportional to the urban core. We follow Ahlfeldt & Wendland (2016) and derive a city-specific hinterland related to the dimension of each urban core by means of the following formula: $Th_i = \alpha_2 * Au_i^{\beta_2}$, where Th_i is the time in minutes for the hinterland, Au_i is the geographical area of the urban core, and parameters α_2 and β_2, again, require some calibration. If one area is linked to two urban cores, it will be associated with the largest FUA, as it represents the highest position in the urban hierarchy.

Our hierarchical procedure avoids overlapping: two urban cores can be connected and form a unique FUA. On the contrary, two alternative FUAs are not supposed to be connected. If they were, they would form a unique FUA (step 2). Still, two FUAs can be contiguous: they will be far enough not to constitute a unique FUA, but they can be close enough so that their respective hinterlands are contiguous (step 3).

2.4 The case study: FUAs in Ecuador and Colombia

We use two South American countries Ecuador and Colombia as our case study. Ecuador has 17 million inhabitants and a total territorial extension of 283,560 km^2, close to the size of Great Britain or Italy, although each of these countries has about 60 million inhabitants. Colombia is much a bigger country, with 49 million inhabitants and 1,141,748 km^2, twice the size of Spain. The urbanization rate is around 65% for Ecuador and 75% for Colombia; the Latin American average is around 70%.

Commuting data is unavailable in Ecuador and, consequently, is the focus of our work. Analyzing Colombia allows us to work with a developing country with available commuting data from its 2005 census, gathered from the *Departamento Administrativo Nacional de Estadística* (DANE). In addition, the Colombian case allows to calibrate the parameters for Ecuador, because they share common characteristics: both Ecuador and Colombia are countries of regions with large disparities and idiosyncratic geographical, economic, and sociocultural characteristics, and roads are the main network connection systems. Large cities are found both in the mountainous areas (Bogotá, Medellín, and Cali for Colombia and Quito for Ecuador) and in coastal areas (Barranquilla and Cartagena for Colombia and Guayaquil for Ecuador). In addition, both countries have an Amazon region, which are less populated than the other two regions. These similarities make them a good pair for comparison purposes. In addition, Ecuador is our best option for several other reasons: it is representative of many developing countries in the world in terms of population size (close to the average size of a country in the world once we exclude the 10 largest and 10 smallest countries); its urbanization rate and population size characteristics allow us to analyze changes in minimum thresholds; and it has not been previously analyzed, thereby presenting an opportunity to expand present knowledge in the applied literature.

We use land cover information, transport networks, and demographic information at the lowest political division: municipalities for Colombia and parishes for Ecuador. The LandScan datasets, developed by Oak Ridge National Laboratory, provide land cover information based on Satellite Imagery. In this regard, we follow OECD, as they also use the LandScan database. The database uses approximately 1 km^2 resolution (30” x 30”) and represents an ambient population (average over 24 hours). It is practically Raster information vectorized into SHP format. The roadway information comes from Google Maps and Open Street databases (2013). Political divisions at the local level come from INEC (*Instituto Nacional de Estadística y Censo*) for Ecuador and DANE (*Departamento Administrativo Nacional de Estadística*) for Colombia.

Colombia has five natural regions: two on the coast (Pacific and Caribe), one on the Andean central highlands (Andes), and two on the plains (Amazonia and Orinoquia). The Landscan datasets report 334,215 grid cells of population density (see Figure A2-1 in the Appendix), with a poorer coverage in the Amazonian region. Ecuador has four natural regions: the coastal plain (*Costa*), the inter-Andean central highlands (*Sierra*), the eastern jungle (*Oriente*), and the Galapagos Islands (*Insular*). The final Landscan dataset considers 122,544 valid grid cells of 1 km^2 of population density. These are mainly concentrated in the coastal plain and inter-Andean central highlands regions (see Figure A2-2 in the Appendix) in two specific urban poles, one located in the coastal plain region (Guayaquil) and the other in the inter-Andean central highlands region (Quito).

In 2013 there were 1,046 parishes in Ecuador, and in 2005 there were 1,120 municipalities in Colombia. The mean (median) of population density is around 120 (35) inhabitants per km^2 in Ecuador and 128 (10) inhabitants per km^2 in Colombia. In line with other countries, the distribution of population over municipalities follows a very lumpy and concentrated distribution. In addition, they are largely spatially heterogeneous.

To perform further robustness analysis in Ecuador, where there is no commuting data, we consider the Survey of Households' Living Conditions (SHLC) of 2014. Even though this survey is not designed to map the commuting patterns of the whole country, it reports information of this variable for a large sample of individuals. We use this source to report the average commuting time in Ecuador. Finally, we use the Ecuadorean National Census of Population 2010 to perform additional robustness checks based of the analysis of internal migration patterns and the computation of commuting flows based on the gravity and radiation models.

2.5 Results

Colombia is the first country we analyze. We can use both the OECD methodology using commuting data and our approach considering road accessibility. Having data for the OECD approach allows us to calibrate several parameters for the second procedure, which we ultimately use for the Ecuadorean case.

We must first decide on minimum thresholds for population density and urban size. Such decision depends on the type of policy considered. In our case, it must allow us to capture the maximum presence of urban settlements in the country, including the less populated regions that may have representative urban settlements. For Colombia, previous examples are Metropolitan Areas with more than 100,000 inhabitants (Duranton, 2015) and FUAs with minimum populations in clusters of 50,000 inhabitants and minimum densities of 1,500 inhabitants per km^2 (Sanchez-Serra, 2016). Duranton (2015) considers a 10% preferred threshold for commuting flows, while Sanchez-Serra (2016) follows the standard OECD criterion of 15%, although he also plays with lower figures, such as 10%, which is the threshold set in Colombia's national methodology to delimit FUAs (DNP, 2012). As less developed countries are usually less urbanized, we lower the minimum threshold for density at 500 inhabitants per km^2 (which represents 2.5% of total grid cells for Colombia); the minimum threshold of population size of the urban core at 25,000 inhabitants; and the minimum threshold for commuting flows at 10% to obtain results for Colombia that will be used to calibrate our method. Such low thresholds allow us to identify urban settlements in most parts of the country; otherwise, small urban settlements would be invisible.

In line with several authors (Puderer, 2008; Adams et al., 1999), we assume that all techniques and thresholds are arbitrary. Nevertheless, our decisions are not far from other studies. ESPON (2005) uses 650 inhabitants per km^2 at the NUTS-5 level (municipalities) to identify level urban areas in Europe. OECD (2015) applies a minimum threshold of 550 inhabitants per km^2 in China. Some authorities have even considered an urban density of 400 inhabitants per km^2 (Demographia, 2015). In the same vein, the minimum size threshold is flexible: Toribio (2008) argues that the typical population size to define a municipality as the central core inside of a Metropolitan Area is 50,000 inhabitants. However, he uses a minimum of population size of 100,000 inhabitants, because he considered that Spain is a big country in demographic terms. The OECD uses 50,000 for Europe, and Gisbert & Martí (2014) used the minimum threshold of 1,500 inhabitants per km^2 and 50,000 inhabitants for urban centers for Spain. Our decisions are consistent with the objective of maximizing the number of FUAs in developing countries, where small and medium cities are expected to grow in the near future (a process that is taking place in Ecuador, as explained in Royuela & Ordóñez, 2018). Later, we analyze our procedure's sensitivity compared to alternative thresholds.

Table 2-1 shows the results of the OECD methodology using commuting flows on the number of FUAs in Colombia based on 500 inhabitants per km^2 as a threshold for density. We present the number of FUAs identified at three different minimum sizes for urban cores: 25,000, 50,000, and 100,000 inhabitants. The results are also presented for two alternative thresholds for commuting flows: 10% and 15%. Sánchez-Serra (2016) identifies 53 FUAs for a minimum population size of 50,000, 15% commuting links, and 1,500 inhabitants per km^2, while we identify 58 FUAs with a lower density threshold (500 inhabitants per km^2). Increasing the minimum population size of urban cores significantly reduces the number of FUAs and increasing the threshold of commuting for merging urban cores results in more isolated FUAs.

With our preferred thresholds, we obtain 76 FUAs in Colombia, which we use to calibrate the parameters of connectivity of step 2 of our methodology. Urban cores resulting from the first step can be linked by a fixed travel time or vary as a function of the area of each urban core. We compute the average travel time of connected urban cores using the OECD methodology that considers commuting data. Appendix A2.2 displays the considered options for obtaining these distances. This average figure is about 40 minutes: on average, urban cores within 40 minutes of travel time belong to the same FUA. An alternative is to allow that the time threshold varies with city size. By using the information of connected urban cores we estimate this expression and get $Tc_i = 13 * {Au_i}^{1/4}$ (details are reported in Appendix A2.3).

Step 3 computes the hinterland of the FUAs. As result of the administrative division of the country, only 19 FUAs report hinterlands adding municipalities to the original urban cores. Larger urban cores have hinterlands, as they usually have better road connectivity. Again, we use this outcome, to calibrate travel time as an expression of accessibility. As in the previous step, we use a fixed travel time or a threshold that depends of the area of the urban core. In the Colombian case, this formula becomes $Th_i = 4.5 * {Au_i}^{1/3}$ (see further details in Appendix A2.4).

Table 2-1. FUAs in Colombia based on commuting flows and travel time approaches in Colombia (Population in thousands)

Min Pop	**Urban cores**	**Used Link**	**Total FUAs**	**Total Pop.**	**Mean**	**Median**	**Min**	**Max**	**St. Dev.**
Commuting based approach									
25	88	Commuting at least 10%	76	27,493	361	83	25	7,606	995
50	64		57	26,791	470	121	50	7,606	1,131
100	35		34	25,237	742	322	101	7,606	1,407
25	88	Commuting at least 15%	80	27,195	339	82	25	7,539	954
50	64		58	26,374	454	116	50	7,539	1,099
100	35		34	24,741	721	328	100	7,539	1,372
Accessibility based approach									
25	88	Fixed	69	27,214	494	149	25	7,654	1,156
50	64	travel	54	26,211	569	190	50	7,608	1,237
100	35	time	32	24,642	794	354	100	7,597	1,449
25	88	Varying	76	27,253	363	90	25	7,703	1,008
50	64	travel	56	26,390	471	121	50	8,674	1,229
100	35	time	34	24,709	726	298	100	7,636	1,410

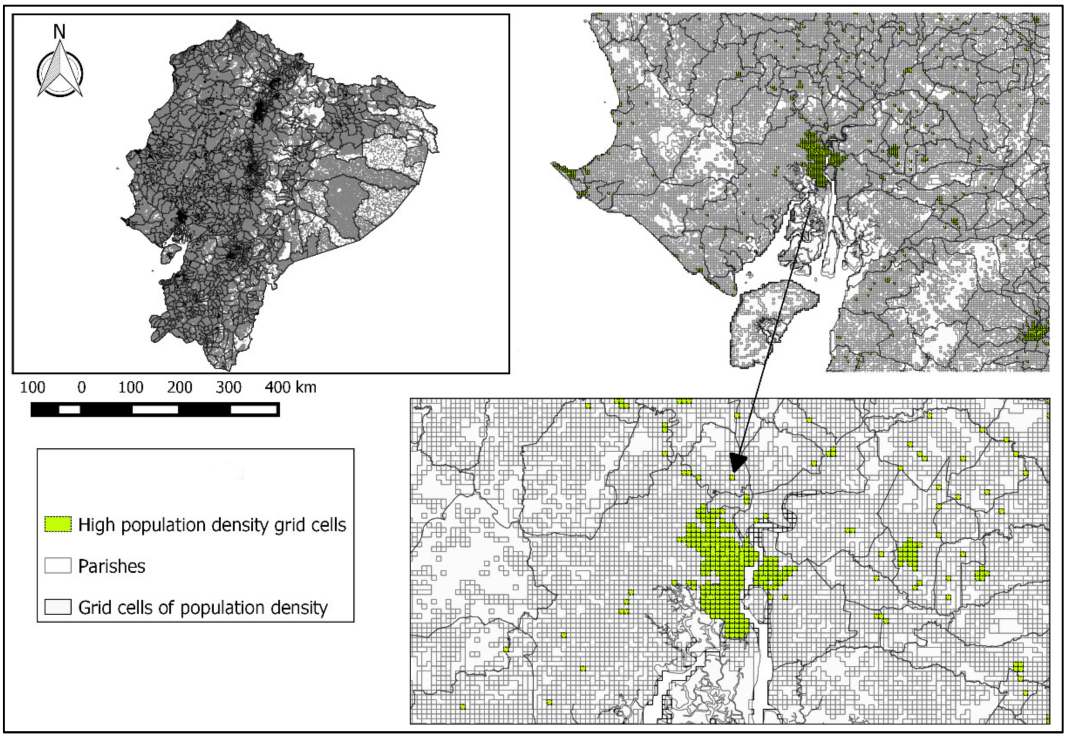
The bottom panel of Table 2-1 displays the results based on road accessibility. We obtain the same number of FUAs than using the commuting-based connectivity approach (76), being the descriptive statistics reasonably close. Such similarities hold while increasing the threshold for population size. We obtain better aggregate summary statistics using a varying travel time approach rather than considering fixed thresholds.

Relying on the calibrated the parameters of our procedure for the Colombian case, we compute the Ecuadorean FUAs. Figure 2-1 displays the map of Ecuador indicating high population density cells (which represent 3% of total) together with a higher detail for the example of the largest city in the country, Guayaquil, which is composed of three administrative boundaries.

Using our preferred thresholds, we identify 34 urban cores in Ecuador, which cover about 50% of total population and 80% of total urban population in the country in the considered year. Given its specific characteristics, we treat the Galapagos Islands as a special case, setting the density threshold at 200 inhabitants per km^2 and a minimum population size of 10,000 inhabitants. Appendix A2.5 displays the descriptive statistics of those urban cores, and the map of the urban cores and the network system.

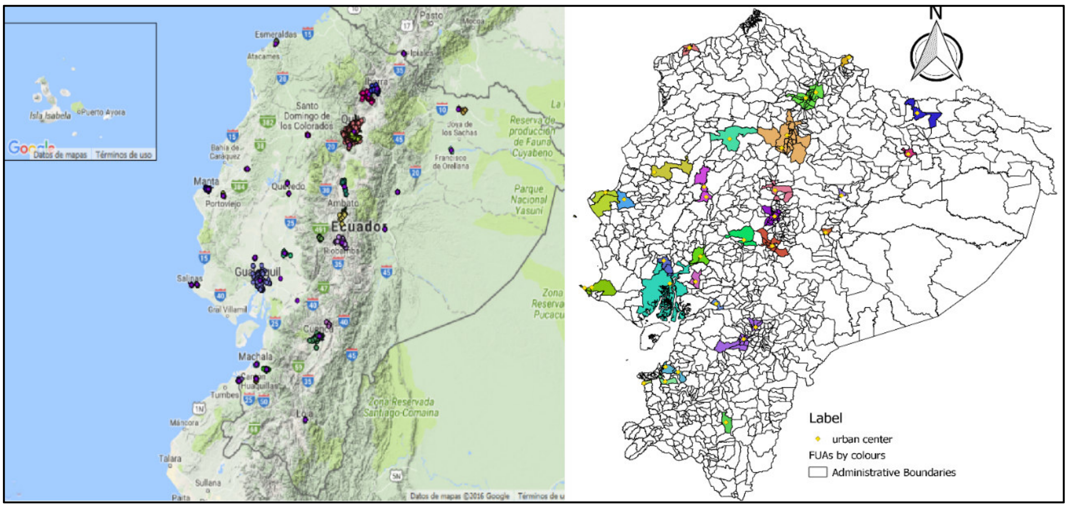
The second step connects non-contiguous urban cores belonging to the same functional area. Every urban core identified above is shaped into a polygon, for which we identify the centroid. We then define the distance matrix by computing the time distance by road from centroid to centroid. In order to verify the travel time threshold for connecting urban cores, we have analyzed the 2014 SHLC. Such survey is only representative at the national and regional level. Similarly, this survey is not designed to capture commuting patterns. It contains information about 110,000 individuals, and around 50,000 are workers. We do not consider commuters within a city, and we disregard workers younger than 15 years old and those who do not return home the same day. Finally, 6,763 workers commute to another city per day, and 3,917 do so by bus, the most popular transportation mode. The median and mean of all commute times are 46 and 68 minutes respectively (the median for bus commuters is 60 minutes, while for car users is about 30 minutes). The global average, then, is close to one hour of travel time, supported with Marchetti's constants that fix the average commute time to approximately one hour (Marchetti, 1994).

Figure 2-1. Grid cells of high population density. Detail for Guayaquil



Like in Colombia, Google Maps does not report actual travel time by public transport in Ecuador, but only by private car, assuming roads are in good condition and traffic is fluid. Developing countries usually have poor quality roads, congested traffic, and bus networks lacking in efficiency. Consequently, we need to translate the 60 minutes by bus inferred from the SHLC into time distance by road reported by Google Maps. We do so by comparing commutes reported at the SHLC with the time indicated by Google Maps. We verify that 30 minutes by private car, reported by Google Maps, is equivalent to one hour by bus. Once we set this threshold, we apply our algorithm based on a hierarchical travel time approach. By applying the clustering algorithm with such thresholds, we merge four high-density urban cores and we ultimately identify 30 FUAs, some of which are polycentric. Allowing for the varying of the travel time, we identify 28 FUAs for Ecuador using the same equation. The SHLC only identifies 326 people commuting between urban cores, and only three urban cores can be connected using this information. In this case, applying the accessibility approach is preferred to use incomplete survey data. Appendix A2.6 shows how we fit Google Maps road distance with survey time distance.

Figure 2-2. Results of FUAs in Ecuador



The final step identifies the hinterland of every FUA, using the equation calibrated above. Any municipality at a lower distance of the threshold is set to be part of the FUA. For instance, under a velocity of 60 km per hour, the threshold for Quito, the urban core with the largest area, greater than 474 km^2, is set at 35 minutes by car, and for the smallest FUA, San Jacinto de Buena Fe, at just 10 km^2, the threshold is set at 10 minutes by car. Figure 2-2 shows the hinterland analysis on the left side and the result in terms of administrative boundaries on the right side (different FUAs by color). Appendix A2.7 reports the detailed list of FUAs.

2.5.1 Sensitivity analysis

This section explores the changes in our results for alternative minimum thresholds in Ecuador. Table 2-2 reports the number of urban cores when the minimum thresholds for density, population size, and travel time are increased. As expected, such increases imply a reduction in the total number of urban cores. No definition should be preferred a priori, although, in our view, in a country where urbanization is taking place the identification of the maximum number of FUAs is preferred.

Table 2-2. Sensitivity test of urban cores based on travel time

Density threshold	Grid cells	Minimum Size Threshold	Initial Number of Urban Cores	Results: number of FUAs			
				Varying travel time	Fixed travel time (minutes)		
					30m	60m	90m
500 inh./km^2	3,699 (3%)	25,000	34	28	30	23	16
		50,000	21	20	20	16	14
		100,000	16	15	15	13	12
1,000 inh./km^2	2,114 (1.75%)	25,000	29	27	28	22	15
		50,000	20	20	20	16	14
		100,000	16	15	15	13	12
1,500 inh./km^2	1,532 (1.25%)	25,000	33	29	31	22	15
		50,000	21	20	20	16	14
		100,000	16	14	15	13	12

Our results display interesting threshold combinations. The highest minimum threshold of population density (1,500 inhabitants per km^2) with a minimum population size of 25,000 inhabitants results in the fragmentation of large urban cores and the creation of new and independent urban cores when compared to a lower threshold for density (1,000). Consequently, we believe that in the Ecuadorean case the chosen lowest minimum threshold of population density is more representative of urban cores across the country.

We also check the influence of fixed versus size varying thresholds for connecting urban cores. Using varying thresholds, we connect two urban cores that are also reported as having significant flows by the SHLC. Consequently, as in the Colombian case, size varying thresholds are preferred over fixed time thresholds.

2.5.2 Robustness checks

In this section, we compare the FUAs obtained for Ecuador using our accessibility approach against urban clusters derived from actual and generated socioeconomic flows, as there is no commuting data available. Next, we describe all considered alternatives to use or generate such flows (additional details for every method are reported in Appendix A2.8):

- ***Survey of Household Living Conditions 2014***: This survey includes information about commuters, although it is not designed as a representative picture at the local level. There is information on 6,763 commuters among around 50,000 workers. It is a matrix of 641 parishes of origin by 540 parishes of destination, but only 2,800 origin-destination pairs have non-zero values. The percentage of commuting flows is obtained from the total outflow of commuters from origin i to destination j divided by total interviewed in i.

- ***Gravitational Approach***: We use a gravitational approach to estimate the full matrix of commuting for the whole country at the local level. The parameters of the gravitational function are obtained by using the commuting information of the SHLC 2014 and the National Census of Population 2010. The specification is a Zero Inflated Negative Binomial model of the between-urban mobility. The considered variables in this model are rescaled commuting flow, total population, and geographical distance.

- ***Radiation model***: We consider the radiation model (Simini et al., 2012), which reports flows between municipalities without any parameterization. This method requires the total outflow of commuters from the origin municipality and population at the origin and destination, which we obtain from the National Census of Population of 2010.

- ***Internal Migration***: We use a matrix of internal migration among parishes between 2005 and 2010, acquired from the 2010 Census. Migration flows within FUAs, which proxy commuting flows, are mixed together with migration between cities. Consequently, we have to differentiate between "movers" and migrants (Zax, 1994). The number of parishes in the migration matrix considers 1,149 origins and 1,211 destinations. We impose a geographical distance restriction between urban cores so that any move beyond this threshold will constitute a migration between FUAs rather than within them. The restriction of distance was 30 minutes by car, which, according to Google Maps, is, on average, 35 km.

Table 2-3 presents descriptive statistics of the flows resulting from the reported alternatives. As expected, they are relatively similar. The rescaled number of commuters from the SHLC 2014 reports several outliers. Similarly, migration flows are heterogeneous compared with what we find in gravity and radiation models.

Table 2-3. Descriptive statistics of commuters

	Obs.	Min	Max	Mean	Median	St. Dev.
Rescaled SHLC	558,902	0	91,403	2.99	0	161.88
Gravity equation	1,024,140	0	4,537	1.54	0	28.71
Radiation model	1,024,140	0	7,563	0.94	0	29.91
Migration flows	1,024,140	1	13,453	12.03	2	98.55

Every described alternative report different flows between municipalities. We use as a starting point the 34 urban cores resulting from the first step of the procedure, which is identified using the minimum density of 500 inhabitants per km^2 and minimum population size of 25,000 inhabitants. We incorporate the computed flows into the OECD procedure to create alternative FUAs, which we compare with those obtained using our accessibility approach. The OECD procedure using commuting flows assumes a minimum threshold of at least 10%, while it is set at least 15% for internal migration (in line with other works comparing these methodologies, Royuela & Vargas, 2009).

Table 2-4 displays the comparison table of FUAs in Ecuador. Column (1) shows the number of identified FUAs. Columns (2) to (6) present descriptive statistics of the population included in those FUAs. Column (7) is the total population of those FUAs and the percentage of population with respect of the country. Differences arise when using computed commuting flows, usually connecting fewer FUAs than our accessibility procedure. On the contrary, when using internal migration flows, more urban cores are connected, as expected, due to the presence of longer distance migrations. Similarly, the migration option captures more population living in FUAs (over 11 million), while the other methods report about 10 million inhabitants. The hinterlands resulting from every method may differ in spatial terms, although the differences in population terms will be small, as every additional spatial unit can be expected to be small.

Table 2-4. Comparative analysis of results among all applied methodologies in terms of population included in each FUA

	FUAs (1)	**Min (2)**	**Max (3)**	**Mean (4)**	**Median (5)**	**St. Dev. (6)**	**Population in FUAs (% of Total) (7)**
Accessibility (varying travel time)	28	37,663	2,812,609	357,320	172,578	663,008	10,004,967 (63.80%)
Commuting SHLC (10%)	31	53,237	2,930,848	340,763	150,258	658,285	10,222,899 (65.15%)
Commuting Gravitational (10%)	33	37,663	2,769,539	295,143	107,129	618,271	9,739,748 (62.07%)
Commuting Radiation (10%)	32	33,186	2,492,869	296,305	161,022	572,811	9,481,786 (60.05%)
Migration flows (15%)	29	59,312	2,558,798	417,070	280,325	634,405	11,260,940 (71.77%)

2.6 Conclusions

This chapter identifies Functional Urban Areas when the researcher has no data on commuting flows. We proxy the OECD methodology by using accessibility and proximity expressed in travel time rather than actual flow data. Our starting point is the use of satellite imagery to identify urban cores. Next, we use travel time in a hierarchical approach to define potential interaction between urban cores and their hinterlands.

We apply our approach to Colombia, and then we extend it to Ecuador, a small country that we believe that can be representative of other developing countries. We test different minimum thresholds to identify cities, and we calibrate our procedure with Colombian records, for which labor commuting data is available. Low thresholds seem to better identify the largest number of cities in a country where urbanization is taking place. We identify 34 urban cores that result in 30 FUAs using a fixed travel time and 28 FUAs using a size-varying travel time, two of them (Quito and Guayaquil) significantly large (2.5 and 2.8 million inhabitants respectively) and the remaining of smaller size. Such areas account for more than 60% of Ecuador’s total population.

We perform robustness checks for Ecuador based on survey and census data. We compute commuting flows resulting from gravitational and radiation models. We also compare our results with algorithms using internal migration flows. All methodologies report similar results, highlighting an important concentration of urban population in those identified urban cores.

Our approach allows researchers, policy makers, and planners to have a better perspective of the integrated cities in the developing world and introduces a feasible methodology, minimizing the need for administrative information. Still, several drawbacks are present. First of all, any approach based on accessibility is actually mixing labor market outcomes with other socioeconomic flows, such as leisure or education commuting. A detailed calibration with labor data of a similar country is advisable to overcome this potential problem. Further research could be applied by adopting the sample of the OECD defined FUAs to obtain global calibrations for some of the parameters of our proposed procedure. In addition, we admit that our approach is based on GIS Google and Open Street Maps assumptions for speed. For example, we do not model explicitly for congestion in larger cities, even though we try to calibrate our approach comparing survey and map travel time to partially overcome this problem. Clearly, both aspects could be tailored with improved data.

References

Adams, J. S., VanDrasek, B. J. & Phillips, E. G. (1999). Metropolitan Area Definition in the United States. Urban Geography, 20(8), 695–726.

Ahlfeldt, G. M. & Wendland, N. (2016). The spatial decay in commuting probabilities: Employment potential vs. commuting gravity. Economics Letters, 143(1), 125–129.

Alonso, W. (1964). Location and Land Use: Towards a General Theory of Land Rent. Cambridge, MA: Harvard University Press.

Arsanjani, J. J., Helbich, M. & Mousivand, A. J. (2014). A Morphological Approach to Predicting Urban Expansion. Transactions in GIS, 18(2), 219–233.

Bretagnolle, A., Giraud, T., Guerois, M. & Mathian, H. (2010). Naming UMZ: methods and results. Technical Report, ESPON 2013 Database. ESPON Database Portal, technical report.

Casado-Díaz, J. M. & Coombes, M. (2011). The delineation of 21st century local labour market areas: a critical review and a research agenda. Boletín de la Asociación de Geógrafos Españoles, 57, 7-32.

Cohen, J. E., Roig, M., Reuman, D. C. & GoGwilt, C. (2008). International migration beyond gravity: a statistical model for use in population projections. Proceedings of the National Academy of Sciences of the United States of America, 105(40), 15269–15274.

Coombes, M. (2004). Multiple Dimensions of Settlement Systems: Coping with Complexity, in: New forms of urbanization: beyond the urban-rural, p. Chapter 16.

Coombes, M.G., Green, A.E. & Openshaw, S. (1986). An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain. Journal of the Operational Research Society, 37, 943-953.

Cörvers, F., Hensen, M. & Bongaerts, D. (2009). Delimitation and Coherence of Functional and Administrative Regions. Regional Studies, 43(1), 19–31.

Davoudi, S. (2008). Conceptions of the City-region: A Critical Review. Urban Design and Planning, 161 (2), 51–60.

Demographia. (2015). Demographia World Urban Areas & Population Projections. Demographia, 14, 132.

Djikstra, L. & Poelman, H. (2012). Cities in Europe, the new OECD-EC definition. Regional focus 01/2012, European Commission.

Departamento Nacional de Planeación (DNP) (2012). Algunos aspectos del análisis del sistema de ciudades colombiano, MISIÓN del Sistema de Ciudades, Bogotá, Colombia.

Duranton, G. (2015). A proposal to delineate metropolitan areas in Colombia. Desarrollo y Sociedad, 75 (1), 223–264.

Duque, J. C., Ramos, R. & Suriñach, J. (2007). Supervised Regionalization Methods: A Survey. International Regional Science Review, 30(3), 165-220.

ESPON. (2005). Potentials for polycentric development in Europe. Project report. ISBN: 91-89332-37-7.

Eurostat. (1999). Regio Database, User's Guide, Methods and Nomenclatures. Official Publication Office, Luxemburg.

Feria, J. M., Casado-Díaz, J. M. & Martínez-Bernabéu, L. (2015). Inside the metropolis: the articulation of Spanish metropolitan areas into local labor markets. Urban Geography, 3638, 1–2.

Ferreira, J. A., Condessa, B., Castro e Almeida, J. & Pinto, P. (2010). Urban settlements delimitation in low-density areas-An application to the municipality of Tomar (Portugal). Landscape and Urban Planning, 97(3), 156–167.

Flórez-Revuelta, F., Casado-Díaz, J. M. & Martínez-Bernabéu, L. (2008). An evolutionary approach to the delineation of functional areas based on travel-to-work flows. International Journal of Automation and Computing, 5(1), 10–21.

Fox, K. A. & Kumar, T. K. (1965). The functional economic area: delineation and implications for economic analysis and policy, Papers and Proceedings of the Regional Science Association, 15, 57–85

Frey, W. H. & Speare, A. (1992). Metropolitan areas as functional communities a proposal for a new definition. Research Report No. 92-245.

Gajovic, V. (2013). Comparative analysis of different methods and obtained results for delineation of functional urban areas. Spatium, 29(1), 8–15.

Gisbert, F. J. G., & Martí, I. C. (2014). El concepto europeo de ciudad: una aplicación para España. Investigaciones Regionales, 30, 145–156.

Guerois, M., Bretagnolle, A., Giraud, T. & Mathian, H. (2012). A new database for the cities of Europe? Urban Morphological Zones (CLC2000) confronted to three national databases of urban agglomerations (Denmark, Sweden and France). Environment and Planning B, 39(3), 439-458.

Herold, M., Goldstein, N. C. & Clarke, K. C. (2003). The spatiotemporal form of urban growth: Measurement, analysis and modelling. Remote Sensing of Environment, 86(3), 286–302.

Karlsson, C. & Olsson, M. (2006). The identification of functional regions: theory, methods, and applications. The Annals of Regional Science, 40(1), 1–18.

Kim, H., Chun, Y. & Kim, K. (2013). Delimitation of Functional Regions Using a p-Regions Problem Approach. International Regional Science Review, 38(3), 235–263.

Kim, K., Chun, Y. & Kim, H. (2016). p-Functional Clusters Location Problem for Detecting Spatial Clusters with Covering Approach. Geographical Analysis, 49 (1), 101–121.

Kim, K., Dean, D. J., Kim, H., & Chun, Y. (2015). Spatial optimization for regionalization problems with spatial interaction: a heuristic approach. International Journal of Geographical Information Science, 30(3), 451–473.

Klapka, P. & Tonev, P. (2013). Functional regions of the Czech Republic: comparison of simpler and more advanced methods of regional taxonomy. Geographica, 44(1), 45–57.

Marchetti. (1994). Anthropological Invariants in Travel Behaviour. Technological Forecasting and Social Change, 47(1), 75–88.

Masucci, A. P., Serras, J., Johansson, A. & Batty, M. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Physical Review, 88(2), 1–9.

Milego, R. (2007). Urban Morphological Zones, Definition and procedural steps. Final Report, Copenhagen: European Environment Agency, ETC Terrestrial Environment.

Muth, R. F. (1969). Cities and Housing. Chicago, IL: University of Chicago Press.

OECD. (2012). Redefining "urban". A New Way to Measure Metropolitan Areas, Paris: OECD Publishing.

OECD. (2015). OECD Urban Policy Reviews: China 2015. OECD Publishing, Paris.

OECD. (2016). Making Cities Work for All. Data and Actions for Inclusive Growth. Paris: OECD Publishing, Paris.

Pham, H. M., Yamaguchi, Y. & Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. Landscape and Urban Planning, 100(3), 223–230.

Puderer, H. (2008). Defining and Measuring Metropolitan Areas: A Comparison between Canada and the United States Geography. Canada.

Rodrigues da Silva, A. N., Manzato, G. G. & Pereira, H. T. S. (2014). Defining functional urban regions in Bahia, Brazil, using roadway coverage and population density variables. Journal of Transport Geography, 36, 79–88.

Royuela, V. & Ordóñez, J (2018). Internal migration in a developing country: A panel data analysis of Ecuador (1982-2010). Papers in Regional Science, 1–23.

Royuela, V. & Vargas, M. A. (2009). Defining Housing Market Areas Using Commuting and Migration Algorithms: Catalonia (Spain) as a Case Study. Urban Studies, 46(11), 2381–2398.

Sanchez-Serra, D. (2016). Functional Urban Areas in Colombia. OECD Regional Development Working Papers, 2016/08, Paris: OECD Publishing.

Schmidheiny, K. & Suedekum, J. (2015). The Pan-European Population Distribution across Consistently Defined Functional Urban Areas. Economics Letters, 133, 10–13.

Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. Nature, 484(7392), 96-100.

Smart, M. (1974). Labour market areas: Uses and definition. Progress in Planning, 2, 239–353.

Tong, D. & Plane, D. A. (2014). A New Spatial Optimization Perspective on the Delineation of Metropolitan and Micropolitan Statistical Areas. Geographical Analysis, 46(3), 230–249.

Toregas, C., Swain, R., ReVelle, C. & Bergman, L. (1971). The Location of Emergency Service Facilities. Operations Research, 19(6), 1363-1373.

Toribio, J. F. (2008). Un ensayo metodológico de definición de las áreas metropolitanas en España a partir de la variable residencia-trabajo. Investigaciones Geográficas, 46, 49–68.

Van de Voorde, T., Jacquet, W. & Canters, F. (2011). Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data. Landscape and Urban Planning, 102(3), 143–155.

Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter?, Computers, Environment and Urban Systems, 56 (2), 86–94.

Veneri (2017). Urban spatial structure in OECD cities: Is urban population decentralising or clustering?. Papers in Regional Science.

Wang, F. & Guldmann, J. M. (1996). Simulating urban population density with a gravity-based model. Socio-Economic Planning Sciences, 30(4), 245–256.

Weber, C. (2000). Urban agglomeration delimitation using remote sensing data, in: Remote sensing and urban analysis: GISDATA 9, 304.

Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sensing of Environment, 117, 34–49.

Zax, J. S. (1994). When is a move a migration? Regional Science and Urban Economics, 24(3), 341–360.

Appendix of Chapter Two

A2.1 Population distribution of Colombia and Ecuador.

Figure A2-1. Colombia: population distribution of high population density. Grid cells with at least 500 inhabitants

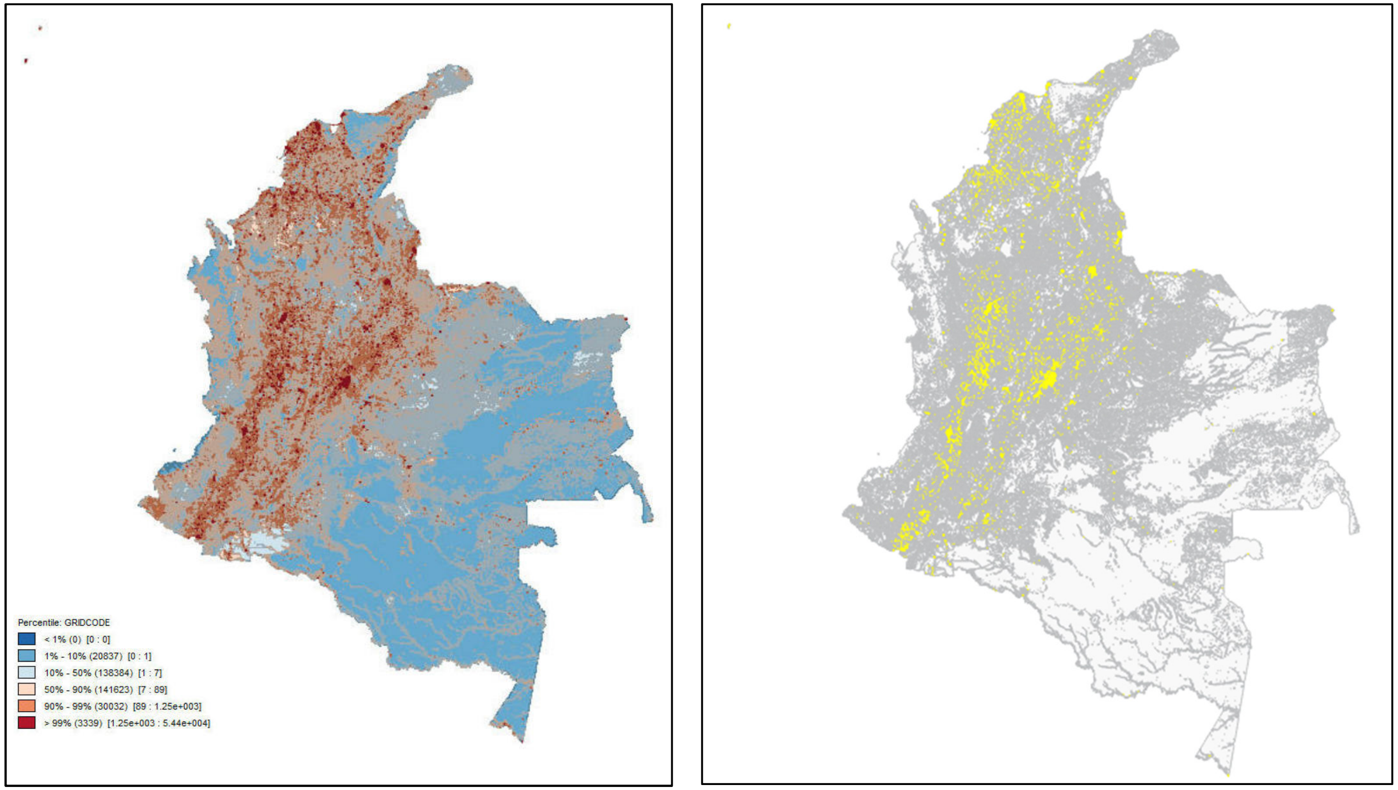
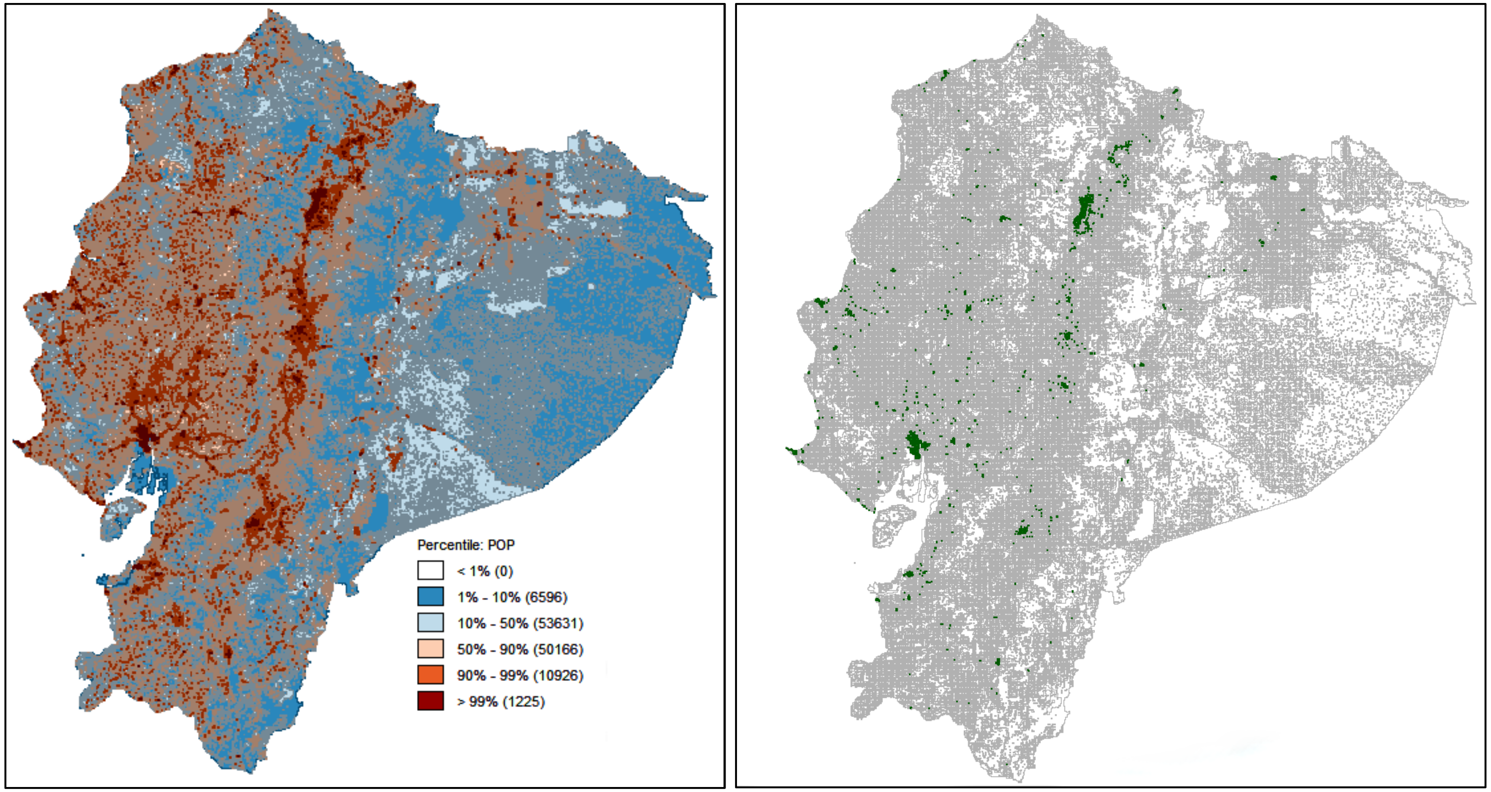


Figure A2-2. Ecuador: population distribution of high population density. Grid cells with at least 500 inhabitants



A2.2 Colombia and Ecuador description

We consider up to three possibilities to compute geographic distances:

1. API Google maps: It is useful when the distance between urban cores there is not computed yet, so it computes at that moment using the Google maps service. We compute these distances by means of the *traveltime3* Stata command. We notice that Google maps service has a limitation in the computation of distance per day, around 5,000 distances. See http://jearl.faculty.arizona.edu/sites/jearl.faculty.arizona.edu/files/traveltime3%20geocode3_b.pdf

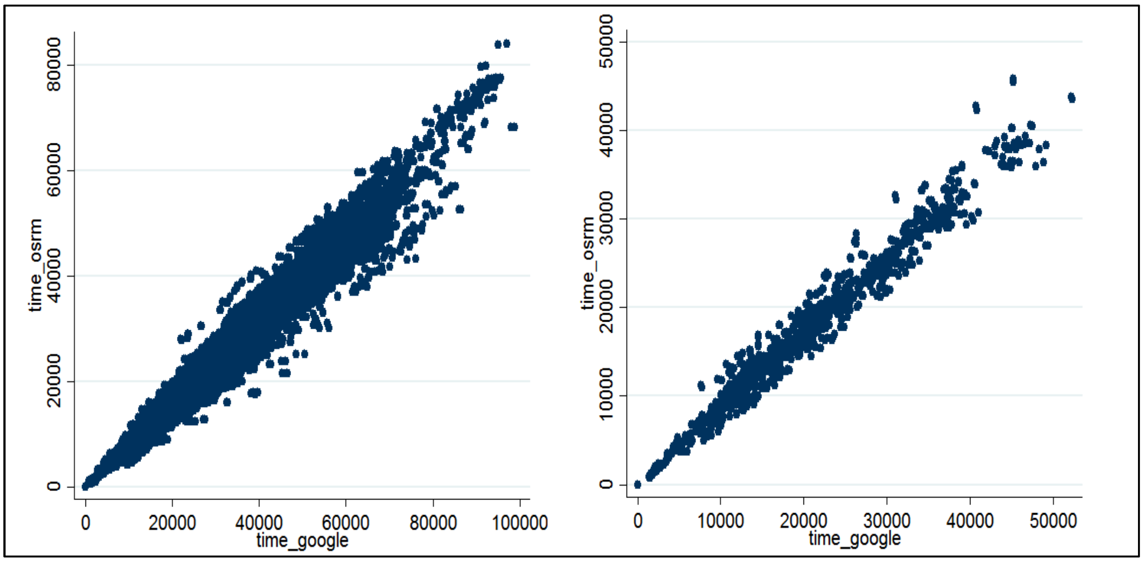
2. Open Street Maps: It works in a similar way, but using the OSRM database. We use *osrmtime* Stata command. While there is not a limitation in the computation of time per day, the database needs to be downloaded, and installed previously (also updated). Consequently, it needs more minimum hardware requirements for working. See https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2691551

3. Origin-destination matrix: we leave open the possibility to upload a self-computed distances matrix, for instance coming from surveys or alternative data sources.

We compare the differences in travel time between the Open Street Maps and Google time. On average Open Street Maps travel time distances are faster. Our preferred option is the use of Google maps. However, its limitation in use per day and the unavailability to download the roads makes OSRM the best complementary data base. Consequently, we suggest using Google time in the second step and OSRM time in the third step. Figure A2-3 shows this comparison for both Ecuador and Colombia.

Figure A2-3. Google maps vs Open Street maps travel time between urban cores

(a) Colombia **(b) Ecuador**



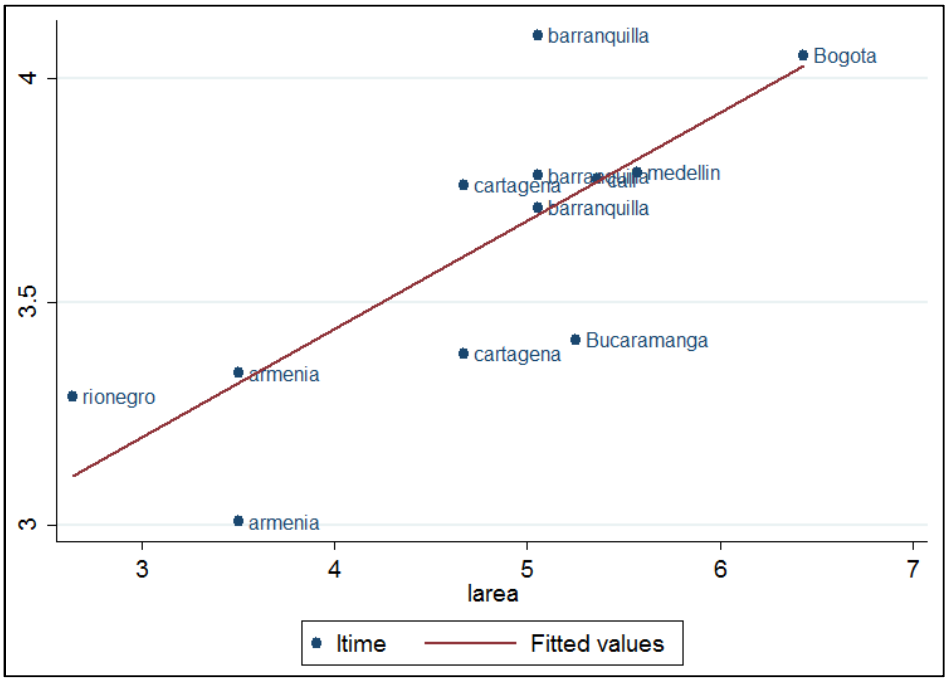
A2.3 Calibration of parameters for connecting urban cores

Table A2-1 reports the 12 Colombian urban cores (origin) that are connected with other urban cores of higher hierarchical level (destination). This information allows us to display the average travel time of connected urban cores that we set at 40 minutes. A fixed travel time can be a good proposal, but it may be not the optimal one. We explore the relationship between commuting patterns and urban size. Figure A2-4 shows the scatterplot between the log of the area of the destination urban core and the log of time between connected urban cores.

Table A2-1. Connected urban cores at 10% commuting flow
(Identified at 500 inhabitants per grid cell / 25,000 inhabitants to be urban core)

Origin ID	Dest ID	Urban Core Origin Name	Population Size Origin	Urban Core Destination Name	Population Size Destination	Origin-Destination Commuting Flow	Origin-Destination Time	Area (size) Destination
5308	5001	Girardota	42566	Medellín	2214494	0.1891	44	263.22
5148	5615	El Carmen	41012	Rionegro	100502	0.1331	27	14.03
8638	8001	Sabanalarga	86631	Barranquilla	1146359	0.1285	60	156.87
8078	8001	Baranoa	51571	Barranquilla	1146359	0.2665	41	156.87
8634	8001	Sabanagrande	25399	Barranquilla	1146359	0.2921	44	156.87
25175	11001	Cha	97896	Bogotá	6840116	0.2301	57	620.78
13052	13001	Arjona	60407	Cartagena	892545	0.1831	43	106.56
13836	13001	Turbaco	63046	Cartagena	892545	0.3362	30	106.56
63401	63001	La Tebaida	33504	Armenia	280930	0.1241	28	33.15
63130	63001	Calarca	73741	Armenia	280930	0.1818	20	33.15
68547	68001	Piedecuesta	117364	Bucaramanga	516512	0.2411	30	190.48
19573	76001	Puerto Tejada	44324	Cali	2119908	0.1137	44	213.54

Figure A2-4. Log(time) vs log(area) between connected Colombian urban cores



We finally regress log of time and the log of the area of the head of the FUA. We have a reasonable adjustment (R^2 about 60%). The constant is 2.473152 and the parameter 0.2417572, both significant at 1%. The final expression is: $Tc_i = 13 * A_i{}^{1/4}$.

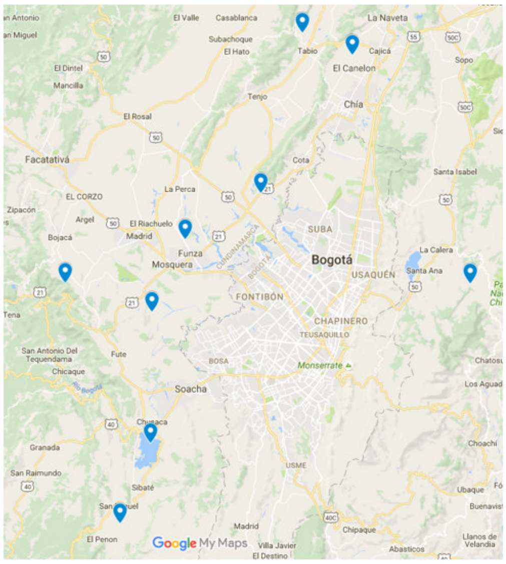
A2.4 Calibration of parameters for computing hinterlands

Figures A2-5 display the hinterlands for five Colombian cities: Barranquilla, Bogotá, Cartagena, Medellín and Cali. Every blue-point reports a municipality that belongs to the hinterland of every FUA. We consider as the hinterland distance as the distance with the farthest municipality of every FUA.

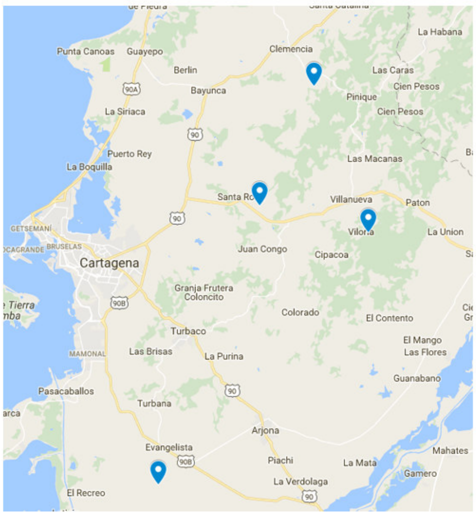
Figure A2-5. Hinterland zones in Colombia

Hinterland of Barranquilla

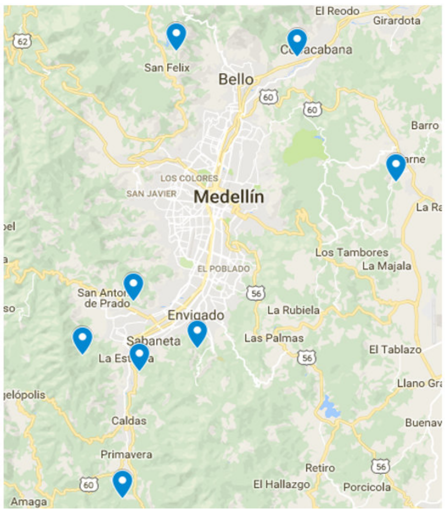
Hinterland of Bogota



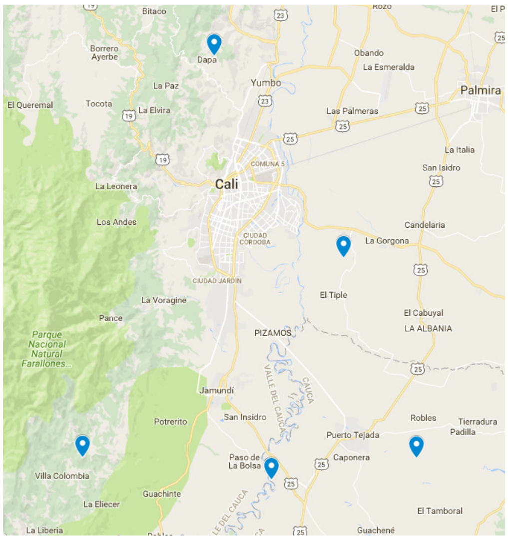
Hinterland of Cartagena



Hinterland of Medellin

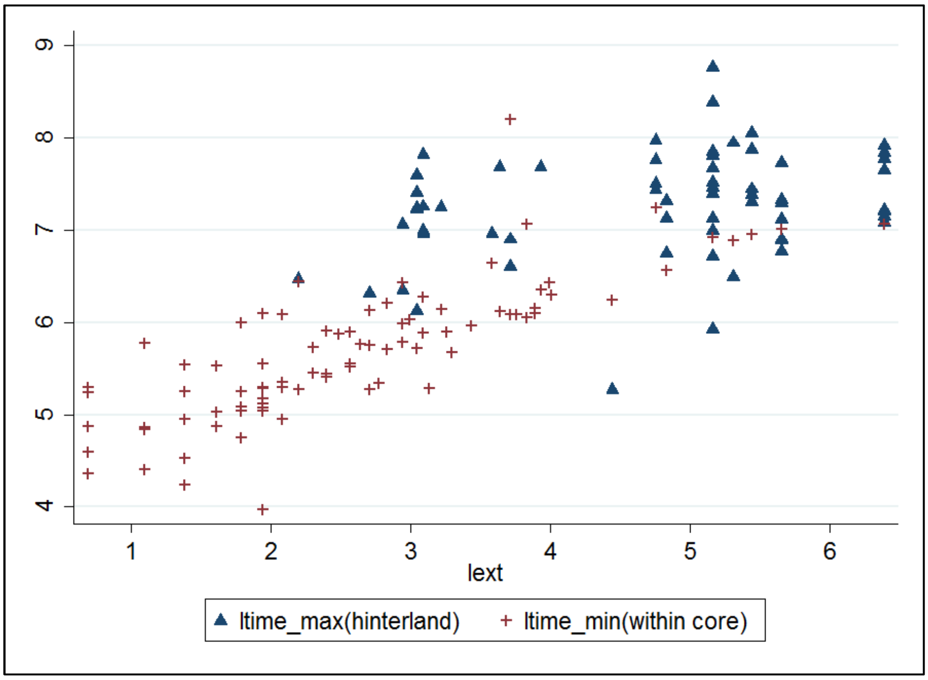


Hinterland of Cali



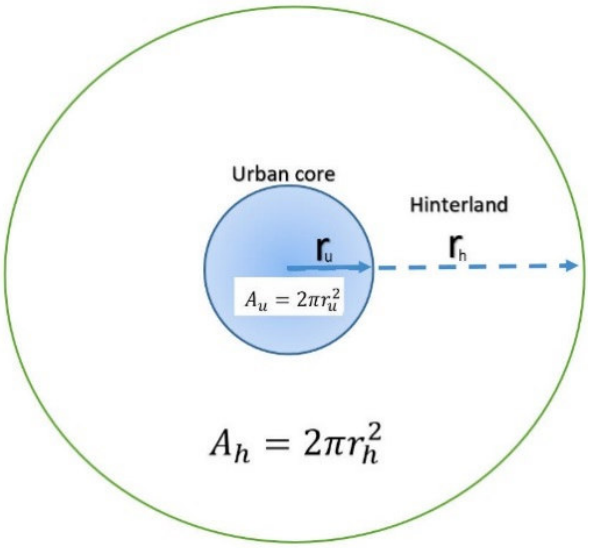
We now consider the relationship between the area of the *urban core*, and the distance of the farthest municipality in the *hinterland*. We assume that every FUA will have a hinterland (low density area) out of the urban core (characterized by high density). The administrative division of space, i.e. municipalities or parishes, makes that these hinterlands are usually within municipalities. We have computed the distance between the centroid of every urban core and the farthest coordinate of the FUA. Figure A2-6 plots the linear relationship between the size of the urban core of *all* FUAs and the maximum distance to every hinterland. Blue triangles represent those FUAs capturing alternative municipalities in the hinterland, while red crosses characterize FUAs where the hinterland is included in a single municipality. Consequently, in practical terms we only have to capture the hinterland of those FUAs adding new municipalities.

Figure A2-6. Scatter plot between the size of the urban core and maximum distance of the hinterland



Then, we look for a relationship where the area of the urban core is used to proxy the size of the hinterland (see figure A2-7.)

Figure A2-7. Hinterland approach



We can say that the hinterland area, A_h, is a function the urban core area, A_u.

$$A_h = \alpha_2 {A_u}^{\beta_2} \tag{A2.4.1a}$$

or

$$\log(A_h) = \ln(\alpha_2) + \beta_2 \ln(A_u) \tag{A2.4.1b}$$

Where, α_2 is an expansion factor and $\beta_2$is an adjustment factor. We may obtain the radius of the hinterland area as a function of the urban core, where the radius measured in distance is equal to time multiplied by a given velocity.

$$r_h = Dist_h = T_h * speed = \sqrt{\frac{A_h}{2\pi}} = \sqrt{\frac{\alpha_2 {A_u}^{\beta_2}}{2\pi}} \tag{A2.4.2}$$

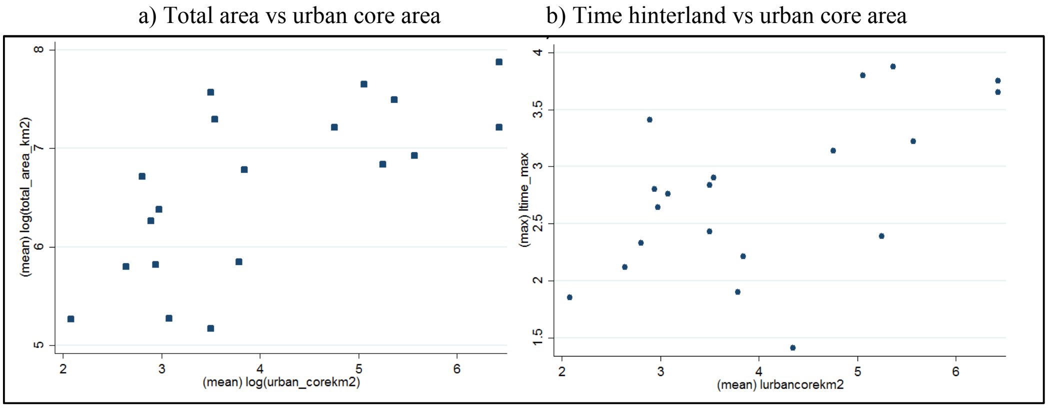
Considering speed is constant, i.e. 60km/h, we get an expression that allows estimating the maximum of travel time as a function of the area of the urban core. The empirical model becomes as:

$$\log(T_h) = \left(\frac{1}{2}\ln\left(\frac{\alpha_2}{2\pi}\right) - \log(speed)\right) + \frac{\beta_2}{2}\ln(A_u) \tag{A2.4.3}$$

$$\log(T_h) = \alpha_2' + \beta_2' \ln(A_u) \tag{A2.4.4}$$

Equation (A2.4.4) is a simple linear equation that allows computing the size of the hinterland as a function of the size (area) of the urban core, what is particularly useful when there is not commuting data of the hinterland, as happens in the Ecuadorean case. To estimate equation (A2.4.4), we need the hinterland generated by urban cores and, for those hinterlands, we need the maximum travel time by urban core.[15] As can be expected, the areas of both urban cores and hinterlands, are not even close to a circle. In addition, administrative boundaries are relatively large compared to real settlements of those municipalities that belong to the hinterland. These characteristics are very close to the Ecuadorean case, where the administrative boundaries are relatively large compared with the municipalities extension as well. Finally, the radius using travel time, generated by using road network measured in extension of Km, tend to be larger than the geographical radius. Figure A2-8.a) shows the relationship between the areas of urban core and urban hinterland, while A2-8.b) shows the relationship between maximum of hinterland travel time and the area of urban core.

Figure A2-8. Relationship between the size of the urban core and the size of the hinterland



Distances were computed using the road network of Open Street Maps with a fixed speed of 60km/h in order to make the computations easier. In the same context, the area was expressed in km^2 and the travel time was recorded in minutes.

Table A2-2 introduces the results of estimate equation (A2.4.4) in column (1), equation (A2.4.1b) in column (2) and the radius of the hinterland against the total size of the hinterland (computed as the total area of all municipalities in the FUA) as robust check in column (3). All parameters are statistically

[15] We use maximum of travel time because the mean or the minimum of the hinterland time do not have a significant slope with the size of urban core.

significant, and their values are the expected values within the confidence of interval. The adjustment of all regressions is quite similar.

Table A2-2. Hinterland estimation

	(1)	(2)	(3)
VARIABLES	**$\ln(time_h)$**	**$\ln(area_h)$**	**$\ln(ltime_h)$**
$\ln(area_u)$	0.334***	0.459***	
	(0.0862)	(0.114)	
$\ln(area_h)$			0.501***
			(0.133)
Constant	1.498***	4.752***	-0.462
	(0.364)	(0.480)	(0.888)
Observations	19	19	19
R-squared	0.469	0.490	0.453

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

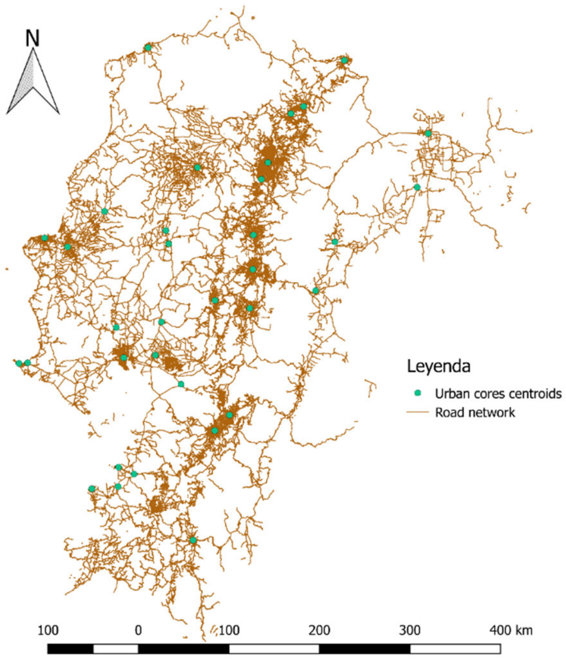
Using the parameters of column (1) find the final expression of the hinterland equation: $T_h = 4.5 * {A_u}^{1/3}$. This time hinterland equation is an equivalent function of the maximum travel time, on average, that an urban core may have according its geographical extension.

A2.5 Description of urban cores and road network structure of Ecuador

Table A2-3. Descriptive statistics of core population (threshold of 500 inhabitants per grid cell)

Reference Name	Pop. Size	Pop. Mean	Pop. Median	Pop. Max	Pop. Min	Pop. St.Dev.	Total cells	Fringe (min)	Area km2	Reference Region
Guayaquil	2553993	8238.69	5008.5	39800	0	9150.31	310	30	297	Coastal
Quito	2166700	4142.83	1753	41536	3	4950.62	523	35	474	Highland
Cuenca	347371	3581.14	1770	39473	92	4809.74	97	21	93	Highland
Manta	294618	3682.73	1910.5	21696	11	4337.59	80	19	70	Coastal
Santo Domingo	286186	8943.31	5531	31110	58	9217.87	32	14	29	Highland
Ambato	276507	2248.02	729	19390	7	3589.86	123	22	113	Highland
Machala	250088	6099.71	4272	43145	91	8935.10	41	15	36	Coastal
Portoviejo	212192	4330.45	1891	35823	112	7233.95	49	16	42	Coastal
Loja	180342	4293.86	1318	36652	392	7853.18	42	15	37	Highland
Esmeraldas	174433	4714.41	1849	19467	28	5388.00	37	15	32	Coastal
Riobamba	169165	4572.03	2008	24266	275	5950.39	37	15	33	Highland
Otavalo	167157	1168.93	893	5528	10	1229.94	143	23	127	Highland
Quevedo	158623	6100.88	2091	37498	563	1474.82	26	13	22	Highland
Libertad	157929	4644.97	2353	34035	0	6560.96	34	14	31	Coastal
Milagro	131806	5272.24	5213	12202	525	3317.68	25	13	22	Coastal
Ibarra	130131	3173.93	1755	19276	0	4062.01	41	15	37	Highland
Latacunga	79710	4195.26	1625	16304	535	4764.16	19	12	16	Highland
Babahoyo	71684	7964.89	2205	32503	819	1376.39	9	10	10	Coastal
Daule	69750	5812.5	1169.5	23606	511	7706.03	12	10	11	Coastal
Tulcán	55855	5585.5	4081.5	25846	599	7258.40	10	10	9	Highland
Nueva Loja	53787	2241.13	1778	5147	14	1536.48	24	13	21	Amazon
Huaquillas	49012	4455.64	3353	15801	1143	4119.98	11	10	9	Coastal
Chone	46159	3077.27	2250	7564	712	2498.53	15	11	13	Coastal
Pto.Orellana	45711	1987.43	1202	11981	3	2568.07	23	6	2	Amazon
Tena	39696	3308	1514.5	13105	223	3954.61	12	10	11	Amazon
Pasaje	39235	5605	3385	15888	892	5164.67	7	9	6	Coastal
Puyo	38318	3831.8	2035.5	11683	591	3962.50	10	10	9	Amazon
La Troncal	36678	4584.75	2986	19000	769	5959.36	8	9	7	Coastal
Santa Elena	35830	3981.11	2891	8839	81	3589.01	9	9	8	Coastal
Santa Rosa	32693	2335.21	1753.5	5987	256	1772.24	14	5	1	Coastal
Azogues	31361	2613.42	677	13855	398	4428.09	12	10	10	Highland
Cutuglahua	27797	1737.31	1241	6319	159	1508.30	16	11	14	Highland
Guaranda	27649	5529.8	5974	10648	1365	3626.97	5	8	5	Highland
S.J. de Buena Fe	25820	2347.27	1574	7580	732	1953.85	11	10	10	Coastal

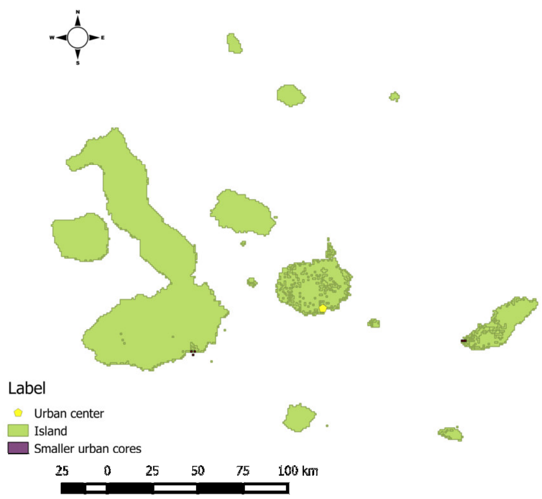
Figure A2-9. Urban cores and the road network system of Ecuador



Galapagos' Islands

For the Insular region (Galapagos Islands), in order to find an urban settlement, we set the minimum density threshold at 200 inhabitants per km^2 and a minimum population size for the urban core at 10,000 inhabitants. As there is no road connection between cities in different islands, we applied a minimum distance between them is around 80 km from the largest urban core. Figure A2-10 shows the urban core in Galapagos.

Figure A2-10. Galapagos' islands



A2.6 Fitting Google maps road distance with survey time distance.

Here we fit Google maps road distance with survey time distance. We compare informed time of commuting at SHLS, from which we know origin and destination, against travel time by car computed using Google maps. The information at SHLS allows for considering the mode of transportation. We exclude marginal transportation modes, such as rides on animals, boats, airplanes, planes and those usual for short distances, such as walking and biking. Table A2-4 displays some descriptive statistics, while Figure A2-11 and Figure A2-12 report respectively the density of travel time associated with Google maps and surveys time.

Table A2-4. Travel time survey vs travel time Google maps

	Google time	**Survey time**
Mean	2161.907	3847.814
Std. Dev.	(1490.121)	(2065.578)

Figure A2-11. Density of Google time and survey time

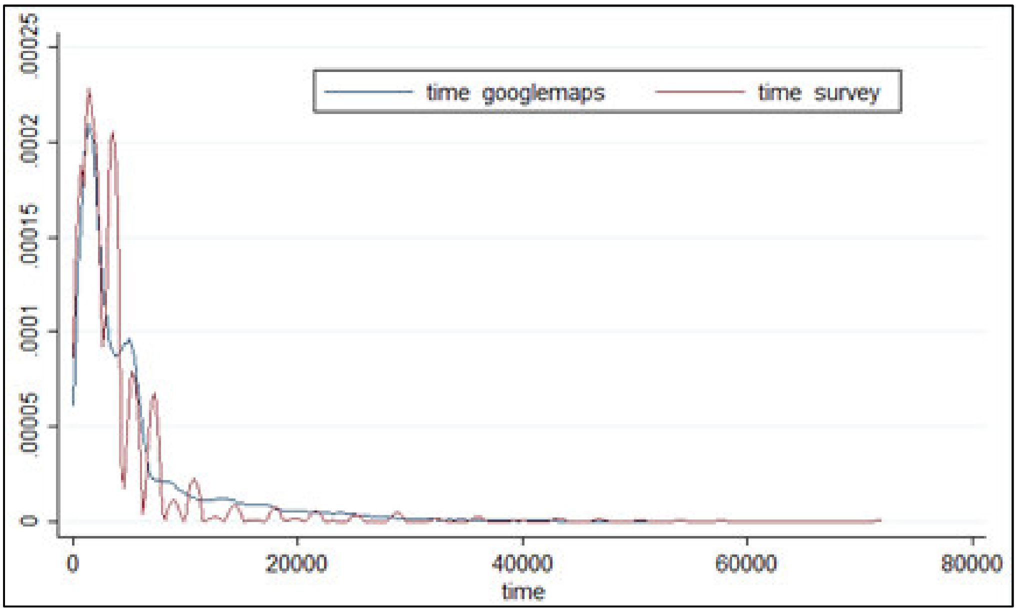
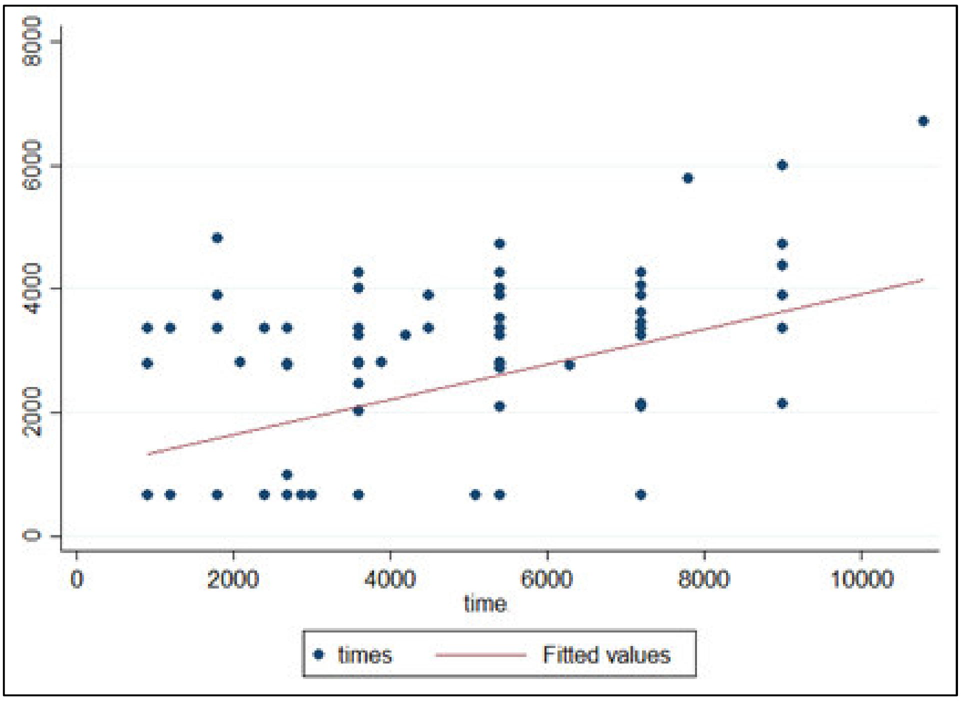


Figure A2-12. Scatter plot of survey time and Google maps in seconds



A2.7 Description of Ecuadorean FUAs

Table A2-5. List of municipalities in the FUAs of Ecuador

Head	Name	Code	Name	Pop	Total Pop
10150	Cuenca	10167	Sinincay	17507	544619
10150	Cuenca	30151	Cojitambo	4070	544619
10150	Cuenca	30150	Azogues	41924	544619
10150	Cuenca	30153	Guapán	9768	544619
10150	Cuenca	10170	Valle	26840	544619
10150	Cuenca	10151	Baños	18602	544619
10150	Cuenca	30250	Biblián	14812	544619
10150	Cuenca	30252	San Francisco de Sageo	1870	544619
10150	Cuenca	10162	Ricaurte	21373	544619
10150	Cuenca	10168	Tarqui	11580	544619
10150	Cuenca	10150	Cuenca	366378	544619
10150	Cuenca	10169	Turi	9895	544619
20150	Guaranda	20157	San Simón (Yacoto)	4569	66680
20150	Guaranda	20158	Santafé (Santa Fe)	1904	66680
20150	Guaranda	20150	Guaranda	60207	66680
30450	La Troncal	30450	La Troncal	48798	48798
40150	Tulcán	40150	Tulcán	65608	65608
50150	Latacunga	50550	San Miguel	33693	158706
50150	Latacunga	50158	Poaló	6218	158706
50150	Latacunga	50150	Latacunga	107129	158706
50150	Latacunga	50153	Guaitacama	10530	158706
50150	Latacunga	50652	Chantilin	1136	158706
60150	Riobamba	60155	Lican	8598	242563
60150	Riobamba	60754	San Andrés	14419	242563
60150	Riobamba	60150	Riobamba	169232	242563
60150	Riobamba	60450	Chambo	12702	242563
60150	Riobamba	60152	Calpi	6985	242563
60150	Riobamba	60750	Guano	17667	242563
60150	Riobamba	60161	San Luis	12960	242563
70150	Machala	70950	Pasaje	58366	324200
70150	Machala	70953	La Peaña	3929	324200
70150	Machala	70150	Machala	261905	324200

70750	Huaquillas	70750	Huaquillas	53237	53237
71250	Santa Rosa	71250	Santa Rosa	57497	57497
80150	Esmeraldas	80166	Tachina	4285	181657
80150	Esmeraldas	80168	Vuelta Larga	3224	181657
80150	Esmeraldas	80150	Esmeraldas	174148	181657
90150	Guayaquil	90750	Eloy Alfaro (Durán)	263970	2790620
90150	Guayaquil	90150	Guayaquil	2466882	2790620
90150	Guayaquil	91650	Samborondón	59768	2790620
90650	Daule	90656	Los Lojas	9894	109872
90650	Daule	90650	Daule	99978	109872
91050	Milagro	91050	Milagro	157608	163499
91050	Milagro	91051	Chobo	5891	163499
100450	Otavalo	100455	San José de Quichinche	9215	370244
100450	Otavalo	100250	Atuntaqui	25603	370244
100450	Otavalo	100650	Urcuquí	5554	370244
100450	Otavalo	100453	González Suárez	6120	370244
100450	Otavalo	100350	Cotacachi	18221	370244
100450	Otavalo	100356	Quiroga	6861	370244
100450	Otavalo	100458	San Rafael	5893	370244
100450	Otavalo	100157	San Antonio	19140	370244
100450	Otavalo	100154	La Esperanza	8042	370244
100450	Otavalo	100457	San Pablo	10764	370244
100450	Otavalo	100254	San Roque	11145	370244
100450	Otavalo	100450	Otavalo	57352	370244
100450	Otavalo	100150	Ibarra	152624	370244
100450	Otavalo	100251	Imbaya	1405	370244
100450	Otavalo	100456	San Juan de Iluman	9332	370244
100450	Otavalo	100451	Dr. Miguel Egas	5308	370244
100450	Otavalo	100452	Eugenio Espejo	7998	370244
100450	Otavalo	100252	S. F. de Natabue	6209	370244
100450	Otavalo	100253	San José de Chaltura	3458	370244
110150	Loja	110150	Loja	200217	200217
120150	Babahoyo	120150	Babahoyo	103837	126355
120150	Babahoyo	120154	Pimocha	22518	126355
120550	Quevedo	120550	Quevedo	173559	230294
120550	Quevedo	121050	S. J. de Buena Fe	56735	230294
130150	Portoviejo	130150	Portoviejo	239695	239695

130350	Chone	130350	Chone	78255	78255
130850	Manta	132150	Jaramijó	21489	338852
130850	Manta	130950	Montecristi	78793	338852
130850	Manta	130850	Manta	238570	338852
150150	Tena	150150	Tena	37663	37663
160150	Puyo	160150	Puyo	41228	41228
170150	Quito	170176	Pintag	19689	2499616
170150	Quito	170163	Guayllabamba	17803	2499616
170150	Quito	170357	Uyumbicho	5152	2499616
170150	Quito	170151	Alangasí	26630	2499616
170150	Quito	170152	Amaguaña	34158	2499616
170150	Quito	170180	San Antonio	35531	2499616
170150	Quito	170551	Cotogchoa	4416	2499616
170150	Quito	170353	Cutuglahua	18730	2499616
170150	Quito	170155	Calderón	167179	2499616
170150	Quito	170177	Pomasqui	31746	2499616
170150	Quito	170356	Tambillo	9304	2499616
170150	Quito	170164	La Merced	9217	2499616
170150	Quito	170186	Zambiza	4411	2499616
170150	Quito	170179	Puembo	14926	2499616
170150	Quito	170170	Nayón	17169	2499616
170150	Quito	170157	Cumbayá	34550	2499616
170150	Quito	170162	Guangopolo	3359	2499616
170150	Quito	170150	Quito	1778016	2499616
170150	Quito	170166	Lloa	1640	2499616
170150	Quito	170156	Conocoto	90124	2499616
170150	Quito	170550	Sangolquí	91024	2499616
170150	Quito	170184	Tumbaco	54844	2499616
170150	Quito	170175	Pifo	18278	2499616
170150	Quito	170165	Llano Chico	11720	2499616
180150	Ambato	180758	Salasaca	6363	333601
180150	Ambato	180156	Izamba	15717	333601
180150	Ambato	180166	Totoras	7444	333601
180150	Ambato	180160	Picaigua	8939	333601
180150	Ambato	180157	Juan Benigno Vela	8047	333601
180150	Ambato	180158	Montalvo	4222	333601
180150	Ambato	180950	Tisaleo	11704	333601

180150	Ambato	180162	Quisapincha	14031	333601
180150	Ambato	180151	Ambatillo	5658	333601
180150	Ambato	180150	Ambato	192693	333601
180150	Ambato	180155	Huachi Grande	11455	333601
180150	Ambato	180751	Benitez	2360	333601
180150	Ambato	180951	Quinchicoto	1411	333601
180150	Ambato	180165	Santa Rosa	22668	333601
180150	Ambato	180152	Atahualpa	11074	333601
180150	Ambato	180163	San Bartolomé	9815	333601
210150	Nueva Loja	210150	Nueva Loja	64041	67098
210150	Nueva Loja	210152	Dureno	3057	67098
220150	Puerto F. de Orellana	220150	Puerto F. de Orellana	49558	49558
230150	Santo Domingo de los Colorados	230150	Santo Domingo de los Colorados	334740	334740
240250	La libertad	240150	Santa Elena	59125	228006
240250	La libertad	240352	José Luis Tamayo	24864	228006
240250	La libertad	240350	Salinas	39205	228006
240250	La libertad	240250	La Libertad	104812	228006

A2.8. Robustness checks

Commuting patterns

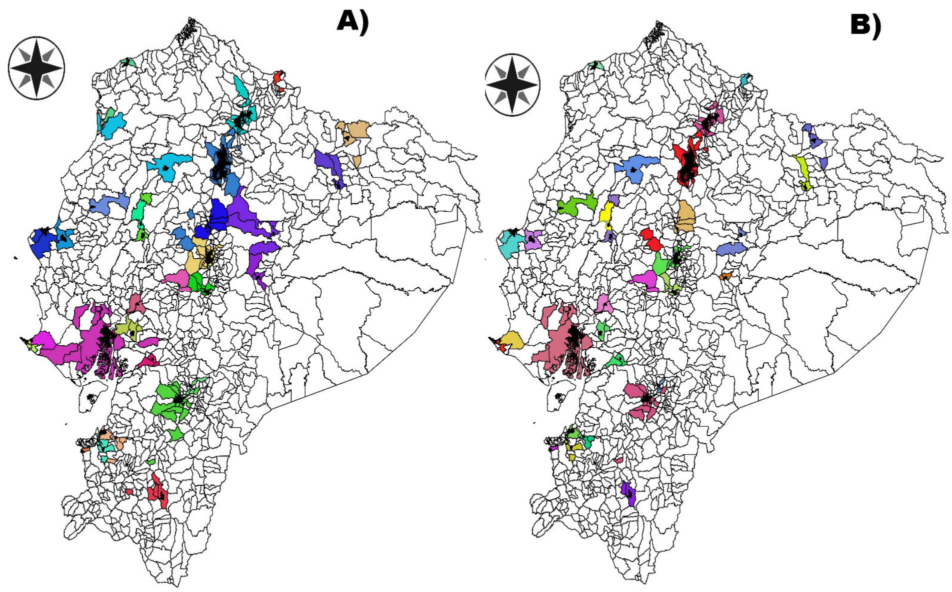
Table A2-6 shows the results of applying the algorithm between urban cores using the SHLC 2014. Urban cores connected in commuting terms are exactly those that were relatively close in travel time terms. Therefore, it gives validation to our proposed based on proximity. A minimum threshold of at least 10% of commuting flow (the same as the preferred threshold for the Colombian case reported by Duranton, 2016) gives the closest approximation to our approach using travel time.

Table A2-6. Sensitivity test of urban cores based on rescaled commuting patterns from SHLC

		Initial	Results / FUAs *(% min. commuting flow)*			
	Size	Urban cores	8%	10%	15%	20%
500 inh./km^2	***25,000***	***34***	30	31	32	32
	50,000	21	20	20	20	20
	100,000	16	16	16	16	16
1000 inh./km^2	25,000	29	26	27	28	28
	50,000	20	19	19	19	19
	100,000	16	16	16	16	16
1500 inh./km^2	25,000	33	27	28	29	29
	50,000	21	19	19	19	19
	100,000	16	16	16	16	16

Figure A2-13 plots the FUAs with hinterlands computed using thresholds of commuting flow at 10% and 15%. In this case the hinterlands were very sensitive to the minimum threshold applied, what can be expected given the poor quality of the commuting data. Similar results of hinterlands are obtained when we use a minimum threshold at 15% and 30 min of travel time using private car.

Figure A2-13. Functional Urban Areas based on commuting patterns derived from the SHLC (A) 10% threshold of commuting (B) 15% threshold of commuting



Gravitational approach

We use the gravity approach under the idea of extending the commuting flows to the whole population matrix of pairs of origin and destination. Using the SHLC 2014, we forecast the total expected number of commuting flows with respect to the total population in each area. In order to do that, we rescaled commuting flows resulting from the survey, multiplying the share of commuters by population size. We use a gravitational exponential decay function devoted to inter-urban mobility; where our dependent variable is the total rescaled commuting flow between origin and destination. This specification is preferred because it has a faster decay function with respect to distance, similar to commuting patterns. An alternative specification can be used to forecast migration patterns. The masses in origin and destination can be total economically active population (pea) or whole population (pop). Distance is measured as straight geographical distance in meters (*Dist*)[16]. The specification is the following:

$$Flow_{orig,dest} = f(Mass_{orig}; Mass_{dest}; D_{orig,dest}) \qquad \text{(A2.8.1)}$$

Flow is the rescaled commuting data from the survey. Mass represent the masses of origin and destination, Dist is the distance. We estimate a linear regression using a zero inflated negative binomial (ZINB) model as OLS overestimates commuters because we have a large number of zeros in the matrix (Westerlund & Wilhelmsson, 2011). In the final estimation we include polynomial extension of origin and destination masses (see results at table A2-7). The flow of commuters was obtained from the ratio between the commuters from origin i to destination j, divided by population of origin i, $\sum F_{ij}/POP_i$.

[16] We preferred using travel time distance because parishes were too large compared with urban settlements, and consequently Google maps or Open Street Maps were reporting incorrect estimates in too many occasions.

Table A2-7. Gravity regression. zero inflated binomial model estimation

Variable	(1) Basic form	(2) Squared form	(3) Cubic form
Count			
Lpop_o	0.413***	1.055***	1.054***
Lpop_d	0.268***	0.091	-1.599***
Distance	-4.211e-06***	-1.01e-05***	-1. 01e-05***
Lpop2_o		-0.030***	-0.029***
Lpop2_d		0.008**	0.170***
Dist2		1.712e-11***	5.810e-11***
Lpop3_d			-0.005***
Dist3			-6.471e-17***
Constant	-0.5387***	-2.789***	3.028
Inflate			
Lpop_o	-0.5842***	0.509***	0.516***
Lpop_d	-.8154***	0.177	6.313***
Distance	2.385e-5***	4.249e-5***	7e-5***
Lpop2_o		-0.055***	-0.056***
Lpop2_d		-0.049***	-0.644***
Dist2		-6.179e-11***	-2.464e-10***
Lpop3_d			0.019***
Dist3			2.746e-16***
Constant	15.248***	4.287***	-16.970***
Lnalpha	-0.426***	-0.505***	-0.527***
Statistics			
N	558,902	558,902	558,902
Lok Lik.	-31246.868	-30396.782	-30049.737
AIC	62511.736	60819.563	60129.474
BIC	62612.84	60965.602	60297.979

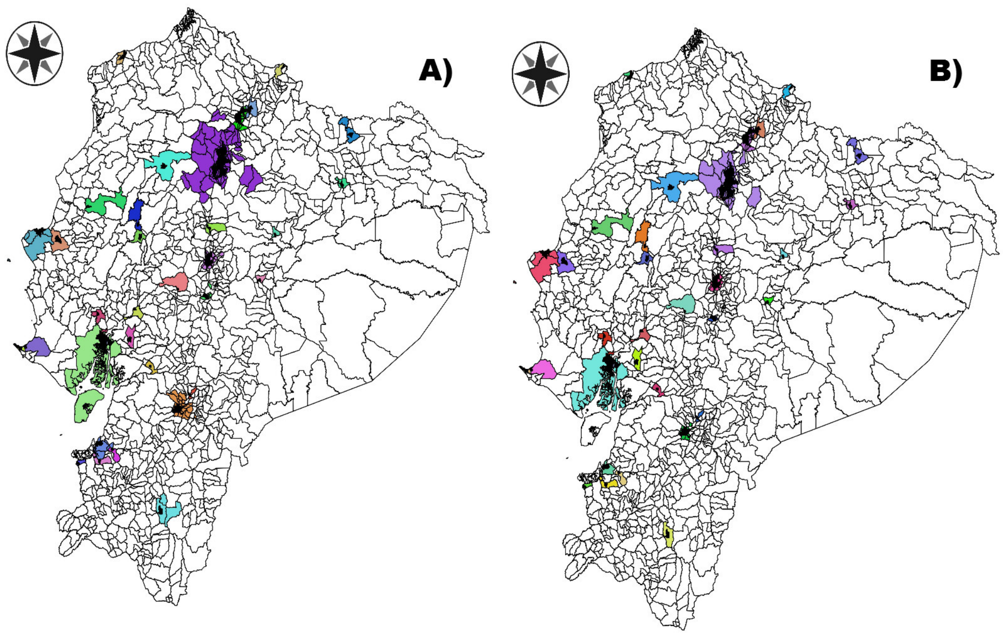
Note: Asteriscs account for significance * p<.05; **p<.01; *** p<.001

Table A2-8 introduces the results of sensitivity test of urban cores. These results are similar to those presented using our travel time proposal and also close to the flows using rescaled commuting resulting from SHLC. Differences arise at lower thresholds, as the gravitational computed flows cannot connect very close urban cores, as other approaches do. Figure A2-14 displays the results considering hinterlands based on thresholds 10% and 15% from commuting flows derived from the gravitational model. Again, hinterlands were very sensitive to those minimum thresholds.

Table A2-8. Sensitivity test of urban cores based on the gravitational approach

		Initial	Results / FUAs (*% min. commuting flow)*			
	Size	Urban cores	5%	8%	10%	15%
500 inh./km^2	*25,000*	*34*	33	33	33	34
	50,000	21	21	21	21	21
	100,000	16	16	16	16	16
1000 inh./km^2	25,000	29	29	29	29	29
	50,000	20	20	20	20	20
	100,000	16	16	16	16	16
1500 inh./km^2	25,000	33	33	33	33	33
	50,000	21	21	21	21	21
	100,000	16	16	16	16	16

Figure A2-14. Functional Urban Areas based on commuting patterns derived of the gravitational model (A) 10% threshold for commuting (B) 15% threshold for commuting



Radiation model

The radiation model for commuting is expressed in equation (A2.8.2).

$$F_{ij} = F_i * \frac{Pop_i * Pop_j}{(Pop_i + w_{i,j})\,(Pop_i + Pop_j + w_{i,j})} \qquad \text{(A2.8.2)}$$

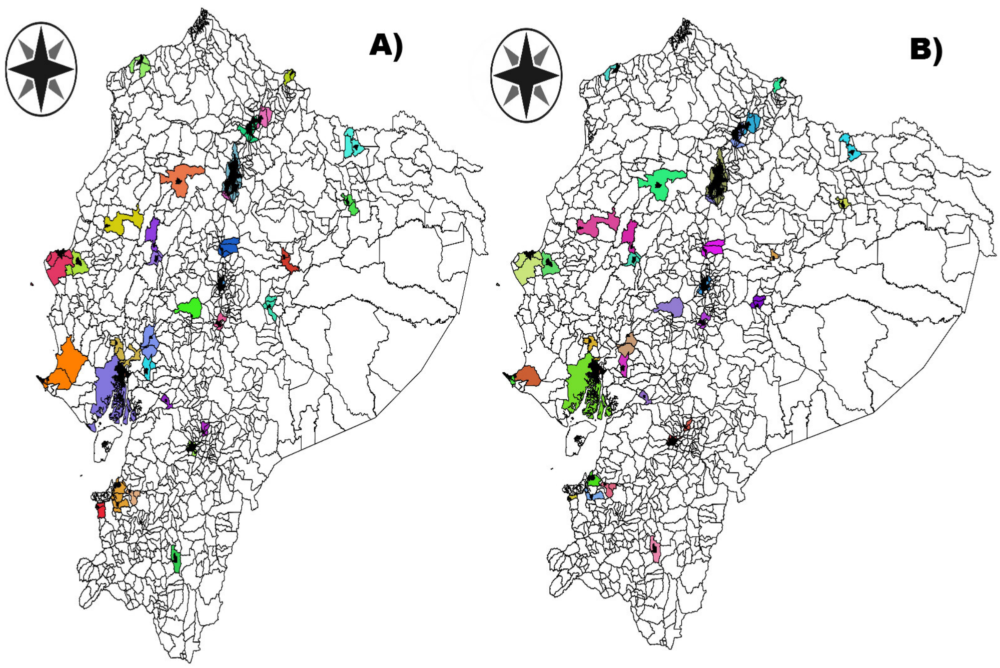
Where F_{ij} is the forecasted commuters from origin i to destination j; F_i is the total outflow of commuters from origin i; Pop_i and Pop_j are the total population in origin i and j destination respectively; and $w_{i,j}$ represents the population contained in a radius given by the distance between origin i and destination j, excluding both the population contained in origin i and destination j. One advantage of this approach is that is parameter free. We use the information at the National Census of Ecuador 2010; this census has a specific question that allows accounting for the proportion of workers commuting out of the parish. Next, we programmed an algorithm in Stata to build the matrix W_{ij}.

We use the forecasted commuters as the source flow for OECD's algorithms. Table A2-9 reports the results and a sensitivity analysis for different thresholds. These outputs are pretty close to the ones derived from the travel time procedure, again at the 10% threshold of commuting. Figure A2-15 displays the FUAs including the hinterlands computed using 10% and 15% of commuting flows derived from radiation model. As before, the hinterland is the most sensitive part of the analysis.

Table A2-9. Sensitivity test of urban cores based on the radiation model

		Initial	Results/FUAs (% min. commuting flow)			
	Size	Urban cores	5%	8%	10%	15%
500 inh./km²	*25,000*	*34*	29	31	32	34
	50,000	21	20	21	21	21
	100,000	16	15	16	16	16
1000 inh./km²	25,000	29	24	26	27	29
	50,000	20	19	20	20	20
	100,000	16	15	16	16	16
1500 inh./km²	25,000	33	27	31	32	33
	50,000	21	20	21	21	21
	100,000	16	15	16	16	16

Figure A2-15. Functional Urban Areas based on commuting patterns of the radiation model (A) 10% threshold for commuting (B) 15% threshold for commuting



Internal migration

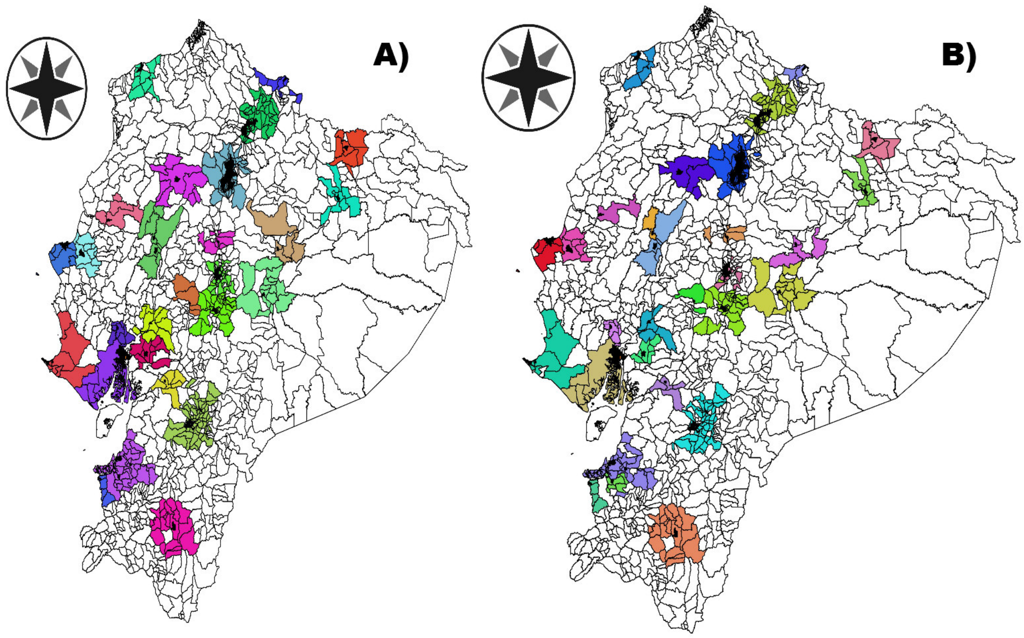
In this case we use internal migration patterns, gathered from the national census of population 2010 of Ecuador. There is information of internal migration between the years 2005 and 2010. The actual matrix is 1,149 parishes by 1,211parishes, as there were several changes in the boundaries of some parishes. We have identified large migration flows between the largest urban poles of the country. Consequently, we have opted for imposing a geographical distance restriction. This allows generating a correct identification of flows that can enter in the algorithm. We opt to use a hierarchical pattern and keep away those urban cores that are relatively far from each other. The restriction of distance was 34,765 meters, which according with Google maps is the distance by car with a half hour of travel time.

Table A2-10 shows the results of the algorithm for different thresholds. The algorithm was successful at connecting cities using a minimum threshold of internal migration, although the patterns are different to the results obtained from travel time and derived commuting flows. In this case, the closest approximation is obtained when using a threshold set at 15% of internal migration. As before, high minimum thresholds make the results more stable. Even if this is a good approach, the results seem very sensible and they were not very similar to commuting patterns. We also present in Figure A2-16, the hinterlands of each FUA at least 15% and at least 20% of internal migration. The results are relatively similar. However, the hinterlands are also too sensitive as the others approaches introduced previously. In this case, our best approximation of the hinterland was using the minimum threshold of at least 20% of internal migration.

Table A2-10. Sensitivity test of urban cores based on internal migration patterns

		Initial	Results / FUAs (*% min. commuting flow)*			
	Size	urban cores	10%	15%	20%	25%
500 inh./km^2	25,000	34	27	29	*33*	33
	50,000	21	20	21	21	21
	100,000	16	15	16	16	16
1000 inh./km^2	25,000	29	26	27	29	29
	50,000	20	19	20	20	20
	100,000	16	15	16	16	16
1500 inh./km^2	25,000	33	27	29	32	32
	50,000	21	21	21	21	21
	100,000	16	15	16	16	16

Figure A2-16. Functional Urban Areas based on migration patterns (A) 10% threshold for migration (B) 15% threshold for migration



Chapter Three: Agglomeration Effects and Informal Employment

3.1 Introduction

With the growth of cities, the analysis of the benefits of agglomeration economies have been object of a wide range of studies. From an 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 Mion and Naticchioni, 2009; Combes et al., 2008; Combes, 2000).

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

A few studies have looked at the importance of agglomeration economies in the developing countries focusing on those characterized by a big geographical extension and large population size such as China, India, Brazil or Colombia (see Chauvin et al., 2017; Combes et al., 2015 and Duranton, 2016).[17] Findings have shown an important role for spatial externalities in fostering productivity and wages, with impacts higher than the one detected for the developed world. However, these results have been uncovered for emerging markets economies, while there is no evidence of studies focusing on more typical developing countries. Moreover, developing countries are generally characterized by a strong presence of the informal economy, which brings to the existence of a dual labor market (Fields, 1990; La Porta and Shleifer, 2014). Albeit the theoretical implications of the presence of informal employment on agglomeration externalities are unclear, assuming

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

the no existence of agglomeration economies for the informal sector is inappropriate (Overman and Venables, 2005). Also, Duranton (2009) points out that formal and informal sectors have strong interconnections, thus suggesting that in both sectors agglomeration effects generate benefits. Nonetheless, there is a lack of more formal studies on this topic (Overman and Venables, 2005).

The aim of this chapter 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, low and increasing urbanization rate, weak industrial activity, and wide presence of informal employment. Our first goal is to understand whether agglomeration externalities exert similar impacts to those detected for the developed world and/or for emerging economies. Secondly, we analyze the importance of the informal sector within this relationship, by 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 work contributes to the literature by directly and comprehensively analyzing how spatial externalities affect informal workers in a typical developing country.[18]

We use repeated quarterly cross-section data coming from the Ecuadorean Labor Surveys (ENEMDU) available from 2005 to 2015, joint with historical information from the censuses in 1950 and 1990. We consider Functional Urban Areas (FUAs hereafter) as unit of analysis, since they represent a suitable economic definition of cities (OECD, 2013). We look at the extent of spatial externalities by analyzing the impact of two measures of

[18] Actually, there are almost no work 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 confronting estimates with all workers with estimates using only formal workers, finding lower impacts for the latter group, and pointing out 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, both articles do not investigate the channels behind the shown differences, which is one of the main purposes of the current analysis.

agglomeration on individual real workers' wages: population size (proxy for urbanization externalities) and sector specialization at the local level.

Two main methodological issues arise in estimating this relationship. First, there might be sorting of workers in the case that more skilled individuals are more likely to be located in highly agglomerated areas. This issue has been taken into account by the empirical literature using individual panel data and performing fixed effects estimations, which may control for unobserved individual heterogeneity (see Matano and Naticchioni, 2012; Mion and Naticchioni, 2009; Combes et al., 2008, among others). However, in this work and similarly to previous studies in developing countries (Chauvin et al., 2017; Duranton, 2016; Combes et al., 2015), we do not have the availability of individual panel data that would allow to control for the unobserved individual heterogeneity. Therefore, we are not able to control for the individual unobserved heterogeneity, throughout our empirical analysis. Nonetheless, on the one hand, we will do our best from a methodological point of view by introducing in the estimation a wide set of individual controls to capture the individual heterogeneity (as in Duranton, 2016, and Glaeser and Resseger, 2010). On the other hand, we will exploit the information coming from a panel subsample of the original data that refers to stayers' workers (not moving across cities) in order to get some insight on the sorting of skilled workers in bigger cities or highly specialized areas.

The second methodological concern is the endogeneity due to the possible simultaneity in the individual choices concerning wages and locations. We will take into account this issue 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), Mion and Naticchioni (2009).

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

When we consider the informal sector in our analysis, results confirm that it strongly matters. First of all, there is evidence of a wage penalization for informal workers, since their earnings are significantly lower (-7%) compared to otherwise similar workers employed in the formal labor market, as expected. When considering the interaction with agglomeration externalities, findings show that informal workers have lower benefits than formal workers. In particular, considering urban agglomeration, informal workers get a 1.6% lower increase in wages than 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 there is a small wage penalty. These outcomes might be partially explained by considering that informal workers are mainly composed by low-educated workers, for whom we detect lower agglomeration benefits compared to high-educated workers.[19]

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 we look at the mechanisms of sorting, matching and learning to explain the outcomes of the analysis (Combes and Gobillon, 2015; Puga, 2010). Considering urbanization externalities, results for formal workers show that the urban wage premium is driven by the sorting of higher (unobservable) skilled workers and by a better quality of job-match in 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 in highly specialized areas, while there is evidence of some wage gains from job changes, that is however lower than in highly populated areas. For informal workers, there is evidence of strong negative sorting in highly specialized areas that helps to explain the wage penalization detected for this category of workers. To conclude, taking into account that formal workers on average have higher skills than informal

[19] This result is in contrast to the one of Duranton (2016), who points out higher benefits for informal workers from spatial externalities. However, and contrary to this analysis, he does not find evidence of higher agglomeration benefits for higher educated workers.

workers, these findings are in line with previous literature pointing out 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 chapter is structured as follows. First, we introduce the literature of reference on spatial agglomeration in developed and developing countries. Then, the data and the variables used for the empirical analysis are described. Next, the methodology and the results are presented. Finally, we derive the conclusions.

3.2 Related Literature

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

Marshall (1890) was the 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 have been 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, that points out that agglomerated areas offer the possibility to achieve a better match between workers and firms; and the sharing mechanism, that refers to the advantages generated by sharing indivisible goods such as facilities and risk in new investments, which decrease individual and firms' costs. These mechanisms are behind both urbanization and specialization positive spillovers, being the former associated to highly dense areas, and therefore across-industries, while the latter take place 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, 2002; Combes, 2000; Ciccone and Hall, 1996, among others), or individual level data (see De la Roca and Puga, 2017; Combes et al., 2010; Mion and Naticchioni, 2009; Combes et al., 2008; Glaeser and Marè, 2001). These studies have taken into account the empirical issues that arise in the identification of the role of such externalities. In particular, they have addressed the endogeneity of the relationship by using an instrumental variable strategy, while the role of the sorting of skilled workers in highly agglomerated areas has been faced by means of a fixed effect strategy when using individual level panel data. Results have shown that workers' sorting captures a large part of the impact imputed to spatial externalities on wages. Further, research has also pointed out the relevance of dynamic gains in the biggest cities (De la Roca and Puga, 2017).

However, these studies have been generally carried out on the developed countries.[20] 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 driven by the growth of cities in developing countries and, interestingly, not in the largest cities (Royuela and Castells-Quintana, 2015). Thus, in line with Glaeser and Henderson (2017), we think that it is 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 considered 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 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 within a range of 11%-14%.[21] Duranton (2016), looking at the case of Colombia, finds an elasticity of wages with respect to population size of 5%, while Ahrend et al. (2015) find a similar magnitude for the Mexican case (4.2%).[22]

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 structure of the data. Nonetheless, they make use of a wide set of individual level controls in order to address as much as possible the

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

[21] They also look at the impact of sectoral specialization at the local level and find elasticities of wages with respect to specialization of around 6%.

[22] 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 at 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.

individual heterogeneity (see Combes and Gobillon, 2015, for a discussion). Furthermore, an important concern in the case of developing countries regards the existence of the informal economy, which implies the presence of a dual labor market that generally accounting for about as a half of the local labor force (La Porta and Shleifer, 2014; Overman and Venables, 2005; Maloney, 2004). It is usually composed 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).[23] Consequently, in our view, a proper analysis of the relevance of spatial externalities in a developing economy should take into account this dimension (Glaeser and Henderson, 2017).

From a theoretical point of view, the literature has not yet reached a consensus about the relationship between agglomeration externalities and the informal economy. 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 be not fast enough to balance out the advantages from agglomeration. Furthermore, they point out that the informal sector itself might contribute to agglomeration economies because of the vitality of this sector, the existence of networks of small firms getting advantage from labor market pooling and the role of the informal economy inside clusters in developing countries. This spillover effect is also stressed by Duranton (2009) who argues that formal and informal sectors are strongly interconnected, so that positive effects might arise within both sectors.

From an empirical point of view, the role of the informal economy within the relation between spatial externalities and wages has been generally neglected. One exception is Duranton (2016) who, by using individual level data for Colombia, provides an indirect test performing estimations of the spatial externalities impact on wages considering either all workers or only workers with a written labor contract. His findings show that spatial

[23] In particular, Fields (1990) points out the existence of a duality also within the informal sector, with an "easy-entry-informal sector" composed mainly by unskilled workers with low wages and protection, and an "upper-tier informal sector" whereas successful workers of the formal sector decide to be on their own, because of the possibility of better wages and working conditions, conditioned by good financial availability and human capital level (as such the sector is not free-entry).

externalities have a lower impact when only workers with written contracts are taken into account (around 3.7% compared to 5/6% when considering all workers), thus inferring higher benefits due to agglomeration economies for informal workers.[24]

3.3 Description of the data and definition of variables of interest

To analyze agglomeration effects on labor productivity in Ecuadorian cities, we use an international harmonized concept of economic cities, namely Functional Urban Areas as in Ahrend et al. (2017).[25] We use 28 FUAs of Ecuador identified in chapter two. Figure A3.1, in the Appendix, presents the FUAs used in this chapter. The most populated FUAs are the capital city, Quito, and Guayaquil with more than 1.5 million of inhabitants each, located in the Andean and Coastal region respectively.[26]

As for wages, we use data coming from the Ecuadorean Labor 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) that contain detailed information on Ecuadorean workers: labor income, worked hours, age, gender, ethnicity, education, occupation, previously migrant status and informal employment status.[27] We have also information on workers' area/city of residence and sector of employment. In order to build the

[24] Other related works are Garcia (2016) who, similarly to Duranton (2016), uses individual level data to study the case of Colombia finding out 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 chapter. Also, Harris (2014), using firm level data for the Nairobi's handicraft industry, finds disadvantages in terms of agglomeration externalities in presence of the informal economy. Further, Bernedo Del Caprio and Patrick (2017) use firm level data for Perú to show lower benefits from agglomeration externalities for the informal economy.

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

[26] Quito and Guayaquil are the largest metropolitan 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).

[27] In particular, the survey is designed to sample specific buildings/houses (*viviendas*) and to provide information on the families living in the sampled buildings/houses. Moreover, the 25% of the sample in each quarter is subsequently re-sampled for four –not consecutive-quarters. Section 3.4.3 provides a more detailed description of the panel dimension of the database.

instruments of the analysis, we join these data with data coming from the population census of 1950 and 1990.[28] Finally, we also merge these data with information on a geological variable associated with the city soil (pH that measure the acidity of the soil) to check the robustness of our IV strategy.

The period of the empirical analysis dates from 2005 to 2015 (44 quarters).[29] We restrict the sample to workers located in the FUAs areas (males and females) aged between 15 and 64 years old who perform one job.[30] We exclude workers employed in the public sector since their wage is set nationally and focus on dependent and self-employed 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, worked hours and real wages per hour.[31] We end up with a sample of 408,197 observations for the period of analysis, around 9,300 per quarter.

Informality is defined according to the last methodology implemented by the INEC (following the 2013 ILO guidelines): informal workers are those employed in firms with less than 100 employees with no tax identification number (*Registro único de contribuyentes*).[32]

---

[28] In 1950, there was the first population census. The data come from the National Institute of Statistics and Census of Ecuador (INEC).

[29] 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).

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

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

[32] 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 include in the formal sector those firms having tax identification number, even in case they have not full accounting records, in order to take into account those cases where this situation is justified by the law. The classification divides workers in formal, informal, dwelling activities and no classified workers. Formal and informal workers account for 90% of our sample, while 4% are dwelling activities and 6% are no classified workers. In this chapter, we only consider formal and informal workers. The new methodology was introduced in 2015 and applied to former surveys until 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. For more details see http://www.ecuadorencifras.gob.ec/institucional/home/ .

We use two measures of agglomeration effects. The first one, total FUAs 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 Matano and Naticchioni (2012), Mion and Naticchioni (2009) and Combes (2000):

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

Where c stands for the city, s for the sector, and t for time. It is the ratio between the share of sectoral employment out of total employment in any city c and the corresponding share at the national level.[33]

Table 3-1 presents the average of the real wages per hour (in US dollars) for the workers' categories considered in in the empirical analysis, at the beginning, middle and final period of the analysis, as well as for the whole-time span of the analysis (2005-2015) joint with their relative presence in the sample.

Considering the composition of the sample (Table 3-1, last column), it is possible to note that males constitute around 60% of the labor force. The dominant ethnicity is mestizo (87%). Around 77% of the sample is constituted by low/medium educated workers (education level at most equal to high school), while about 52% of 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 accounts for just 26% of the total workers' sample (whose 16% in manufacturing), while construction for 10%.

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

As for the dual labor 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 dependent workers. As for firm size, 86% of workers are employed in firms with less than 100 employees. Further, considering the regions, 53% of workers is located in the Andean region, 43% in the Coastal, while the remaining 4% in the Amazonian region. Looking at the area of residence, around 89% of the workers live in an urban area of the FUA.[34] Finally, 31% of workers declare to have not always lived in the city they currently reside.

As for wages (Table 3-1, columns 1 to 4), it is possible to note a general increase in real wages per worked hour throughout the analyzed period. In addition, there is evidence of strong heterogeneity according to the considered dimensions: 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 level of education and with the job skill intensity, as well as with firm size and previously migrant status of the worker. Considering the dual labor market, wages are significantly higher (about 60% more) for formal workers than for the informal ones. Further, wages are lower for self-employed than for dependent workers.[35] As for the economic sectors, apart from the mining and quarrying sector, wages are significantly higher in the service sector with respect to manufacturing. In terms of regions, wages are higher in the Amazonian region (due to oil extraction), followed by the Andean region. Last, and as expected, wages are significantly higher in urban areas than in rural ones.

[34] Across the interconnected economic areas of the FUAs there might be rural areas.
[35] This is due to the fact that self-employed are mostly informal workers.

Table 3-1. Descriptive statistics. Real wage per hour and sample composition by workers' categories

	Mean real wage per hour (US dollars)				**Sample composition**
Workers information	**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

Continue on next page…

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 Observations	29,351	37,329	54,208	408,197	

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

3.4 Empirical Analysis

3.4.1 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 similarly to Duranton (2016). The specification is the following:

$$\begin{aligned} Log(w_{i,t}) = \alpha + \delta \log(Pop_{c(i),t}) + \theta \log(Spec_{c(i),s(i),t}) \\ +X_{i,t}\beta + \mu Dfirmsize_{i,t} + S_{s(i)} + A_{c(i)} + R_{c(i)} \\ + T_t + \varepsilon_{i,t} \end{aligned} \tag{3.1}$$

Where $\log(W_{i,t})$ is the log of real wage per worked hour of 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-digits level), and job category (self-employed/dependent worker) dummies. We also include a firm size dummy (higher or lower than 100 employees) $Dfirmsize_{i,t}$, sector dummies (ISIC 3.1, at two-digits level[36]) $S_{s(i)}$, dummies for the natural regions of Ecuador (Andean, Coastal

[36] In order to be able to instrument the specialization variable, and due to a no always unique correspondence between sectors classification over time (ISIC2 for the census 1990 and ISIC3.1 for the ENEMDU data), we made the choice to group the three ISIC31 sectors 50, 51 and 52 in a single group representing the entire Wholesale and Retail sector. This has prevented us from a significant loss of observations. Nonetheless, we have run a robustness check by repeating the estimation without this grouping (therefore not using around 17,000 observations), and results remain consistent.

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 (3.1) might be affected by two main issues. First, there might be sorting of skilled workers in highly agglomerated areas. Generally, this issue is taken into account by introducing individual fixed effects in the econometric specification. In our data, we cannot properly address this point since we would need a panel dataset in order to control for individual unobserved heterogeneity. Nonetheless, we do our best by introducing a wide set of workers' control variables in line with Duranton (2016) and Glaeser and Resseger (2010). Later, in Section 4.3 we will better address this point by exploiting the panel dimension of a subsample of our data. Second, there might be an endogeneity issue 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 intuition is that lagged levels of total population size and specialization are correlated to the current levels of spatial variables, but they are supposed 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 and Gobillon, 2015; Combes et al., 2010).

Table 3-2. OLS and IV regressions of wage on the spatial variables.
Dependent variable: log of real worker's wage per hour

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) 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)

Continue on next page…

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

Notes: 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-digits level, while sector according to ISIC 3.1, two-digits 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 3-2 shows the results of the pool estimation of equation (3.1). Columns (1) and (2) show the OLS estimates, where in column (1) as controls variables we only insert time, region and sector dummies, while in column (2) we introduce the full set of control variables included in equation (3.1). When no conditioning for workers' o firms' characteristics (first column of Table 3-2), it is possible to see 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. Once we introduce all the set of control variables specified in equation (3.1) the elasticity of wages with respect population strongly drops, passing from 6.5% to 3.9%, while the one of specialization remain below these figures (1.6%). Nonetheless, they are still sizeable effects, which point out that agglomeration externalities are at work also in a typical developing country such as Ecuador, even if the magnitude of the impacts is lower with respect to those uncovered for the emerging developing countries in case of population size (see Chauvin et al., 2017; Duranton, 2016; Combes et al. 2015). The magnitude of the specialization index is consistent with that uncovered from other studies based on a similar definition of sector specialization at the local level (Matano and Naticchioni, 2012; Mion and Naticchioni, 2009).

Column (3) of Table 3-2 presents the IV estimates. Results show a no significant difference with respect to the OLS estimates: the elasticity of wages with respect to population is 3.7%, while the one for specialization is around 1%. This is consistent with previous empirical finding pointing out that endogeneity does not appear to be a main concern in the analysis of the agglomeration impacts on wages (Matano and Naticchioni, 2012; Combes et al., 2010; Melo et al. 2009). Column (4) of Table 3-2 presents the same estimates of column (3) with the addition of the soil pH as an extra instrument in order 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 look for any non-linearity in the relationship between spatial variables and wages. Results show a strong increase in the elasticities of wages with respect to city size, which passes from 3.8% to 9.1%. This outcome points out the existence 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 best cities in exploiting these agglomeration advantages.

3.4.2 Agglomeration effects and informality

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

$$\begin{aligned} Log(w_{i,t}) = {} & \alpha + \delta \log(Pop_{c(i),t}) + \theta \log(Spec_{c(i),s(i),t}) \\ & + \gamma \log(Pop_{c(i),t}) * Inf_{i,t} + \lambda \log(Spec_{c(i),s(i),t}) * Inf_{i,t} \\ & + \rho Inf_{i,t} + X_{i,t}{}'\beta + \mu Dfirmsize_{i,t} + S_{s(i)} + A_{c(i)} + R_{c(i)} \\ & + T_t + \varepsilon_{i,t} \end{aligned} \tag{3.2}$$

Where $\mathrm{Inf}_{i,t}$ is a dummy that takes on a value of 1 whether the worker is employed in the informal sector and 0, otherwise. This variable is interacted with our two measures of agglomeration: $\log(Pop_{c(i),t}) * \mathrm{Inf}_{i,t}$ and $\log(Spec_{c(i),s(i),t}) * \mathrm{Inf}_{i,t}$. All other variables are the same as defined in equation (3.1). 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. Therefore, the significance of the interaction terms will give us an insight of the spatial agglomeration impact on wages in presence of a dual economy.

Table 3-3 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 (3.2) are included in the estimation. First of all, looking at the coefficient related to the informal dummy, it is possible to observe that on average, and in line with the descriptive statistics, informal workers are harshly penalized in terms of wage. In fact, the percentage difference in the expected wage between formal and informal workers is around -6/-7% in the both OLS and IV estimates (columns (1) to (4)).

When taking into account the heterogeneity in the agglomeration impacts between formal and informal workers, OLS results show a general penalization for informal workers in the 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 3-3 shows the IV estimates. Compared to the OLS estimates, we can 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, and still decreases by 1.6% for informal workers. As for specialization, the difference with respect to the OLS estimates is important, since when taking into account the endogeneity of the relationship between wages and spatial variables, there are no longer benefits arising from working in locally specialized areas for informal workers. On the contrary, there is a small wage penalization (around -1.8%).[37] In the next section we will provide a possible explanation for this evidence. For formal workers the wage elasticity with respect to specialization is 2.4%.

Column (4) of Table 3-3 shows the IV estimates using the soil pH as an extra instrument. Instruments of agglomeration variables are exogenous as confirmed by the Hansen test p-value (0.295), while 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

[37] By 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.

elasticity of wages with respect to total population size is 7.7% for formal workers and 5.8% for the informal ones. For what concerns specialization, we find now a small wage premium for formal workers, although still 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, what informs about the strong effect of informality in the two largest cities.

Summing up, these results show that informal workers do not enjoy the same wage premium from agglomeration externalities than formal workers and that they are strongly penalized in the largest cities. These findings are at odd with the one of Duranton (2016) for Colombia, who, by performing an indirect test, point out 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 from agglomeration externalities. Although the same mechanism may be at play also in Ecuador, in this case (and contrary to the case of Colombia), there is evidence of higher returns to agglomeration externalities for more educated workers (Table A3-1, in the Appendix), a finding in line with Bacolod et al. (2009), Glaeser and Resseger (2010) and Wheeler (2001). Since more educated workers are more likely to be employed into formal occupations,[38] this can (at least partially) explain the higher returns to spatial externalities for formal workers we detected.[39] Next, we will provide a more detailed picture of the mechanisms through which agglomeration externalities impacts formal and informal workers' wages.

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

[39] In fact, the higher returns to agglomeration externalities for higher educated workers likely offset more than proportionally -in case at play- the mechanism pointed out by Duranton (2016) for informal workers.

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

	(1) **OLS**	**(2)** **OLS**	**(3)** **IV**	**(4)** **IV**	**(5)** **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)

Continue on next page…

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

Notes: 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-digits level, while sector according to ISIC 3.1, two-digits 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).

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

In this section, we aim at shedding light on the channels behind the impacts uncovered so far for formal and informal workers: we analyze the role of sorting, matching and learning mechanisms on the spatial wage dynamics (Combes and Gobillon, 2015; Puga, 2010). To this purpose, we exploit the information of a panel subsample of our original dataset. The ENEMDU survey is designed to sample families living in specific buildings/houses (*viviendas*), whose each quarter a 25% is resampled, following a 2-2-2 panel structure: a building/house may be sampled for two following quarters, then left for the two successive quarters, and be back to be sampled for other two quarters. This means that if a household has not changed place to live, it might be interviewed at most four times, covering a total time span of six quarters. We take the opportunity of the survey structure to identify households' workers sampled more than once and analyze their wage dynamics.[40] As in previous sections, we focus on workers aged between 15 - 64, with information on formality and informality consistent over time. We delete outliers[41] and extreme observations in terms of real wages, hours worked and real wages per worked hour. By this means, we end up with a panel of 79,902 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 are being analyzed for a developing country, by means of a panel dataset. Nonetheless, we claim 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 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.

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

[41] In particular, we identify outlier workers as those reporting not consistent information over time about age, education, gender and previous migration status.

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 regressed on the set of observable characteristics used as control variables in the main analysis. Table 3-4 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)).

Looking at the skills distribution across low and high populated areas (Table 3-4, top panel), there is evidence of positive sorting. In fact, the average of the fixed effect is 0.047 in high-populated areas, greater than the one in low-populated areas, -0.095. It is remarkable that this sorting effect actually essentially holds for formal workers who witness an increase in fixed effects of 0.129 passing 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 however marginal compared to the one 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 due to population is not essentially driven by individual sorting.

As for specialization, Table 3-4 (bottom panel) shows that, if any, on average there is a mild evidence of negative sorting of skilled workers along the specialization dimension, since the difference between high and low specialized areas is marginal (-0.021). When analyzing separately the different workers' categories, it is possible to see that there is no sorting for formal workers, while the informal ones display a strong 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 (that might more than offset other positive channels).

Summing up, the picture on skills sorting shows that individuals with higher unobservable skills are generally employed in formal jobs and sort in larger cities, thus likely contributing to explain part of their urban wage premium.

Table 3-4. 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

We now aim at investigating 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 that generate while remaining employed in the same job category. To analyze this issue, we use the information on the specific type of occupation of the worker, as we cannot know if he or she remains or changes the firm. Therefore, our analysis will evaluate the impact on wages of a change in type of occupation, that might proxy a better match between workers and firms in terms of tasks, while no addressing the impact of a 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. Our analysis, then, is focused on the search for a wage premium derived from a specific matching linked to the type of occupation, which may occur across and also within a firm. We build a dummy signaling job-change with a value equal to 1 when a worker changes occupation. We consider 2- and 3-digits level occupations defined according to the ISCO88 classification.[42] 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 for the analysis (equation 3.1), now expanded with a job-to-job dummy, also interacted with the informal status of the worker, as follows:

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

$$Log(w_{i,t}) = \alpha + \delta \mathrm{JTJ}_{i,t} + \theta(JTJ_{i,t} * Inf_{i,t}) + \gamma Inf_{i,t}$$
$$+X_{i,t}\beta + \mu Dfirmsize_{i,t} + S_{s(i)} + A_{c(i)} \qquad (3.3)$$
$$+R_{c(i)} + T_t + \varepsilon_{i,t}$$

where i denotes the worker, c the city, s the sector and t the time. $Log(W_{it})$ is the log of real wage per hour of the 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 the worker i at time t. The other control variables are the same as in equation (3.1) and (3.2)), but for the occupation dummies that are now introduced at 2- or 3-digit level according to the relative considered job-to-job digit level. Our parameters of interest are δ and θ , which report the elasticities of wages with respect to a job-change (compared to a no job-change) and its differential impact in case the worker is informal. We run an OLS regression, because for most individuals we have the availability of only two observations, with one lost when the JTJ dummy is computed. Standard errors are clustered at the individual level. With this specification we can see whether wages react positively to a change in occupation, thus suggesting wage increases due to better quality match, compared to the case of remaining in the same kind of occupation. If, on the contrary, there is a penalization due to the 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 understand the channels behind the spatial wage premium. In addition, we decided to focus on workers not changing their formal or informal status when having a job-change. This way we do not to mix the impact of an occupation change with the one of a change in formality status.[43]

[43] Workers who change status between formal and informal or the other way around represent 18% of the total number of workers. We carried out a robustness check using all sample observations and results remain robust, even if the impacts are lower in magnitude (Table A3-2 in the Appendix).

Table 3-5 shows the results. Columns (1), (2) and (3) present the estimates when defining the occupation change at the 2-digits level, while columns (4), (5) and (6) when defining the occupation change at the 3-digits level.[44] 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-digits level. At 3-digits level the wage premia 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 explanation behind the urban wage premium for this category of workers. In high-specialized areas, the dummy for the job change has still a positive impact, but smaller (around 2/2.7%) and similar to the average.

As for informal workers, patterns are different: on average a change in occupation penalizes wages compared to remaining in the same job: the total effect is -2.8% in column (1) and -3.9% in column (4)). Moreover, the wage penalty increases as the size of the FUAs get larger (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 significantly 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 in high density areas skilled workers gain relatively more from a job change, while unskilled workers get advantage from positive knowledge spillovers.

[44] As previously mentioned, the data have a break into the ISCO classification between 2012 and 2013. We have 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 until 2012 and ISCO08 from 2013) to compute the job change, while at the same time excluding the observations affected during the change period. Results are shown in Table A3-3 in the Appendix and remain robust to the considered classification.

To sum up, the analysis of the channels through which spatial wage premiums arise are different 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 opposite, for informal workers the gains from agglomeration are mostly driven by positive knowledge spillovers when employed in the same occupation.

As for specialization, the channels driving previous results seem to be still driven by job-quality matching for formal workers, while for informal workers there is no significant difference between job-matching effects and learning effects within occupation. In addition, there is evidence of negative sorting for this category of workers, such that those characterized by lower unobservable skills sort into highly-specialized areas, which might help to explain the net negative impact detected for this workers' category.

Table 3-5. Job-change impacts on wages by formal/informal workers and differently populated/specialized areas

	(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)

Continue on next page…

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-digits level. From column (1) to (3) the job change is considered at 2-digits level, from column (4) to (6) the job change is considered at 3-digits level.

3.5 Conclusions

In this chapter, 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 uncovered positive and significant impacts on wages deriving from spatial agglomeration. In particular, the elasticity of wages with respect to total population size is as large as 3.8%, while the one with respect to specialization is 0.9%.

Moreover, we have also addressed the role of the informal economy in this relationship, by 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 labor 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 case of specialization, also cancelled. This points out the importance of considering the role of the duality of the labor market in addressing the relationship between wages and spatial externalities.

Furthermore, we have also tried to identify the channels through which these wage gains occur across workers' categories. Results show that for formal workers there is evidence of sorting and of higher gains deriving from a better job-match in highly-populated areas, as opposed to learning advantages for informal workers. In highly-specialized areas, the wage premium due to matching effects still hold for formal workers, while the learning and matching effects are similar for informal workers. In addition, the latter appear to be negatively sorted in these areas, thus explaining the wage penalization uncovered for this category of workers along the specialization dimension. Taking into account that formal workers are relatively more high-skilled and high-educated than informal workers, these findings are in line with previous literature pointing out better gains due to good matches for high skilled workers and higher advantages due to learning externalities for low skilled workers (Matano and Naticchioni, 2016).

References

Au, C.C. & Henderson, V. (2006). Are Chinese Cities Too Small?. Review of Economic Studies, 73 (3), 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.

Carlsen, F., Rattsø, J. & Stokke, H. E. (2016). Education, experience, and urban wage premium. Journal of Regional Science and Urban Economics, 60 (1), 30-49.

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 (1), 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 (ed.) 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 Handbook of Regional and Urban Economics, 5 ed. J.V. Henderson and J.-F. Thisse (Amsterdam: Elsevier-North-Holland), 247–348.

D' Costa, S. & Overman, H. G. (2014). The urban wage growth premium: Sorting or learning?. Journal of Regional Science and Urban Economics, 48 (1), 168–179.

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, (ed) 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 Handbook of Regional and Urban Economics, 4 ed. J.V. Henderson and J.-F. Thisse (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. (Ed.), 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, David 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. & Shleifer, A. (2014). Informality and Development. Journal of Economics Perspectives, 28 (3), 109-126.

Lall, S.V., Shalizi, Z. & Deichmann, U. (2004). Agglomeration Economies in the Indian Industry. Journal of Development Economics, 73 (2), 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.

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. & Venables, A. (2005). Cities in the Developing World. EP 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. & Strange W.C. (2004). Evidence on the nature and sources of agglomeration Economies. In Handbook of Regional and Urban Economics, 4 ed. J.V. Henderson and J.-F. Thisse (Amsterdam: Elsevier-North-Holland), 2119–2171.

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

United Nations. (2014). World Economic Situation and Prospects, WESP. United Nations Publication Sales No. E.14.II.C.2, ISBN: 978-92-1-109168-7, 198.

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

Appendix of Chapter Three

Figure A3-1. FUAs in Ecuador

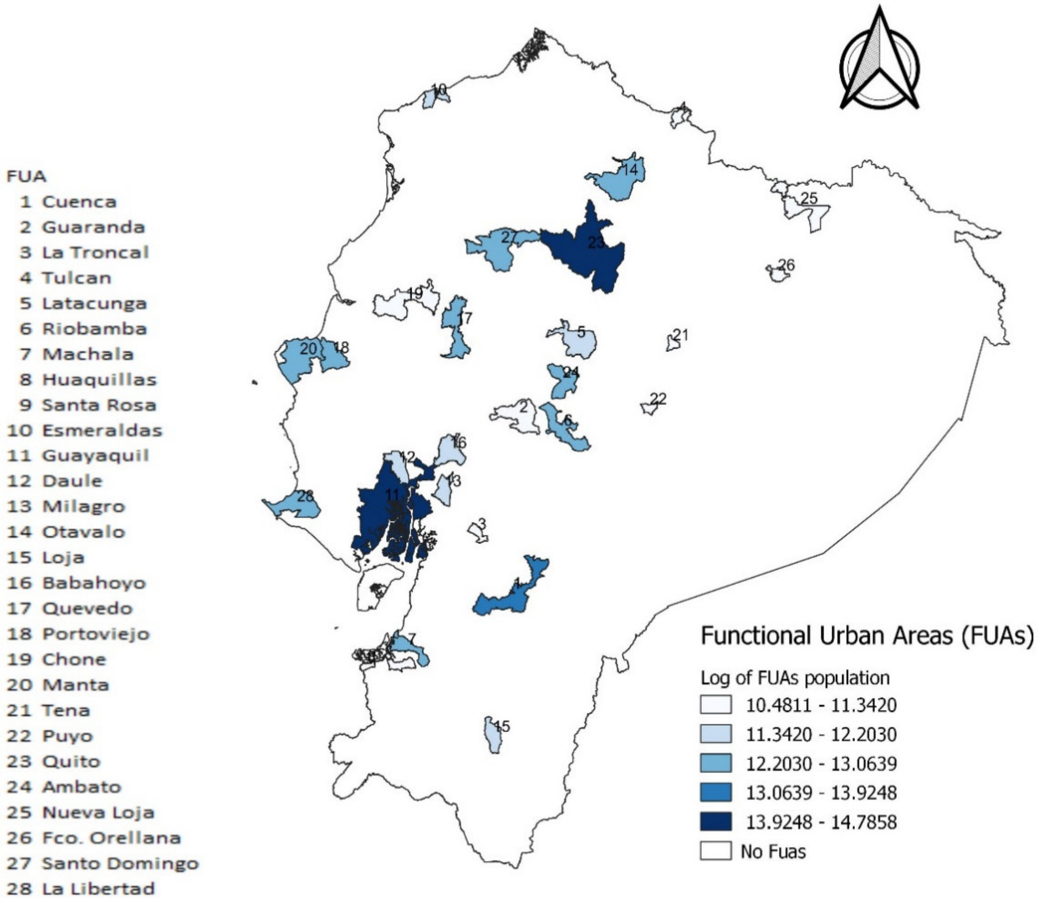


Table A3-1. OLS and IV regressions of wage on the spatial variables and interaction with high education.
Dependent variable: log of real worker's wage per hour

	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) 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***

Continue on next page…

		(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-digits level, while sector according to ISIC 3.1, two-digits 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 A3-2. Job change impacts on wage by formal/informal workers and size area, including workers who change status between formal and informal when having the 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-digits level), migrant status, rural area, macro-areas and time. From column (1) to (3) the job change is considered at 2-digits level, from column (4) to (6) the job change is considered at 3-digits level.

Table A3-3. Job change impacts on wage 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-digits level), migrant status, rural area, macro-areas and time. From column (1) to (3) the job change is considered at 2-digits level, from column (4) to (6) the job change is considered at 3-digits level.

Chapter Four: Agglomeration Effects and Slums

4.1 Introduction

As we have seen in previous chapters, cities are centers of production, enjoying a list of the advantages derived from the three main sources of agglomeration: learning, sharing and matching. From this point of view, the demand for density comes then from workers and firms, supply and demand of production. On the contrary, costs of cities are usually seen from the consumption side: workers face higher rents in large cities, they commute longer, suffer higher crime rates and pollution. As Glaeser et al. (2001) stressed, cities are usually seen as being good for production and bad for consumption.

Nevertheless, such view is far from reality, as individuals can also enjoy the agglomeration economies described by Duranton and Puga (2004) from a consumption point of view. As cities get bigger they might play a superior function in the territorial system (the neoclassical supply-oriented dynamic approach, Camagni et al., 1989, Royuela and Suriñach, 2005) and can enjoy a large variety of consumer goods and personal services or better and superior public services, such as universities or large and good hospitals. These can be linked to the concept of sharing. Glaeser et al. (2001) also list as a consumption advantage of cities the way they allow for facilitating enjoyable social contact or finding a couple, what can be easily linked with the idea of matching (Costa and Kahn, 2000, Puga, 2010). Social learning has been also demonstrated in improved efficiency in capital-intensive services such as utility systems or public works, in larger cities (Holzer et al., 2009).

Nevertheless, some of these advantages become disadvantages once city sizes is beyond a certain threshold. For instance, Holzer et al. (2009) report that for labor-intensive services, such as police work, bigger cities are less efficient. In the same vein, agglomeration economies can be fully exploited only if personal contact is facilitated for instance by improved connectivity (Castells-Quintana and Royuela, 2018). Finally, it is not certain that larger cities do always provide a better and more pleasant physical setting or upgraded public services. Over the last half century, we have seen a massive growth of cities in developing countries characterized by informal settlements, also called slums. According to UN-Habitat (2003) a household is a slum-dweller if it lacks one or more of the following five elements: access to adequate drinking water, access to adequate sanitation, housing with adequate space, housing with adequate structure to protect against climatic conditions and secured tenure. Approximately, 924 million people live in slums or informal settlements, and they provide shelter to a third of urban residents (UN-Habitat, 2015).

Figure 4-1 presents the association between urbanization and the slum rate for 138 countries in 2014.[45] The left panel displays a negative association between the urbanization rate and the slums rate: countries with higher urbanization rates present lower slum rates. However, the right panel of Figure 4-1 plots a positive association between urban primacy (percentage of population in the largest city with respect to the total urban population of the country) and the slum rate. In this case, a higher concentration of the urban population in the largest city seems associated to a higher presence of slum rate.

Figure 4-1. Urbanization rate and slum rate in the world

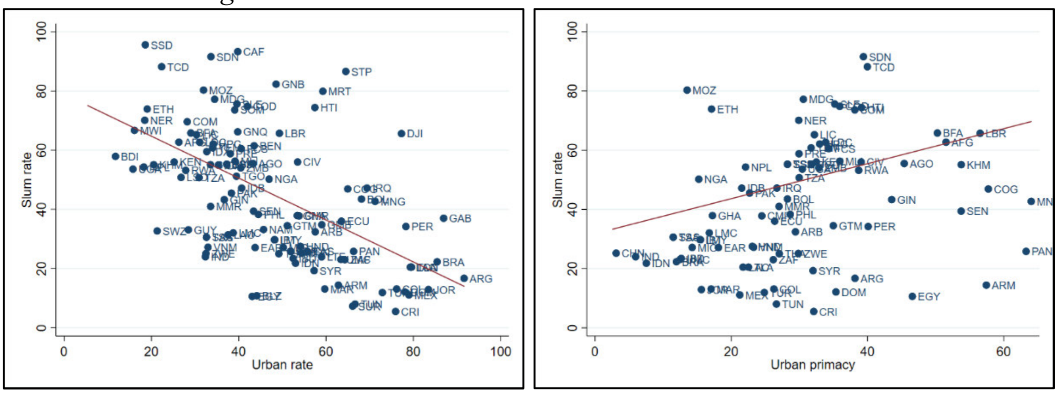
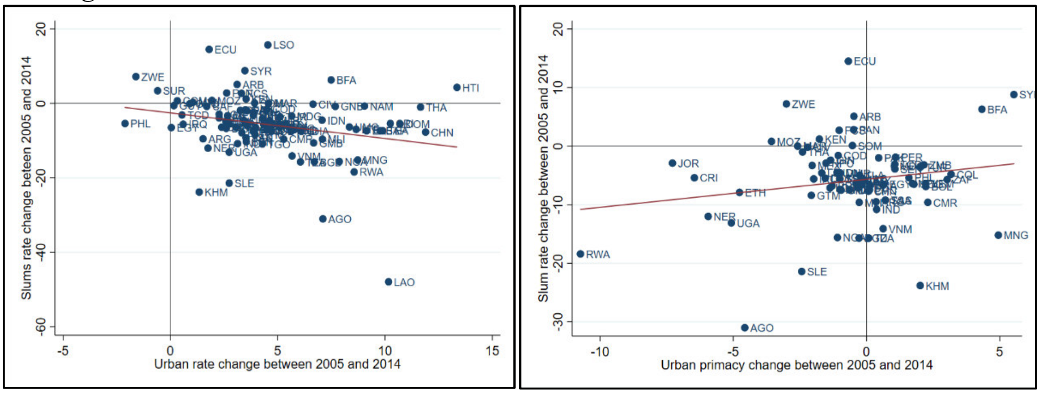


Figure 4-2 shows the association between the growth of both urbanization measures and the evolution (growth between 2005 and 2014) of the slums rate. Again, the left panel reports a negative association while the right panel presents a positive one. It is reasonable, then, to investigate the link between the size of cities and the impact on the type of housing that cities produce, and households consume.

[45] The Information for the Figure 4-1 and the Figure 4-2 are gathered from the UN-Habitat and the World Bank, data available at http://urbandata.unhabitat.org/ and https://data.worldbank.org/.

The analysis provided in this chapter is focused in Ecuador. There is no official national identification of slums in this country. Nevertheless, UN-Habitat provides information of slums for two years, 2005 and 2014. In 2005 Ecuador had a slums rate around 20%, which climbed to 35% in 2014, an opposite trend to what happens on average in Latin America, where the sums rate of Latin America decreased from around 26% in 2005 to 20% in 2014.

Figure 4-2. Variation in the urbanization and slum rate between 2005 and 2014.



The main aim of this chapter is to analyze the association between urban agglomeration and the development of slums. In order to do that, we first review the literature on urban quality of life in general and slums in particular. Next, we analyze the theoretical arguments behind the association between city size and urban well-being. In the subsequent sections we review our case study, Ecuador, the empirical approach, the results and finally the main findings and conclusions.

4.2 Related literature

The framework of agglomeration economies has mainly focused on the production side, arguing that, on average, firms are more productive in larger cities (Combes et al., 2008; Duranton and Puga, 2004). However, too little attention has been paid to the role of cities as centers of consumption. The role of cities as promoters of consumption is as important as their role for production.

Glaeser et al. (2001) set up a framework on which cities produce four particularly amenities. The first one is a larger and richer variety of services and goods. The second is the aesthetics and physical setting of the cities, which plays an important role for cities to become more attractive. The third amenity is the quality of public services such as good schools, universities or hospitals. The fourth amenity is the speed, which is linked to the sufficient capability of cities for decentralization and increasing commuting distance in an efficient way. Cities can prevail and grow as far as benefits are higher than costs of the agglomeration. And this is true both for production and for consumption. Consequently, we understand that cities' attractiveness is not only a matter of productivity but also a subject of the expected quality of life for individuals, including both poor and rich people.

Quality of life (QoL) is not an easy concept to address. We follow here the concept offered by the International Society for Quality-of-Life Studies: 'Quality of life usually refers to the degree to which a person's life is desirable versus undesirable, often with an emphasis on external components, such as environmental factors and income' (Diener & Ryan, 2009, p. 401). Despite the stronger emphasis on objective issues (circumstances of a person's life) rather than his or her reaction to those circumstances, several scholars also include subjective and psychological elements (Royuela et al., 2003; Liu, 1978).

A very important feature of QoL is its multidimensionality, which results in the need of building a composite index of the concept. As stressed by Saisana (2014), multidimensional measures give signals of society's performance in complex fields. There is a vast list of such indices in the field of well-being and QoL, including the well-known Human Development Index (HDI) and more recently the one developed by the OECD (2013b).[46]

OECD (2013b) analyses the concept of well-being by looking at two main components. The first one looks at *current* well-being, proxying material conditions (income and wealth, jobs and earnings, housing conditions) and quality of life (health status, work life balance, education and skills, social connections, civic engagement, governance, environmental quality, personal security, and subjective well-being). The second component, *future* well-being, considers resources that drive well-being over time and that are

[46] Other multidisciplinary indices analyze aspects such as poverty and deprivation, which are also linked to QoL as they measure the lack of basic needs for a group of the society (Durán and Condorí, 2017; Cabrera-Barona et al., 2016; Havard et al., 2008; Narayan, 2000). For technical aspects linked with the construction of composite indices see Athanassoglou (2015) and Decancq and Lugo (2013).

measured by types of capital, including human, economic, social and natural capital.

According to Combes et al. (2012) cities increase amenities as they grow in size, while Glaeser et al. (2016) argue that megacities prevail thanks to the role of amenities, which enjoy spillovers both across time and space.[47] Mitra and Nagar (2018) remembers how in the size-productivity literature, infrastructure is a major driver of agglomeration economies: "Enhanced government, private and household investments result in improved living standards including accessibility to various amenities" (Mitra and Nagar, 2018, p. 273). Negative externalities arise in the form of high costs of transport and land, and also in difficulty to access housing. Several works display a nonlinear relationship: Frick and Rodríguez-Pose (2017) use a panel of 113 countries between 1980 and 2010 and show that as population concentration grows the benefits of agglomeration declines. In the same line, and also using data at the national level, Castells-Quintana (2016) reports that the channel that brakes such positive link is the lack of adequate infrastructures.

What is not addressed in the urban economics literature, though, is why such negative externalities arise. Why do cities face problems in providing adequate access to housing when they grow in size. The development literature, though, has addressed this aspect. Fox (2014) explains three conditions for the slums creations. The first one is a demographic explanation due to the fast urbanization growth. Cities with fast population growth face an excessive demand in the short run, while housing supply is typically inelastic, what results in the creation of informal settlements with inadequate conditions.

The second is an economic explanation: urban poverty. Even if housing supply were elastic, poor households would not be able to access such dwellings or there will be no developer willing to invest in housing meeting international standards, as there will be no demand with the adequate purchasing power.

The third argument proposed by Fox (2014) is the institutional explanation, which he links to underinvestment in housing and infrastructure

[47] They use as an example the aesthetics of one's residence or other amenities potentially produced collectively, such as a community club.

stock. Examples of institutional failures harming the development of housing are rigid land use regulations, poorly defined and enforced property rights. He also cites that institutional issues are a factor inhibiting public investments because informal settlements, generally illegal in planning regulations and property right, discourage public investments in infrastructure either because such settlements are ineligible for investment or because public authorities fear that public investment would constitute tacit recognition of legitimate those informal settlements and therefore, encourage more informal settlements. He also stresses the importance of the historical and political dynamics to shape urban development trajectories: most developing countries were colonies in which the interest was more in obtaining their own benefits than establishing a strong foundation for urban development by means of investment and improving institution quality. Thus, he shows that colonial era investments and institutions reflected in the contemporary variation in slums rate.

The creation of slums not only reflects an incapacity of the city to cope with a diverse demand for social-economic needs. It also directly affects the overall quality of life of living in cities due to the lack of planning and management of the urban areas, and of course, the coverage of basic infrastructure offered by the cities. Marx et al. (2013) explain that most people born in the slums have hardly improved their living standards over time. In this line, slums in developing countries are considered a form of poverty trap for most of their residents. The importance of the basic infrastructure and connectivity plays a relevant role for the correct work of cities and the generation of agglomeration effects because of the lack of covering basic needs reduces the capacity of cities to develop, attract talent and investment, and pick down the agglomeration effects (Castells-Quintana, 2016; Castells-Quintana & Royuela, 2018; World Bank, 2011). Still, also here the literature shows conflicting results.

As for the economic process, Frankenhoff (1967) and Turner (1969) argue that slums are the product of and the vehicle for activities that are essential in the process of modernization. Glaeser (2011) also supports the idea that slums might bring opportunities to poor and non-educated workers to benefit from agglomeration effects. According to their view, slums due to fast urbanization in developing countries are a pre-phase of development. As time goes by, cities can adjust to the presence of the consolidate slum settlements

(United Nations, 2015b): in a first phase the migrants cannot afford to buy or rent decent housing, but as they become part of the urban economy, their benefits from agglomeration (e.g. higher income, better education) would let them to improve their initial condition until to reach the standards of the society. Alike to that vision, Harris-Todaro (1970) models suggest restricting the mobility from rural to urban areas as a solution for slums. Lipton (1977) and Bates (1981) consider the pro-urban bias policies as a cause for slum, as government allocate a disproportionate share of public resources in urban areas, fostering rural-urban migration. These arguments promoted the development of pro-rural and spatially balanced regional policies, what Fox labels as ‘anti-urban policies’.

Several works have criticized such stream of thought (Lall et al., 2006, Ravallion et al., 2007) under the argument that there are no visible gains of such restrictions. Still, according to Fox, the stronger counter argument is the negligible effects of such policies to prevent rural-urban migration in developing countries.

The recent literature emphasizes the need to deepen the attention on the association between agglomeration effects and slums (Mitra & Nagar 2018; Glaeser & Henderson, 2017; Combes & Gobillon, 2015). We do not know if the channels of agglomeration vary in the presence of slums and if they work properly. In other words, slum areas as new extension of cities characterized by the lack of basic infrastructure would have an impact on the sharing, matching and learning mechanisms for production in cities, and of course, would impact cities’ amenities because it increases the need of local public investments in services, aesthetics and physical setting of the cities and a huge goal for connectivity of the new zones with the center. The empirical evidence of this association is still scarce. Brueckner (2013) shows that higher income and education lead to occupancy of dwellings with less slum characteristics, and this effect is reinforced with lower levels of fertility. Galiani et al. (2017) show that improving physical slum conditions have an important impact on the well-being of extremely poor people. Mitra and Nagar (2018) explore the relationship between city size and slums characteristics and find a negative relationship between city size and slums characteristics, although they increase at some point for the largest cities. In our view, then, there is space for further investigation on this association.

4.3 Case study and data

This chapter uses 28 FUAs of Ecuador identified in chapter two. The FUAs cover the idea of economic integrated cities and, as such, they are preferred in the literature of city size (Ahrend et al., 2017; Veneri, 2017; 2016; OECD, 2013a; 2012). In the Appendix, Figure A4-1 shows the urban primacy structure in Ecuador.

The urban concentration is presented in two main urban areas: Guayaquil and Quito. The Ecuadorian urban system is mostly dominated by small FUAs and their urban structure is similar to the full sample of FUAs of the OECD countries and Colombia.[48] Also, the highest urban growth in Ecuador occurred between 1960 and 1980 (Obaco and Díaz, 2018; Villacís et al., 2001). Royuela and Ordoñez (2018) analyze the migration rate in the provinces of Ecuador in the period 1982-2010. They confirm the concentration of the population in two main provinces that contain the country's main cities, Guayaquil and Quito. Moreover, they also mention that the urbanization trend of the two largest cities has weakened, to the extent that provinces with the greatest influx of migrants are not necessarily the most populated. Hence, small and medium size cities are becoming increasingly important.

To analyze slums in the case study, we gather the information from the National Institute of Statistics and Census of Ecuador (INEC).[49] We use the population censuses in 1990, 2001 and 2010 to evaluate the changes in the slums indicators the across-censuses in Ecuador. The censuses will give information on many household's characteristics across 30 years. We have around 5.3 million of observations of households available for the period of analysis.

Although Ecuador does not provide national information on slums, the UN-Habitat cites that around 35 per cent of the urban population of Ecuador live in slums. According to UN-Habitat (2003) a household is considered as being in slum condition if it has one or more of the following conditions:

[48] See the FUAs information at http://www.oecd.org/cfe/regional-policy/functionalurbanareasbycountry.htm and chapter two.
[49] The data are available at http://www.ecuadorencifras.gob.ec/banco-de-informacion/. The first census of population was in 1950. However, the full censuses digitalized and available to download are from 1990 to 2010.

(1) Access to safe water: sufficient amount of water (20 liters per person and day), at an affordable price (less than 10 % of total household income), available without being subject to extreme effort (less than one hour a day of walking time);
(2) Access to improved sanitation: access to an excretal disposal system, either in the form of a private toilet or a public toilet shared with a reasonable number of people;
(3) Sufficient living area: less than three people per habitable room;
(4) Structural quality/durability of dwellings: a house built on a non-hazardous location and with a permanent structure adequate to protect its inhabitants from extreme climatic conditions; and
(5) Security of tenure: the right to effective protection by the state against arbitrary unlawful evictions.

Not all the specific conditions mentioned by UN-Habitat are available in the censuses. In this chapter, we approximate the concept of slums defined by the UN-Habitat by means of the available information, which considers four main blocks:

(A) Lack of access to safe water in a household: considers the lack of pipe water and lack of public water supply in the households;
(B) Lack of improved sanitation: considers the lack of sewage system, lack of power supply, and lack of public garbage service;
(C) Overcrowding rate: three or more people per bedroom.
(D) Non-durable housing materials: considers the use of non-durables materials in roof, walls and floor of the households.

Table 4-1 summarizes the measures (with associated concepts) used to proxy slums characteristics with the censuses data form Ecuador. In detail, we use a total of nine censuses variables. We define the slum rate as the percentage of households with at least one of these characteristics. Under this scheme, we find that the 93% of households in rural areas (no FUAs) have at least one slum characteristic, while 62% of households in FUAs have at least one slum characteristic. Moreover, the slum rate in the rural areas have slightly decreased over the period of analysis, from 96% in 1990 to 93% in 2010, while the urban areas such decreased was faster, from 67% in 1990 to 62% in 2010. A detailed proportion of the slum characteristics is presented in the Appendix (Tables A1 to A3). Finally, the slums distribution is not even across regions: the Andean highlands region presents around 50% of slum rate, the Coastal region has around 72% of slum rate and the Amazon region displays 82% of slum rate. These figures strongly contrast with that reported by UN-Habitat for Ecuador in 2014. Patel et al. (2014) explain that varying definitions of a slum might result in different population estimates of the slums' incidence. These authors show the differentiated rates obtained in India by using the UN-Habitat criterion versus the approach considered by the Indian census, which assumes a geographical continuous criteria of slum households (Census of India, 2001).

Table 4-1. UN-Habitat concepts of slums and available census information.

N.-	Measure	Concept	
1	No piper water	Not access to water trough pipes	Block A
2	No public water supply	Not access to public water supply	
3	No sewerage system	Not access to sewerage system	Block B
4	No energy supply	No energy supply	
5	No garbage collection	No public garbage collection	
6	Overcrowding	Three or more people per bedroom	Block C
7	Non-durable roof materials	Roofs made of straw or worse	Block D
8	Non-durable wall materials	Walls made of wood or worse	
9	Non-durable floor materials	Floors made of cane or worse	

Note: Information gathered from the Ecuadorean censuses in 1990, 2001 and 2010.

In this work we have decided to go beyond to the traditional dichotomy identification of slum conditions. We follow Patel et al. (2014) and measure the severity in the slum condition. We build an index to measure the different levels of slums characteristics.[50] First, we identify every slum condition by

[50] The use of indexes is widely used in the literature (Decancq and Lugo, 2013; OECD, 2013b; OECD, 2008). The basic structure of an index can be considered as, $I = \sum \omega_i X_i$, where ω_i is the weights associated to each variable X. The problem of using an index is divided in two main concerns: the variables to include in the index and the weights to compute the index. The main idea of an index is to cover a good set of indicators that are related to what is expected to measure that can be summarized into one index. The variables for the index will depend on the analysis to carry out, while the weight shows the relative importance of the components within that index (Athanassoglou, 2015). However, the weights for an index are always in discussion as the final score of an index could be very sensible to its weights. An alternative methodology is to "allow the data to talk", as in

means of a dummy variable. To be clear, if the household has any of the slum characteristics presented in Table 4-1, the indicator will have a value equal to one, and zero otherwise. Using the initial vector of housing characteristics, we apply two alternative approaches to build the index.

- We build an additive slum severity index (ADD) by means of a simple sum of all dummies capturing the slums characteristics.[51] The index goes from zero, total absence of any slums characteristic, to nine, the maximum of slum's characteristics. This approach considers evenly distributed weights, meaning that the lack of any of these basic services is as bad as the other characteristics.
- The second considered alternative is the use of Principal Components Analysis (PCA). This methodology aims at capturing most of the variance of a whole set of variables into a few indicators. After standardizing the variables, we apply the PCA. We use the first two components, which are the ones with eigenvalues larger than one, which capture around 50% of the total variance (Table A4-4 in the Appendix shows the result from the rotated principal components). The first component, PC1, covers most of the variance of blocks A, B, and C, and we label it as the *lack of basic infrastructure coverage index*. The second component, PC2, is labelled as the *bad housing physical quality index* as it covers mostly the variance of the block D. In both cases, and similar to the interpretation of the slum severity index (ADD), they can be interpreted in negative terms, this is: the higher the indexes, the worse the living conditions.

The correlation between the slum severity index (ADD) and the lack of basic infrastructure coverage index (PC1) is around 94%, while the correlation between the slum severity index (ADD) and the bad housing physical quality index (PC2) is 53%. Finally, the Table A4-5 in the Appendix presents some descriptive statistics of the indexes at the city level. On average, an Ecuadorean household living in urban areas presents at least one slum characteristic during the period of analysis. Figure A4-2 in the Appendix plots the relative frequencies of the slum severity index.

Principal Components Analysis (PCA), which combine variables into a set of indicators order by the largest share of variance of the original data they explain. There are also available different methodologies to build indexes (see OECD, 2008).

[51] This procedure has been applied before for Colombia (Duque et al., 2012).

Figure 4-3 introduces the association between the indices and city size, capturing the incidence of slums and the average size of cities. The left and right panels show the association between urban size and, respectively, the slum severity index (ADD), and the lack of basic infrastructure coverage index (PC1). A negative association between the city size and both the indexes is found, showing how, larger cities offer, on average, higher material housing standards of living. This association is in line with the literature, reporting that larger cities tend to offer, on average, higher quality of life.

Figure 4-3. Average of slum severity index and the lack of basic infrastructure coverage index on the city population

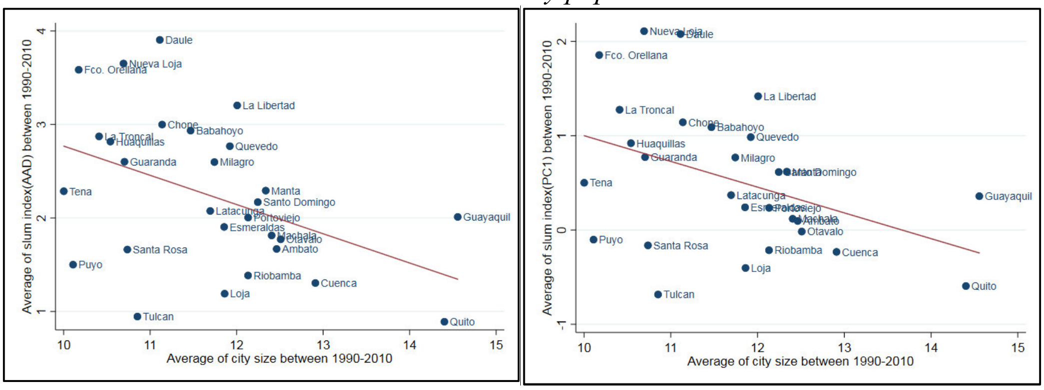
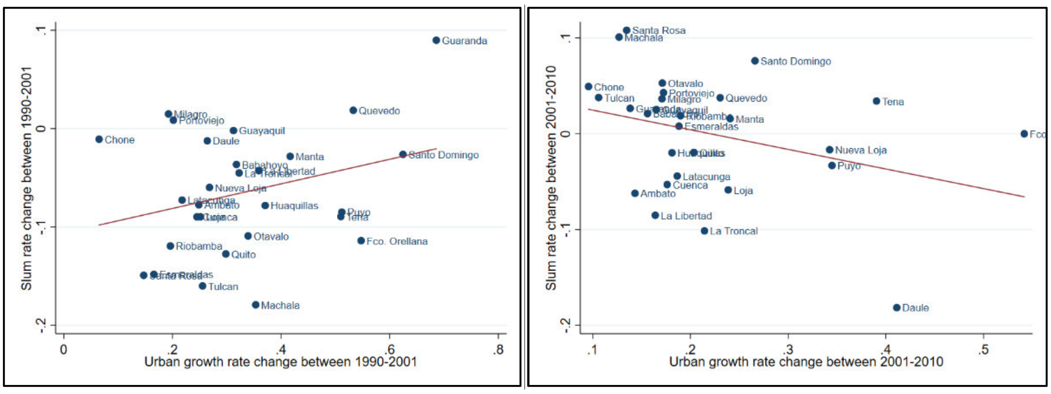


Figure 4-4. Variations in the urban and proportion of slum households during the period 1990 – 2001-2010



Next, we look at the growth rates in Figure 4-4. In the period 1990-2001 the urban growth was parallel to an increase in the slum severity index (ADD). On the contrary, between 2001 and 2010 urban growth was linked to an average decrease in this index. In both panels of Figure 4-4 we can see how most of the urban growth is mainly driven by the population growth of small cities, while there is no clear pattern behind the evolution of the slum severity index.

4.4 Empirical approach

Once described the evolution of urbanization and slums, our next step is to analyze in depth the causes behind the evolution of the slums indexes on Ecuadorean cities. We focus on analyzing the agglomeration effects on the slum indexes. But, we also explore two of the factors driving the creation of slums proposed by Fox: poverty and rapid urban growth. First, we analyze agglomeration effects and poverty. In order to do that, we analyze the individual characteristics of the households together with city size. We pay special attention to education level because it is a good proxy for income, thus it addresses poverty in urban areas at least in a part. The specification is the following:

$$Slum\ Index_{i,c,t} = \delta\ Lpop_{c,t} + \beta' X_{i,c,t} + Prov_c + T_t + e_{i,c,t} \qquad (4.1)$$

Where, $Slum\ Index_{i,c,t}$ for the household i, in city c, at time t, refers to each of the slum indexes presented above: the slum severity index (ADD), the lack of basic infrastructure coverage index (PC1), and the bad housing physical quality index, (PC2). $Lpop_{c,t}$ is the log of city population, which parameter, δ, captures the agglomeration effects on the slum indexes. $X_{i,c,t}$ is the vector of socio-demographic household's characteristic related to the head of the household. We use female, age, four levels of education (literacy or lower, primary, secondary, and university or higher), marital status (free-union, single, married, divorced, and widow), number of members in the household younger than 15 years old (Children). We also control migration identifying the households that lived in a different city 5 years before the census (Migration). $Prov_c$ and T_t are province and time fixed effects, respectively. Finally, to aim at overcoming endogeneity problems we use the population of Ecuador in 1950 as instrument for the current city size (Combes at el., 2010; Ciccone and Hall, 1996).

As for the impact of rapid urban growth on the incidence of slum, we focus on the largest city of Ecuador, Guayaquil. We analyze the physical growth of Guayaquil by means of comparing the presence of new areas created across the three considered censuses using Geographical Information System (GIS). We use the 2010 household's data and analyze the incidence of slum index in the new and old areas of Guayaquil. In the year 2010, there are 652,068 households that live in Guayaquil. By comparing the maps, we can identify the areas of Guayaquil created between 1990 and 2010.[52] We plan to estimate the following regression at the household level:

$$Slum\ Index_i = \theta_1 D_{1990\_2010} + \theta_2 D_{2001\_2010} + \beta' X_i + e_i \qquad (4.2)$$

Where, again, $X_{i,c,t}$ is a vector of individual characteristics related to the head of the households as introduced above. D_{1990_2010} is a dummy variable that identifies all the areas of Guayaquil created between 1990 and 2010. Thus, the parameter θ_1 evaluates the differences in means of the 2010 slum indexes of 122,514 households that live in areas created between 1990 and 2010 with respect to 529,554 households that live in the already existing areas of Guayaquil in 1990. Instead, D_{2001_2010} is a subset from the previous dummy where it receives the value of one for the "newest" areas of Guayaquil between 2001 and 2010. Thus, adding the parameters θ_1 and θ_2 would allow to evaluate the differences in means of the 2010 slum index of 61,557 households that live in areas created between 2001 and 2010, with respect to the 2010 slum indexes of remaining households that live in existing areas of Guayaquil.

[52] We are not able to trace the blocks over the censuses, as they change the codes. Besides, the only available digitalized map is the one of 2010. Consequently, we only have been able to identify which areas present in the 2010 census that were not present in 2001 and in 1990. Moreover, Guayaquil is also a good particular case, because it presents higher heterogeneity in deprivation zones compared with other cities of high urban primacy (Obaco & Ballas, 2018).

4.5 Results

4.5.1 Agglomeration effects

Table 4-2 presents the results of the estimates of equation (4.1). Columns (1) to (6) reports the results using the slum severity index (ADD) as dependent variable. All specifications are OLS with the exception of the last two columns, which report IV estimates. Finally, the last column excludes the two largest cities, Guayaquil and Quito, from the estimation.

Column (1) of table 4-2 presents the regression of the slum severity index on the log of city size, province and time dummies only, therefore still not addressing the role of poverty. The agglomeration effects are negative and significant, meaning that doubling city size implies a decrease of the slum severity index of 0.34. The subsequent columns include individual controls for the head of the household. Demographic controls (columns 2 and 3) are significant and increase the regression adjustment. Education variables, which capture poverty are added in column (4). As expected, the higher the education level, the stronger the reduction in the slum incidence, as expected. Consequently, we can argue that poverty is an important factor of slumness. Nonetheless, the parameter associated to population is still significant and negative, although its impact is around 25% lower. Column (5) presents the IV estimates considering as instrument the log of city population in 1950. The magnitude of the parameter of the log of population slightly decreases, in line with what we found in the previous chapter.

The control variables present the expected signs. Being female or being older is associated with lower level of the slum severity index. This can be potentially associated with international emigration episodes of the male head of households, what has been found as a factor improving housing conditions (Díaz et al., 2018). An increase in the fertility measure, the number of children in the household, is related to a higher level of slum severity index, in line with the development literature, which finds that having more children impede capital accumulation. The migration dummy captures those head of households who lived in a different city five years ago. Contrary to the Harris-Todaro model, this variable presents a negative parameter, which implies that, on average, migrants enjoy better housing conditions than native population. As we have reported, education is the most important determinant of the slum severity index. Households with a head holding a university degree or higher are expected to lower by 2.5 points the slum severity index, with respect to the people with literacy or lower. This means a difference of almost three points in the slum severity index. Finally, column (6) reports the IV estimate once we exclude the largest cities of Quito and Guayaquil. We find that the magnitude of the agglomeration effects slightly decreases by 0.013 points, what suggests evidence of congestion in the two largest cities in the slum severity index.

Table 4-3 presents the estimates for the alternative slum indexes from column (1) to column (6). Table 4-3 follows the same structure of table 4-2, but the results of the lack of basic infrastructure coverage index (PC1) are presented from column (1) to column (3), while the results of the bad housing physical quality index (PC2) are presented from column (4) to column (6). As for the lack of basic infrastructure index, we find a pattern similar to that found for the ADD index in table 4-2 (the correlation between these two variables was over 0.9). The magnitude of the agglomeration effects is around -0.21 on this index (see column (2) in table 4-3). Female, age and education level are negatively associated with the level of the lack of basic infrastructure coverage index. Again, education level has the largest impact on the lack of basic infrastructure coverage index. More children in the households are positively associated with a higher level of the lack of basic infrastructure coverage index. Similarly, migrated households are negatively associated with the level of lack of basic infrastructure coverage index. When we exclude Quito and Guayaquil from the regression, column (3) in table 4-3, the magnitude of the agglomeration effects increases. This would suggest that for this index, the largest cities contribute more to the coverage of basic infrastructure to the households. This is with the idea that larger cities offer better basic infrastructure.

As for the bad housing physical quality index (PC2), the results are consistent to those previous estimation, but to a lesser extent (column (5) in table 4-3). Agglomeration effects are again negatively associated with higher level of the bad housing physical quality index, but its magnitude is lower than in the previous estimations. The other covariates are also presenting lower magnitude that in the other two indexes, with the exception of having more children in the households, which magnitude remains similar. In this regression, the R-squared is lower than what we found for the other indexes. Once we exclude Quito and Guayaquil, the negative effect of agglomeration effects on this index increases (see column (6)). This result suggests the largest two cities contributes to have worse housing physical characteristics in this index. The bad housing physical quality index (PC2) is more related to the own capability of the people to improve the physical structural material of their households. Thus, we can observe the difference composition of the two indexes, the lack of basic infrastructure coverage index (PC1) and the bad housing physical quality index (PC2) on the role of agglomeration effects.

Our findings point out that poverty is a very important driver of the existence of slums. Households' characteristics account for a good proportion of the explained variance on the indexes. Still, even controlling for them, population size presents a significant and negative parameter, what calls for studying other drivers of the incidence of slums.

Larger cities seem to offer better possibilities to households in terms of infrastructure, but it is harder for these households to improve houses with better quality materials.

Table 4-2. Regression of indexes on the individual and city size
Dependent variable: slum severity index

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	OLS	OLS	OLS	OLS	IV	IV
Lpop	-0.340***	-0.319***	-0.320***	-0.243***	-0.237***	-0.250***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Female dummy		-0.127***	-0.129***	-0.181***	-0.181***	-0.230***
		(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Age		-0.008***	-0.008***	-0.021***	-0.021***	-0.019***
		(5.30e-05)	(5.36e-05)	(5.20e-05)	(5.21e-05)	(8.23e-05)
Children		0.131***	0.130***	0.090***	0.090***	0.079***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Migration			-0.162***	-0.148***	-0.147***	-0.247***
			(0.003)	(0.003)	(0.003)	(0.004)
Marital status category		yes	yes	yes	yes	yes
Education (Literacy or lower as base)						
Primary education				-1.036***	-1.039***	-1.173***
				(0.004)	(0.004)	(0.005)
Secondary education				-1.870***	-1.871***	-2.098***
				(0.004)	(0.004)	(0.005)
University or higher				-2.522***	-2.523***	-2.716***
				(0.004)	(0.004)	(0.005)

Continue on next page…

Province dummy	yes	yes	yes	yes	yes	yes
Time dummy	yes	yes	yes	yes	yes	yes
Constant	5.974***	6.740***	6.787***	7.554***	7.481***	7.688***
	(0.023)	(0.023)	(0.023)	(0.021)	(0.021)	(0.054)
Observations	5,431,589	5,421,154	5,421,154	5,356,678	5,323,835	2,383,150
R-squared	0.137	0.175	0.175	0.292	0.290	0.274
Underidentification test					0.00	0.00
Weak identification test					1.520e+07	731375

Notes: Standard errors clustered at household level in parentheses *** p<0.01, ** p<0.05, * p<0.1. Characteristics related to the head of the household. Differences in observations related to missing values.

Table 4-3. Regression of indexes on the individual and city size

Dependent variable: lack of basic infrastructure coverage index (PC1) and the bad housing physical quality index (PC2)

VARIABLES	(1) PC1-OLS	(2) PC1-IV	(3) PC1-IV	(4) PC2-OLS	(5) PC2-IV	(6) PC2-IV
Lpop	-0.215***	-0.208***	-0.170***	-0.082***	-0.070***	-0.135***
	(0.002)	(0.002)	(0.004)	(0.001)	(0.001)	(0.003)
Female dummy	-0.157***	-0.157***	-0.199***	-0.071***	-0.071***	-0.095***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.001)	(0.002)
Age	-0.016***	-0.016***	-0.014***	-0.007***	-0.007***	-0.007***
	(4.74e-05)	(4.75e-05)	(7.41e-05)	(3.80e-05)	(3.81e-05)	(6.24e-05)
Children	0.045***	0.044***	0.033***	0.040***	0.040***	0.038***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Migration	-0.100***	-0.098***	-0.186***	-0.057***	-0.057***	-0.089***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Marital status categories	yes	yes	yes	yes	yes	yes
Education (Literacy or lower as base)						
Primary education	-0.844***	-0.846***	-0.941***	-0.442***	-0.443***	-0.541***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)
Secondary education	-1.545***	-1.546***	-1.719***	-0.680***	-0.680***	-0.815***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)
University or higher	-2.015***	-2.016***	-2.159***	-0.919***	-0.920***	-1.055***
	(0.004)	(0.004)	(0.005)	(0.003)	(0.003)	(0.004)

Continue on next page…

Province dummy	yes	yes	yes	yes	yes	yes
Time dummy	yes	yes	yes	yes	yes	yes
Constant	5.238***	5.156***	4.611***	1.763***	1.621***	2.670***
	(0.020)	(0.020)	(0.047)	(0.017)	(0.017)	(0.041)
Observations	5,356,678	5,323,835	2,383,150	5,356,678	5,323,835	2,383,150
R-squared	0.266	0.262	0.254	0.092	0.093	0.087
Underidentification test	0.00	0.00	0.00	0.00	0.00	0.00
Weak identification test		1.520e+07	731375		1.520e+07	731375

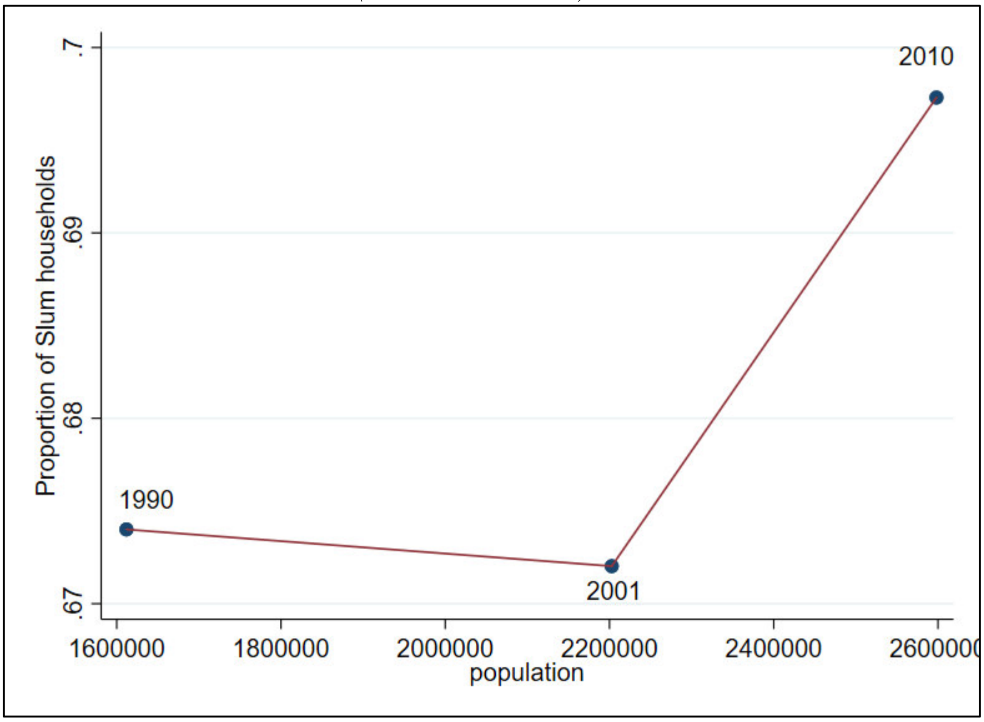
Notes: Standard errors clustered at household level in parentheses *** p<0.01, ** p<0.05, * p<0.1. Characteristics related to the head of the household. Differences in observations related to missing values.

4.5.2 A rapid urban growth: analysis of Guayaquil

To analyze a rapid urbanization as a determinant of slums, we decide to study the magnitude of the slum indexes in the creation of new areas in a city. According to UN-Habitat (2003), slums are generally associated with new areas in the borders of the city. The rapid urban growth named by Fox (2014) as a factor for the creation of slums, is usually associated with the spatial expansion of cities. We have also seen in the introduction of this chapter that rather than urbanization, it is urban primacy a driver of the slum existence. We analyze the largest city of Ecuador, Guayaquil. For our purpose, we only use the urban area associated to the FUA of Guayaquil, which is composed of three municipalities: Guayaquil, Durán and Samborondón. Consequently, we discard the use of the hinterland of the FUA, as this area can be out of the urban core.

Figure 4-5 shows the trend between the urban growth and the slum rate during the period 1990 – 2010. Despite the urban expansion between 1990 and 2001, we observed a small decrease in the proportion of slum households. On the contrary, the proportion of households living in dwellings with some slum characteristics increases considerably between 2001 and 2010.

Figure 4-3. Urban population and slum rate of households in Guayaquil (Period 1990-2010)



To explore deeper into the relationship between urban growth and slum creation, we identify the zones of expansion in the urban areas of Guayaquil during the period 1990 – 2010. Figure 4-6, on the left side, shows the new areas created in Guayaquil between 1990 and 2010, while on the right side of the Figure 4-6 presents the slum severity index in 2010.[53] Clearly, areas created between 1990 and 2010 display the highest levels of slum characteristics compared with the areas already existing in 1990.[54]

Figure 4-4. Spatial distribution of Guayaquil: new areas have been created since 2001 on the left side and the slum severity index of Guayaquil in 2010 on the right side.

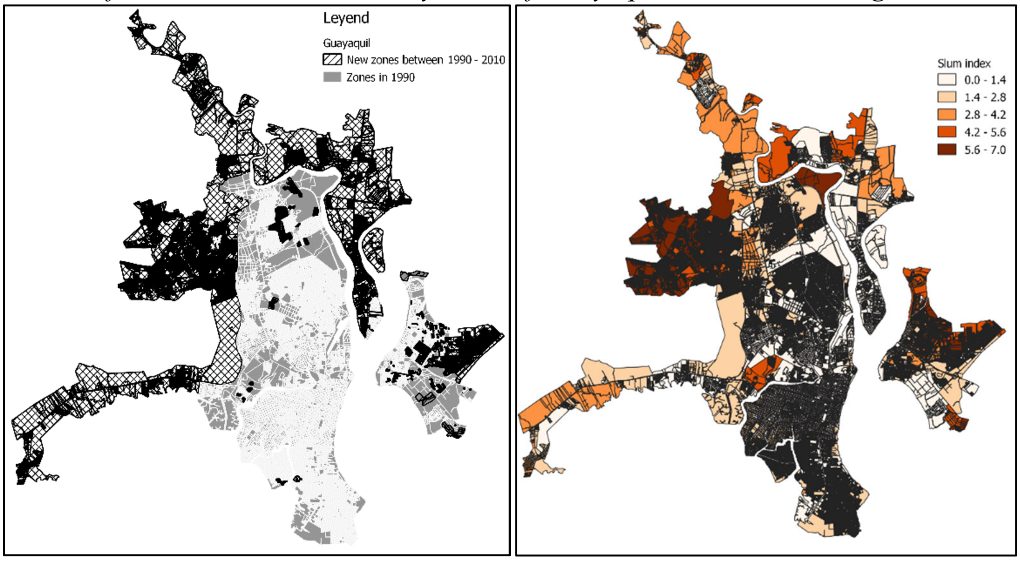


Table 4-4 introduces the slum characteristics between the identifying urban area of Guayaquil. The columns of table 4-4 describe the population, number of households, slum rate and the slum severity index for the urban areas of Guayaquil, in the year 1990 and 2010. Then, in the bottom of table 4-4, we split the household's characteristics in the year 2010 for the existing areas in 1990 and the areas created between 1990 and 2010. The new areas of Guayaquil created between 1990 and 2010 have a large difference than the

[53] Figure A4-3, in the Appendix, shows the different extensions created in the city of Guayaquil by censuses, independently.

[54] Figure A4-4, in the Appendix, introduces the spatial distribution for the lack of basic infrastructure coverage index and the bad housing physical quality index, while figure A4-5 introduces the distribution of the population of Guayaquil at census blocks of 100 households on average.

existing areas in 1990. Considering that the slum characteristics based on the census of 1990 of Guayaquil are the same existing areas in 1990, but evaluated in the 2010, we can observer that there is an improvement from 67.40% in 1990 to 64.20% in 2010 for the existing area of Guayaquil. There is also an improvement in the slum severity index for this area, from 2.38 in 1990 to 1.27 in 2010. Instead, the areas evaluated in 2010 that were created between 1990 and 2010 have a stronger slum component. These slum characteristics are even higher for the areas created between 2001 and 2010.

Table 4-4. Slums characteristics of new areas of Guayaquil

	Population	Households	Slum rate %	Slum Severity Index
Year 1990	1,521,997	340,827	67.40	2.38
Year 2010	2,453,274	665,647	69.72	1.83
Differentiating between areas in the year 2010				
Existing areas in 1990	2,000,805	529,554	64.20	1.27
New areas:				
between 1990 and 2010	452,469	122,514	90.71	4.01
between 2001 and 2010	214,238	61,577	89.93	4.85

After observing the relationship between the new areas of Guayaquil and the level of slum associated with them, we estimate equation (4.2) where the objective of the dummies is to capture the differences in means of the slum indexes in 2010 among the areas created over time.

Table 4-5 presents the estimation of the equation (4.2) that considers the new extension of Guayaquil on the 2010 slum severity index. All the estimates are OLS. Column (1) introduces the dummy D_{1990_2010}, which parameter θ_1 captures the differences in means of the 2010 slum severity index of the 122,514 households living in areas created between 1990 and 2010 with respect to the 529,554 remaining households living in existing areas in 1990. The 2010 slum severity index in the areas of Guayaquil created between 1990 and 2010 have on average around 3 points higher in the slum severity index than the existing areas of Guayaquil in 1990. Column (2) includes in the previous specification the dummy D_{2001_2010}, where adding the parameters θ_1 and θ_2 show the differences in means of the 2010 slum severity index in the areas that are created between 2001 and 2010 compared with the other areas. Therefore, 61,577 households identified with this dummy have on average around 1 point more in the slum severity index than the areas of Guayaquil created between 1990 and 2010. But, the households that live in areas of Guayaquil created between 1990 and 2010 still have a higher slum severity index than the households living in the existing areas in 1990. Columns (3) and (4) include sequentially a list of individual controls. As we saw above, individual characteristics are an important explanatory source of the variance of the slum severity index. Once controlling for individual characteristics, the new built areas still present a significantly higher level of the slum severity index with respect to the existing areas in 1990. The overall effect for the new built areas in 2010 ($\theta_1 + \theta_2$ =2.891) is about 45% higher of what we find in the expanded area ten years earlier (θ_1 =1.987). At some point, these differentiated parameters allow for interpreting that it takes a long period (ten years) to decrease the severity of the slum condition by a third.

As for the controls, compared with what we studied above, we observe the same signs and significance of the parameters, although with different values, what can be interpreted in terms of differentiated patterns of sorting between and within cities, or at least within Guayaquil. What remains, though, is the additional explanatory power of the individual controls, which increase the coefficient of determination by some 0.15 in both exercises. Education level has the largest impact on the index, while the more children in the households have a lower effect on this index.

Finally, we also estimate the models for the two additional indexes capturing slums dimensions. The result of the lack of basic infrastructure coverage index (PC1) is presented in column (5), while the bad physical quality index (PC2) is presented in column (6). The results are similar to previous results, although signaling different speeds. After ten years of the new establishment, the dwellings present a significant relative improvement in the physical quality index (halved in such period)), stronger of what we find for the lack of basic infrastructure (just a third). This can be interpreted as a faster individual effort to improve the housing conditions than the public investment in developing basic infrastructure. This can be interpreted on an important role of the political economy channel in the existence of slums.

This evidence suggests that the new areas and expansion of the city of Guayaquil is based on slum creations. It is 122,514 households that live in these areas created in Guayaquil between 1990 and 2010 present on average higher slum indexes than the 529,554 households that live in existing areas of Guayaquil in 1990.

Table 4-5. Regression of the slum indexes on new areas of Guayaquil

VARIABLES	(1) ADD-OLS	(2) ADD-OLS	(3) ADD-OLS	(4) ADD-OLS	(5) PC1-OLS	(6) PC2-OLS
D1990_2010	2.733***	2.450***	2.244***	1.987***	2.113***	0.286***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.003)
D2001_2010		0.883***	0.850***	0.904***	1.081***	0.268***
		(0.010)	(0.010)	(0.010)	(0.010)	(0.005)
Female			-0.038***	-0.056***	-0.035***	-0.040***
			(0.004)	(0.004)	(0.004)	(0.003)
Age			-0.009***	-0.018***	-0.012***	-0.008***
			(0.0001)	(0.0001)	(0.0001)	(7.49e-05)
Children			0.220***	0.171***	0.080***	0.079***
			(0.002)	(0.001)	(0.001)	(0.001)
Migration			-0.232***	-0.226***	-0.102***	-0.139***
			(0.010)	(0.009)	(0.008)	(0.005)
Marital status			yes	yes	yes	yes

Continue on next page…

Education (Literacy or lower as base)						
Primary education				-0.317***	-0.227***	-0.156***
				(0.010)	(0.009)	(0.006)
Secondary education				-0.922***	-0.657***	-0.425***
				(0.010)	(0.009)	(0.006)
University or higher				-1.809***	-1.194***	-0.891***
				(0.010)	(0.009)	(0.006)
Constant	1.273***	1.243***	1.807***	3.128***	0.665***	0.997***
	(0.002)	(0.002)	(0.007)	(0.013)	(0.011)	(0.008)
Observations	652,068	652,068	652,068	638,204	638,204	638,204
R-squared	0.131	0.350	0.415	0.496	0.524	0.235

Standard errors clustered at household level in parentheses *** p<0.01, ** p<0.05, * p<0.1. Characteristics related to the head of the household. Differences in observations related to the missing values.

4.6 Conclusions

This chapter analyzes the relationship between city size and the minimum of quality of life in the cities of Ecuador. To carry out with this task, we develop a series of three indicators of slum characteristics: the slum severity index (ADD) based on adding dummy variables, the lack of basic infrastructure coverage index based on the first component from Principal Component Analysis (PC1), and the bad housing physical quality index based on the second component from the Principal Component Analysis (PC2). The variables for the analysis are obtained from the censuses of the population in Ecuador.

For all built indexes, a negative association is presented with respect to city size indicating that larger cities offer the better minimum standards of living. Socio-economic characteristics, such as education are associated with a better quality of life, while the more children in the households are associated with a worse quality of life measured trough the indexes.

We also focus on a particular case to analyze a rapid urbanization as a determinant of slums. We analyze Guayaquil, which is the largest city of Ecuador. We clearly find a positive association between the new areas of Guayaquil have a high slum characteristic. In fact, all the new areas of Guayaquil, after the year 1990, have on average higher level of slum characteristics. But, the most recent areas of Guayaquil created between 2001 and 2010 have on average higher slum indexes than the other areas. Thus, the minimum of quality of life in Guayaquil seems to improve slowly over the time.

References

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.

Athanassoglou, A. (2015). Multidimensional welfare rankings underweight imprecisions: a social choice perspective. Social Choice Welfare (44), 719 – 744.

Ballas, D. (2013). What makes a "happy city"?. Cities, 32, 539–550.

Bates, Robert H. (1981). Markets and States in Tropical Africa: The Political Basis of Agricultural Policies. Berkeley, CA: University of California Press.

Brueckner, J. (2013). Slums in developing countries: New evidence for Indonesia. Journal of Housing Economics, 22(4), 178-290

Cabrera-Barona, P., Murphy, T., Kienberger, S. & Blaschkle, T. (2015). A multi-criteria spatial deprivation index to support health inequality analyses. International Journal of Health Geographics, 14 (1), 1-14.

Camagni, R., Diappi, L. & Leonardi, G. (1989). Urban growth and decline in a hierarchical system: a supply-oriented dynamic approach. Regional Science and Urban Economics, 16, 145–160.

Castells-Quintana, D. (2016). Malthus living in a slum: Urban concentration, infrastructure and economic growth. Journal of Urban Economics, 98, 158-173.

Castells-Quintana, D. & Royuela, V. (2018). Spatially blind policies? Analyzing agglomeration economies and European Investment Bank funding in European neighboring countries. The Annals of Regional Science, 60(3), 569–589.

Census of India. (2001). Metadata and brief highlights on slum population. Available at http://www.censusindia.gov.in/2011-Common/CensusData2011.html

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

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. (2010). Estimating Agglomeration Effects with History, Geology, and Worker Fixed-Effects. In Glaeser, E. L. (ed.) Agglomeration Economics, (Chicago, IL: Chicago University), 15–65

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica, 80(6), 2543–2594.

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

Costas, D. & Hahn, M. (2000). Power couples: changes in the locational choice of the college educated: 1940-1990. Quarterly Journal of Economics, 115 (4), 1287-1315.

Decancq, K. & Lugo, M. A. (2013). Weights in Multidimensional Indices of Wellbeing: An overview. Econometric Reviews, 32(1), 7–34.

Díaz, J.P., Royuela, V. & Romaní, J. (2018) Migration Effects on Living Standards of the Left Behind. The Case of Overcrowding Levels in Ecuadorian Households, AQR working paper.

Diener, Ed. & Ryan, K. (2009). Subjective Well-Being: A General Overview. South African Journal of Psychology. 39 (4), 391-406.

Duque J.C., Royuela V. & Noreña M. (2012). A Stepwise Procedure to Determinate a Suitable Scale for the Spatial Delimitation of Urban Slums. In: Fernández Vázquez E., Rubiera Morollón F. (eds) Defining the Spatial Scale in Modern Regional Analysis. Advances in Spatial Science (The Regional Science Series). Springer, Berlin, Heidelberg, 237 – 254.

Durán, J. & Condorí, M. (2017). Deprivation Index for Small Areas Based on Census Data in Argentina. Social Indicators Research, 1-33.

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

Frankenhoff, C. A. (1967). Elements of an Economic Model for Slums in a Developing Economy. Economic Development and Cultural Change, Chicago 16, (1), 27.

Frick, S. & Rodríguez-Pose, A. (2017). Big or Small Cities? On city size and economic growth. Growth and Change, 49 (1), 4-32.

Fox, S. (2014). The Political Economy of Slums: Theory and Evidence from Sub-Saharan Africa. World Development, 54, 191-203.

Galiani, S., Gertler, P., Undurrage, R., Cooper,R., Martínes, S. & Ross, A. (2017). Shelter from the storm: Upgrading housing infrastructure in Latin American slums. Journal of Urban Economics (98), 187 - 213.

Glaeser, E. L. (2011). Triumph of the city: How our greatest invention makes us richer, smater, greener, healthier, and happier. Penguin Books. ISBN-13: 978-0143120544.

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., Kolko, J. & Saiz, A. (2001). Consumer city. Journal of Economic Geography, 1, 27-50.

Glaeser, E. L., Ponzetto, G. A. M. & Zou, Y. (2016). Urban networks: connecting markets, people, and ideas. Regional Science, 95(1), 17–59.

Harvard, S., Deguen, S. Bodin, J. Louis, K. & Laurent, O. (2008). A small-are index of socioeconomic deprivation to capture health inequalities in France. Social Science and Medicine, 67 (12), 2007-1016.

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

Holzer, M. Fry, J., Charbonneau, E. & Van Ryzin, G. (2009). Literature Review and Analysis Related to Optimal Municipal Size and Efficiency. Local Unit Alignment, Reorganization, and Consolidation Commission, SPAA papers, 44.

Lall, S. V., Selod, H. & Shalizi, Z. (2006). Rural–urban migration in developing countries: A survey of theoretical predictions and empirical findings. World Bank Policy Research Working Paper No. 3915. Washington, DC: The World Bank.

Lipton, M. (1977). Why Poor People Stay Poor: Urban Bias in World Development. The Scandinavian Journal of Economics, 80.

Liu, B.C. (1978). Variations in social quality of life indicators in medium metropolitan areas. American Journal of Economics and Sociology, 37(3), 241-260.

Mitra, A. & Nagar, J. (2018). City size, deprivation and other indicators of development: Evidence from India. World Development, 106 (2), 273–283.

Marx, B. Stoker, Stoker, T. & Suri, T. (2013). The economics of slums in the developing world. Journal of Economic Perspective, V. 27 (4), 187-210.

Narayan, D. (2000). Voices of the Poor: Can Anyone Hear Us?. World Bank Publication, DC.

Obaco, M. & Ballas, D. (2018). Deprivation Index for Small Areas Based on Census Data in Ecuador. Mimeo.

Obaco, M. & Díaz, J.P. (2018). Urbanization in Ecuador under the FUAs definition. Mimeo.

OECD. (2008). Handbook on Constructing Composite Indicators: Methodology and user guide. OECD Publishing: Paris, 3, 162.

OECD. (2012). Redefining "Urban". A new way to measure metropolitan areas. Paris: OECD publishing.

OECD. (2013a). Definition of Functional Urban Areas: for the OECD metropolitan database. Paris: OECD publishing.

OECD. (2013b). How's Life? 2013: Measuring Well-being. Paris: OECD publishing.

Patel, A. Koizumi, N. & Crooks, A. (2014). Measuring slum severity in Mumbai and Kolkata: A household-based approach. Habitat International, 41, 300-306.

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

Ravallion, M., Chen, S., & Sangraula, P. (2007). New evidence on the urbanization of global poverty. Population and Development Review, 33(4), 667–701.

Royuela, V. & Ordoñez, J. (2018). Internal Migration in a developing country: A panel data analysis of Ecuador (1982 – 2010). Papers in Regional Science, 97 (2), 345-367.

Royuela, V. & Suriñach, J. (2005). Constituents of Quality of Life and Urban size. Social Indicators Research, Springer, 74: 549-572.

Royuela, V., Suriñach, J. & Reyes, M. (2003). Measuring Quality of Life in small areas and over different periods of time, analysis of the province of Barcelona. Social Indicators Research, 64, 51–74

Saisana M. (2014). Addiction Severity Index. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht.

Turner, J. F. (1969). Uncontrolled urban settlement: problems and policies. Urbanization: development policies and planning, International Social Development Review, 1, United Nations, New York, 1968.

UN-Habitat. (2003). The challenge of slums. United Nations Human settlements program, Earthscan Publications on behalf of UN-Habitat.

UN-Habitat. (2015). Slum Almanac 2015 - 2016, UNON, Publishing Services Section, Nairobi,ISO 14001:2004-certified.

United Nations. (2015a). World Economic Situation and Prospects, WESP. United Nations Publication Sales No. E.14.II.C.2, ISBN: 978-92-1-109168-7: 198.

United Nations. (2015b). World Urbanization prospects: 2014 revision. Department of Economic and Social Affairs, Population Division, (ST/ESA/SER.A/366).

Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter?, Computers, Environment and Urban Systems, 56, 86-94.

Veneri, P. (2017). Urban spatial structure in OECD cities: Is urban population decentralising or clustering?. Papers in Regional Science, 1-27.

Villacís B., Carrillo D. (2011). Estadística Demográfica en el Ecuador: Diagnóstico y Propuesta. Instituto Nacional de Estadística y Censos (INEC). Quito – Ecuador, 86

World Bank. (2011). World Development Report 2011: Conflict Security and Development. The World Bank, Washington.

Appendix of Chapter Four

Table A4-1. Percentage of households living in slum conditions in 1990, disaggregated by conditions

	A		B			C	D		
City	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Ambato (1)	0.24	0.27	0.06	0.41	0.45	0.39	0.01	0.11	0.16
Babahoyo (2)	0.35	0.40	0.24	0.56	0.66	0.57	0.09	0.53	0.15
Chone (2)	0.32	0.35	0.31	0.51	0.60	0.46	0.10	0.51	0.14
Cuenca (1)	0.19	0.22	0.09	0.39	0.36	0.36	0.00	0.11	0.13
Daule (2)	0.54	0.58	0.29	0.84	0.86	0.69	0.15	0.60	0.19
Esmeraldas (2)	0.35	0.41	0.08	0.55	0.54	0.42	0.00	0.32	0.04
Fco. Orellana (3)	0.87	0.95	0.45	0.55	0.95	0.51	0.07	0.73	0.13
Guaranda (1)	0.13	0.16	0.15	0.47	0.45	0.41	0.02	0.02	0.28
Guayaquil (2)	0.35	0.36	0.03	0.44	0.46	0.45	0.00	0.25	0.04
Huaquilla (2)	0.23	0.28	0.11	0.91	0.92	0.56	0.00	0.15	0.18
La Libertad (2)	0.89	0.98	0.12	0.49	0.98	0.56	0.03	0.30	0.12
La Troncal (2)	0.66	0.86	0.20	0.56	0.88	0.57	0.02	0.26	0.10
Latacunga (1)	0.23	0.27	0.14	0.56	0.55	0.46	0.04	0.06	0.28
Loja (1)	0.06	0.10	0.06	0.26	0.31	0.43	0.00	0.07	0.10
Machala (2)	0.26	0.29	0.06	0.59	0.46	0.45	0.00	0.26	0.06
Manta (2)	0.26	0.30	0.08	0.51	0.57	0.46	0.01	0.18	0.06
Milagro (2)	0.15	0.20	0.08	0.58	0.70	0.50	0.04	0.32	0.10
Nueva Loja (3)	0.77	0.87	0.49	0.73	0.88	0.54	0.03	0.66	0.07
Otavalo (1)	0.17	0.20	0.18	0.44	0.47	0.47	0.02	0.04	0.28
Portoviejo (2)	0.12	0.13	0.07	0.34	0.41	0.42	0.04	0.34	0.13
Puyo (3)	0.12	0.15	0.06	0.27	0.37	0.36	0.00	0.37	0.01
Quevedo (2)	0.38	0.43	0.16	0.50	0.69	0.51	0.05	0.26	0.07
Quito (1)	0.15	0.20	0.04	0.21	0.27	0.33	0.00	0.02	0.06
Riobamba (1)	0.09	0.13	0.09	0.38	0.39	0.36	0.03	0.04	0.17
Santa Rosa (2)	0.07	0.08	0.08	0.56	0.35	0.54	0.00	0.19	0.07
Santo Domingo (2)	0.27	0.32	0.10	0.39	0.45	0.46	0.00	0.14	0.03
Tena (3)	0.24	0.29	0.26	0.42	0.57	0.46	0.03	0.50	0.06
Tulcán (1)	0.06	0.07	0.07	0.16	0.26	0.46	0.01	0.06	0.09
Total	0.25	0.29	0.07	0.39	0.43	0.42	0.01	0.16	0.09

Note: Average of the period of analysis.
(1) Andean cities, (2) Coastal cities and (3) Amazon cities.

Table A4-2. Percentage of households living in slum conditions in 2001, disaggregated by conditions.

	A		B			C	D		
City	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Ambato (1)	0.13	0.22	0.04	0.37	0.35	0.32	0.02	0.07	0.14
Babahoyo (2)	0.22	0.36	0.10	0.40	0.66	0.49	0.04	0.31	0.14
Chone (2)	0.27	0.37	0.21	0.37	0.61	0.41	0.07	0.44	0.23
Cuenca (1)	0.05	0.11	0.03	0.25	0.23	0.27	0.03	0.09	0.10
Daule (2)	0.50	0.50	0.07	0.68	0.88	0.60	0.04	0.33	0.22
Esmeraldas (2)	0.14	0.11	0.05	0.16	0.42	0.37	0.03	0.23	0.13
Fco. Orellana (3)	0.35	0.38	0.19	0.40	0.63	0.47	0.06	0.64	0.10
Guaranda (1)	0.16	0.40	0.16	0.62	0.59	0.42	0.04	0.08	0.38
Guayaquil (2)	0.20	0.20	0.03	0.17	0.49	0.41	0.03	0.16	0.13
Huaquilla (2)	0.20	0.53	0.09	0.32	0.75	0.45	0.03	0.15	0.26
La Libertad (2)	0.30	0.31	0.09	0.15	0.88	0.51	0.04	0.22	0.23
La Troncal (2)	0.33	0.76	0.05	0.17	0.51	0.47	0.03	0.13	0.08
Latacunga (1)	0.08	0.22	0.07	0.53	0.41	0.36	0.02	0.03	0.19
Loja (1)	0.04	0.16	0.05	0.19	0.21	0.34	0.03	0.09	0.09
Machala (2)	0.11	0.14	0.03	0.17	0.30	0.38	0.04	0.17	0.12
Manta (2)	0.29	0.31	0.06	0.22	0.54	0.41	0.04	0.18	0.13
Milagro (2)	0.18	0.32	0.06	0.39	0.78	0.42	0.03	0.19	0.15
Nueva Loja (3)	0.56	0.76	0.14	0.33	0.55	0.46	0.04	0.37	0.08
Otavalo (1)	0.06	0.13	0.07	0.31	0.30	0.39	0.03	0.04	0.21
Portoviejo (2)	0.21	0.19	0.04	0.28	0.44	0.35	0.04	0.31	0.21
Puyo (3)	0.13	0.17	0.05	0.17	0.28	0.31	0.03	0.37	0.05
Quevedo (2)	0.26	0.38	0.11	0.29	0.77	0.46	0.05	0.18	0.13
Quito (1)	0.04	0.08	0.02	0.09	0.15	0.26	0.02	0.03	0.06
Riobamba (1)	0.05	0.12	0.04	0.29	0.24	0.29	0.02	0.03	0.12
Santa Rosa (2)	0.06	0.09	0.04	0.26	0.32	0.41	0.04	0.15	0.14
Santo Domingo (2)	0.37	0.47	0.06	0.23	0.34	0.41	0.03	0.15	0.09
Tena (3)	0.13	0.23	0.12	0.30	0.42	0.44	0.03	0.42	0.08
Tulcán (1)	0.03	0.04	0.03	0.10	0.09	0.35	0.03	0.05	0.08
Total	0.14	0.19	0.04	0.20	0.37	0.36	0.03	0.12	0.12

Note: Average of the period of analysis.
(1) Andean cities, (2) Coastal cities and (3) Amazon cities.

Table A4-3. Percentage of households living in slum conditions in 2010, disaggregated by conditions.

	A		B			C	D		
City	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Ambato (1)	0.05	0.16	0.02	0.17	0.26	0.19	0.00	0.05	0.35
Babahoyo (2)	0.16	0.30	0.08	0.26	0.64	0.39	0.00	0.28	0.43
Chone (2)	0.26	0.44	0.13	0.24	0.60	0.33	0.05	0.33	0.35
Cuenca (1)	0.03	0.08	0.01	0.09	0.19	0.16	0.00	0.15	0.18
Daule (2)	0.45	0.47	0.07	0.38	0.73	0.40	0.01	0.28	0.38
Esmeraldas (2)	0.06	0.06	0.05	0.04	0.27	0.28	0.00	0.20	0.40
Fco. Orellana (3)	0.18	0.23	0.11	0.10	0.60	0.34	0.01	0.53	0.24
Guaranda (1)	0.13	0.32	0.09	0.54	0.57	0.31	0.01	0.39	0.48
Guayaquil (2)	0.15	0.16	0.07	0.08	0.40	0.35	0.00	0.13	0.47
Huaquilla (2)	0.05	0.16	0.07	0.10	0.61	0.37	0.00	0.15	0.80
La Libertad (2)	0.15	0.15	0.10	0.04	0.58	0.44	0.02	0.23	0.69
La Troncal (2)	0.09	0.23	0.04	0.08	0.44	0.36	0.01	0.10	0.63
Latacunga (1)	0.05	0.24	0.04	0.28	0.34	0.23	0.01	0.07	0.46
Loja (1)	0.02	0.10	0.01	0.09	0.15	0.22	0.00	0.16	0.22
Machala (2)	0.08	0.17	0.02	0.09	0.24	0.32	0.00	0.11	0.52
Manta (2)	0.27	0.32	0.06	0.07	0.47	0.33	0.01	0.13	0.59
Milagro (2)	0.07	0.30	0.05	0.21	0.85	0.36	0.01	0.18	0.58
Nueva Loja (3)	0.35	0.61	0.05	0.13	0.39	0.34	0.00	0.25	0.50
Otavalo (1)	0.02	0.14	0.02	0.11	0.23	0.25	0.00	0.20	0.51
Portoviejo (2)	0.23	0.26	0.06	0.12	0.37	0.26	0.01	0.19	0.45
Puyo (3)	0.07	0.11	0.02	0.07	0.25	0.21	0.00	0.29	0.19
Quevedo (2)	0.16	0.29	0.05	0.09	0.80	0.39	0.01	0.15	0.66
Quito (1)	0.01	0.04	0.01	0.03	0.08	0.16	0.00	0.06	0.25
Riobamba (1)	0.04	0.18	0.03	0.22	0.23	0.17	0.00	0.05	0.36
Santa Rosa (2)	0.03	0.05	0.03	0.12	0.22	0.33	0.00	0.13	0.62
Santo Domingo (2)	0.30	0.46	0.03	0.09	0.30	0.29	0.00	0.09	0.61
Tena (3)	0.08	0.22	0.05	0.13	0.35	0.31	0.01	0.26	0.43
Tulcán (1)	0.01	0.02	0.01	0.06	0.06	0.23	0.00	0.12	0.29
Total	0.10	0.15	0.04	0.09	0.30	0.27	0.00	0.12	0.40

Note: Average of the period of analysis.
(1) Andean cities, (2) Coastal cities and (3) Amazon cities.

Table A4-4. Results from the PCA

Results from the censuses (obs.5,269,680)

Component	**Eigenvalue**	**Proportion**	**Cumulative**
Comp1	3.08	1.85	0.34
Comp2	1.23	0.29	0.48
Comp3	0.94	0.06	0.58
Comp4	0.88	0.03	0.68
Comp5	0.85	0.15	0.78
Comp6	0.70	0.06	0.85
Comp7	0.63	0.17	0.92
Comp8	0.47	0.23	0.97
Comp9	0.23	.	1.00

After rotation:

Variable	**PC1**	**PC2**	**Unexplained**
No piper water	0.45	-0.23	0.30
No public water supply	0.46	-0.25	0.28
No sewerage system	0.42	-0.11	0.43
No energy supply	0.26	0.11	0.78
No garbage collection	0.38	-0.17	0.51
Overcrowding	0.23	0.11	0.82
Non-durable roof materials	0.12	0.69	0.37
Non-durable wall materials	0.32	0.37	0.52
Non-durable floor materials	0.17	0.45	0.66

PC1=lack of basic infrastructure coverage index

PC2=bad of housing physical material index

Table A4-5. Average of the indexes by city

City	ADD		PC1		PC2		
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Obs.**
Ambato (1)	1.56	1.82	0.03	1.68	-0.13	1.03	195856
Babahoyo (2)	2.76	2.15	0.99	2.01	0.50	1.63	66038
Chone (2)	2.88	2.48	1.12	2.26	0.79	2.03	42401
Cuenca (1)	1.19	1.69	-0.31	1.50	-0.13	1.13	291042
Daule (2)	3.56	2.33	1.87	2.19	0.52	1.79	45420
Esmeraldas (2)	1.74	1.73	0.09	1.65	0.15	1.17	94826
Fco. Orellana (3)	2.96	2.23	1.25	2.22	0.74	1.47	19523
Guaranda (1)	2.64	2.23	0.87	1.93	0.37	1.33	33103
Guayaquil (2)	1.85	1.96	0.23	1.88	0.06	1.18	1452275
Huaquilla (2)	2.59	1.70	0.72	1.56	0.33	1.29	25768
La Libertad (2)	2.81	1.87	1.02	1.86	0.39	1.49	99328
La Troncal (2)	2.61	1.81	1.02	1.81	0.09	1.29	23427
Latacunga (1)	1.96	1.97	0.31	1.71	0.03	1.20	85611
Loja (1)	1.12	1.52	-0.43	1.28	-0.12	1.04	99991
Machala (2)	1.66	1.73	-0.01	1.60	0.09	1.20	173260
Manta (2)	2.17	1.93	0.52	1.81	0.14	1.29	150962
Milagro (2)	2.51	1.73	0.74	1.53	0.28	1.40	88550
Nueva Loja (3)	3.29	2.12	1.78	2.16	0.38	1.37	32160
Otavalo (1)	1.65	1.81	-0.10	1.51	0.16	1.21	194508
Portoviejo (2)	1.93	1.95	0.23	1.72	0.30	1.48	120948
Puyo (3)	1.40	1.59	-0.17	1.50	0.13	1.14	19086
Quevedo (2)	2.64	1.85	0.86	1.76	0.35	1.54	104260
Quito (1)	0.80	1.27	-0.67	1.13	-0.26	0.93	1436225
Riobamba (1)	1.33	1.69	-0.23	1.44	-0.12	1.10	135636
Santa Rosa (2)	1.58	1.54	-0.21	1.34	0.18	1.22	31833
Santo Domingo (2)	2.13	1.87	0.58	1.85	0.01	1.14	155141
Tena (3)	2.09	2.06	0.33	1.91	0.44	1.35	14970
Tulcán (1)	0.89	1.28	-0.71	1.00	-0.10	1.08	37532
Total	1.59	1.85	0.00	1.70	0.00	1.19	5269680

Note: Average of the period of analysis.
(1) Andean cities, (2) Coastal cities and (3) Amazon cities.

Figure A4-1. Urban primacy in Ecuador

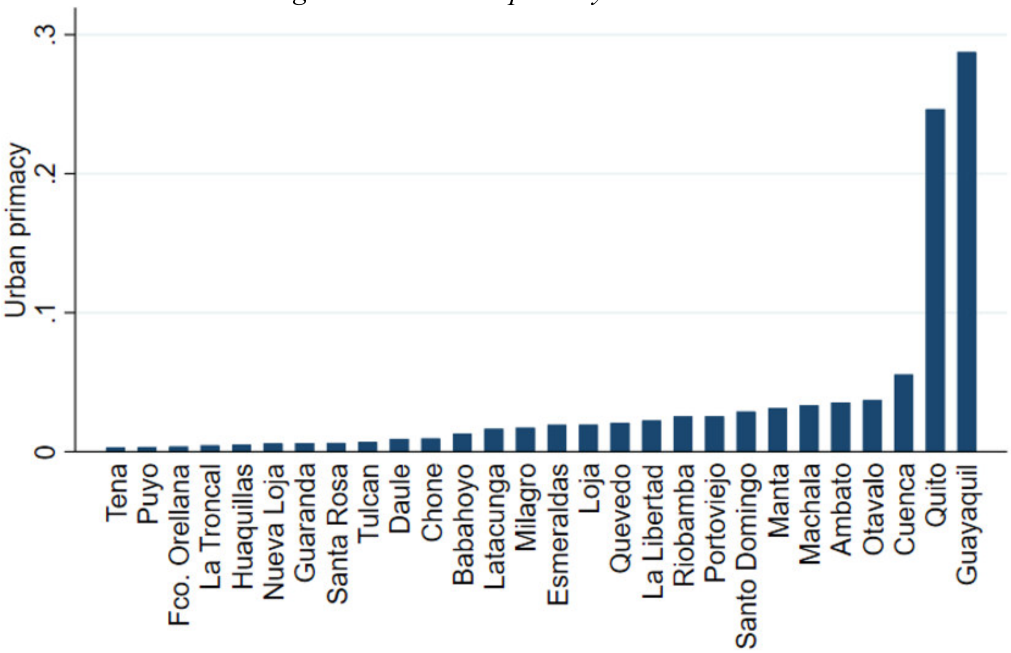


Figure A4-2. Relative frequencies of the slum severity index (ADD).

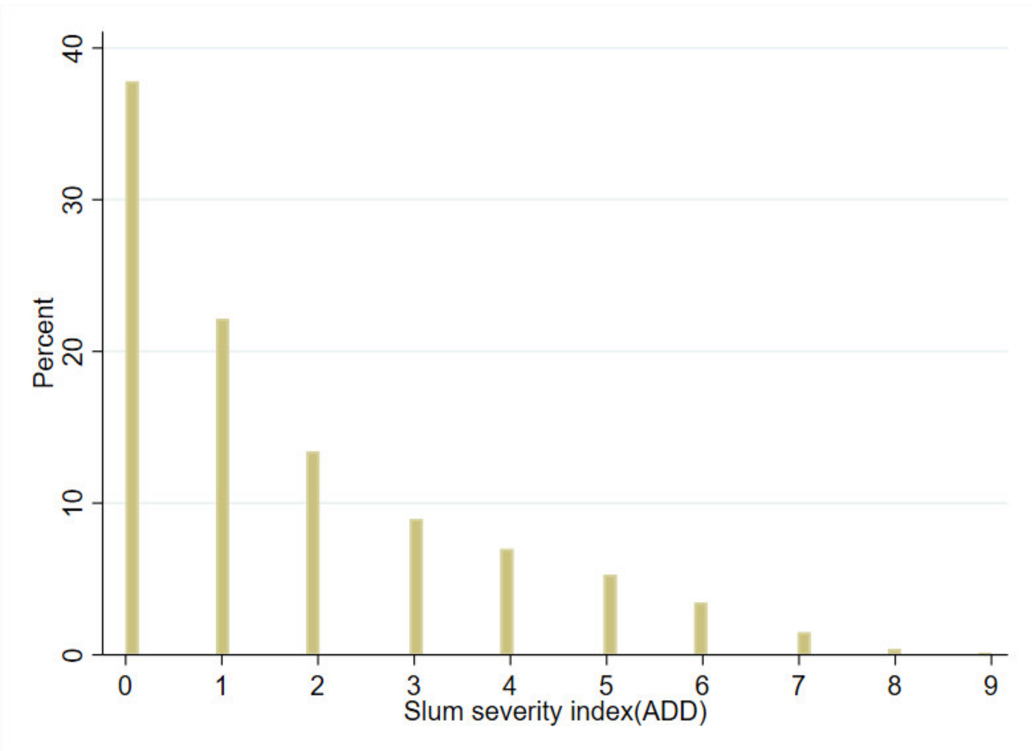


Figure A4-3. New extension in Guayaquil

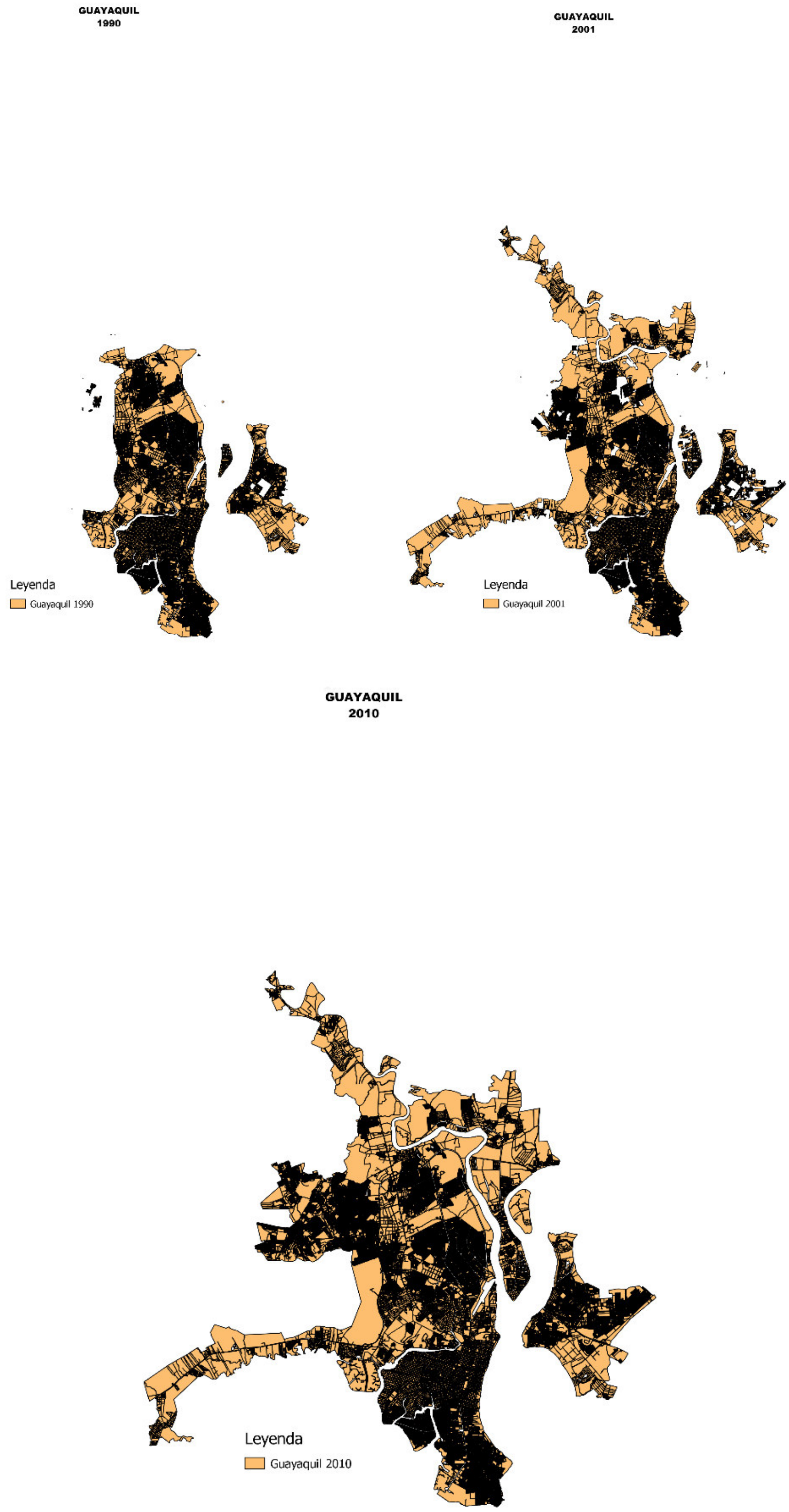
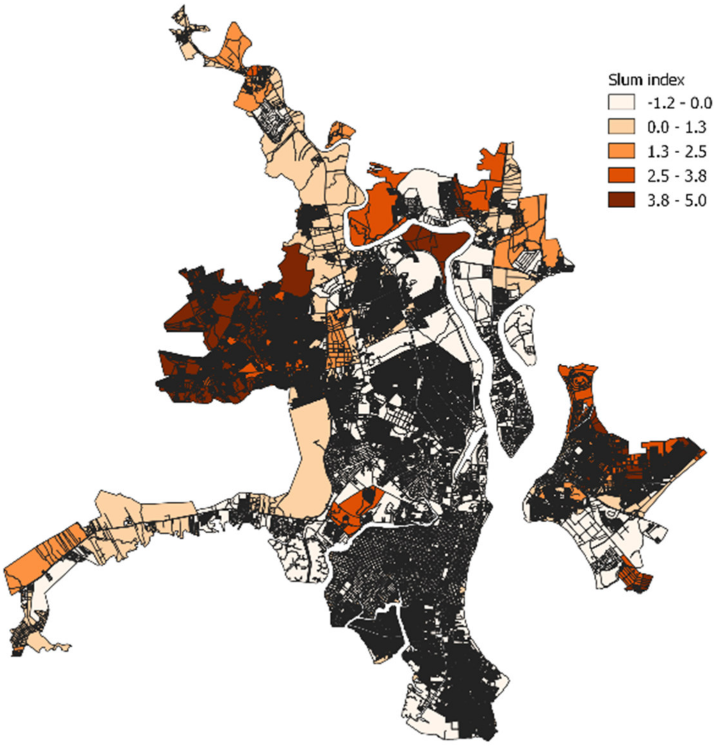


Figure A4-4. Distribution of the slum severity index of Guayaquil in 2010.
PC1: Lack of basic infrastructure coverage index



PC2: Bad housing physical quality index is shown on the right side.

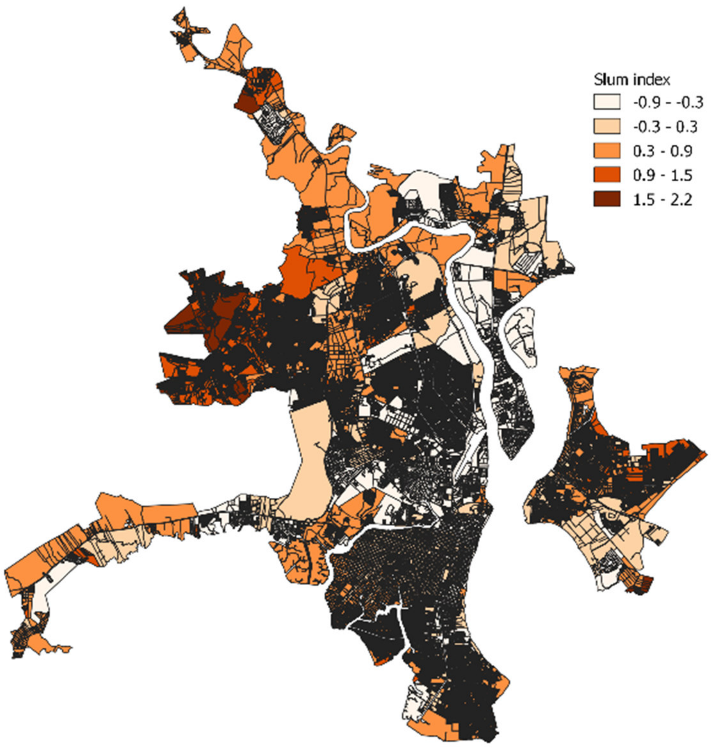
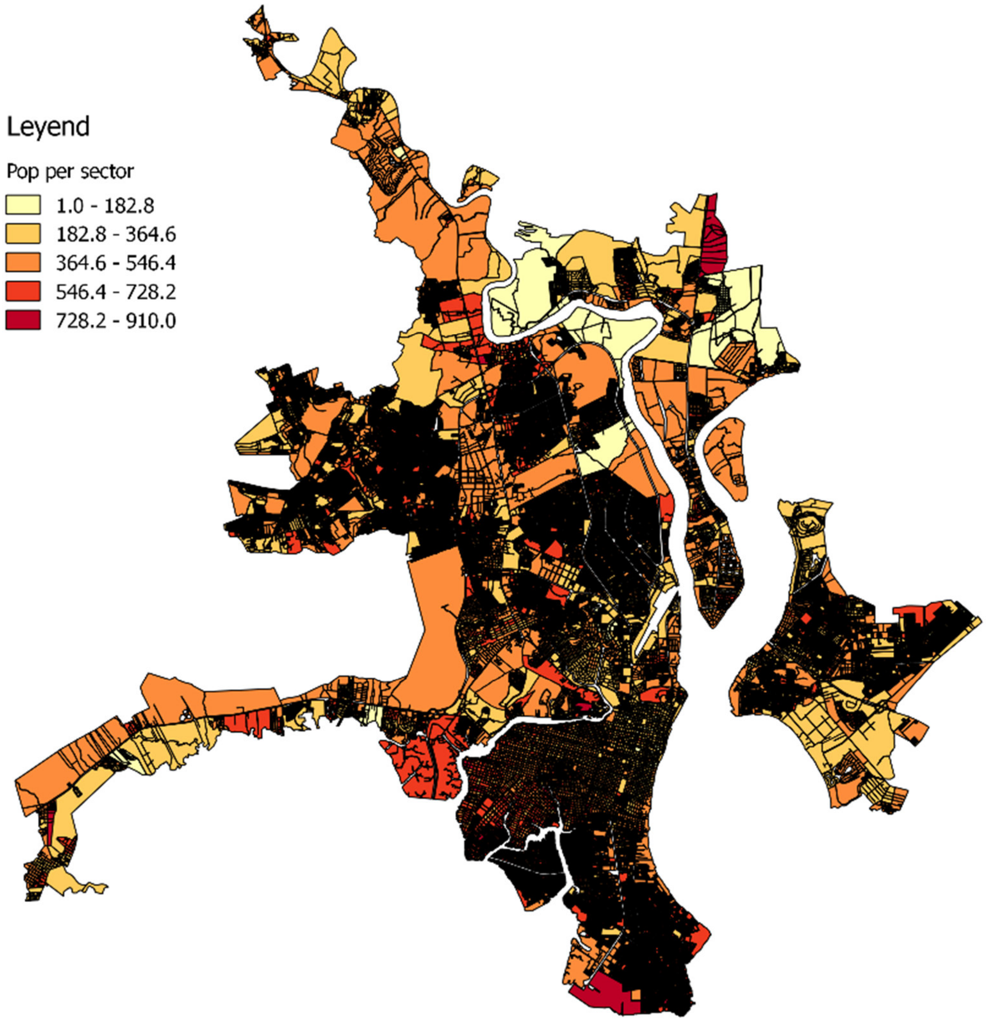


Figure A4-5. Spatial distribution of the population of Guayaquil in 2010



Chapter Five: Conclusions, Policy Implications and Further Research

5.1 Summary and main findings

This thesis focuses on studying agglomeration economies in Ecuador. Chapter one introduces the international literature on agglomeration economies and explains the motivation for analyzing Ecuador as case study. The literature of agglomeration economies has focused on analyzing big emerging markets such as Brazil, India and China. Thus, Ecuador presents a set of unexplored characteristics in the literature such as small in geographical extension and population size (low population density as well), with industrialization and urbanization rates below the Latin America's averages. It has also shown a rapid urbanization since 1960. Finally, it has an important share of informal workers in cities and high incidence of slum. Although Ecuador does not provide national statistics, UN-Habitat estimates that in 2014 around 35% of the population lived in slums. Chapter one also presents the research goals of this thesis, which are analyzing the effects of agglomeration on labor productivity and the quality of life enjoyed by urbanites. In order to achieve these aims, though, first it is needed to identify the units of analysis, as Ecuador does not have an economic definition of cities. Consequently, an instrumental but key objective of this dissertation is the definition of real cities, the spatially economic approach of urban areas. The identification of cities is important *per se*, but it also helps to reduce the bias that can arise by using administrative boundaries of cities as unit of analysis.

We devote chapter two for identifying the economic boundaries of the cities in Ecuador. We identify the so-called Functional Urban Areas (FUAs) in Ecuador. We apply the concept of FUAs used by the OECD in order to use a standardized economic concept of cities across countries. The traditional FUAs methodology is based on the identification of urban cores, by means of a minimum of population density and population size, and the use of commuting flows to connect urban cores and surrounded areas to define the final boundaries of the FUAs. As many other developing economies, Ecuador has not the needed data to identify FUAs as in the traditional approach. In this thesis we propose a new approach based on a population density and a varying travel time model to connect urban cores and define the hinterlands of the FUAs. This approach only uses satellite data through the usage of Landscan, Open Street Maps and Google maps. The first one gives information of population density while the other two give information of the road network structure of the country. The final extension of the FUAs is based on the geographical extension of the urban cores. This effect seems more realistic as larger urban areas has on average larger influence zones than small urban areas. This model solves the problem of the lack of standard data to identify FUAs because it only uses satellite data. However, a varying travel time model needs some preliminary calibration. The parameters for this model are obtained from considering the case of Colombia, a country for which there is commuting data, and has similar geographical characteristics, and a road network structure like Ecuador. However, more data are needed to generalize this approach. As main findings of chapter two, we identify in Ecuador 34 urban cores based on a minimum of population of 500 inhabitants per squared kilometer and 25,000 inhabitants in urban cores. If we impose the minimum used by the OECD –1,000 inhabitants per squared kilometer and 50,000 inhabitants in urban cores–, we are only able to identify 20 urban cores, as we are not able to identify urban areas in low density areas, such as the Amazonian region, where there are important urban settlements that are representative of the Amazonian region. Consequently, we believe that using lower thresholds are important in countries with low density areas. A series of robustness checks and sensitivity analysis are performed to validate our approach. Our approach presents results similar to those obtained when the FUAs are estimated using a survey of commuting, a gravity model, a radiation model and migration patterns. The sensitivity test shows that the proposed model is more stable when we use higher thresholds to define urban cores. But, higher thresholds in the low densely country can also hide important urban settlements that are growing up and are representative of many regions. Our preferred number of FUAs is 28 that are identified using the half of the minimum thresholds applied by the OECD.

In chapter three, we use individual wages of workers as an indicator of labor productivity to analyze agglomeration effects in Ecuador. We use the 28 FUAs identified in chapter two as spatial units of analysis. Labor micro-surveys (ENEMDU Surveys) offer micro level information of workers for the period 2005 – 2015. We follow the standard methodology used in the literature of agglomeration economies, and we perform pool OLS and IV estimates to study agglomeration effects. We use as instruments the historical population and soil pH of the FUAs. Although, we cannot use a panel of workers as it is usually recommended in the current literature of agglomeration effects, we use a large set of individual workers' characteristics such as age, gender, ethnicity, education level, occupation and industry classification among others to control for individual heterogeneity as it has been done on the estimation agglomeration effects in other developing economies. Our variables of interest are the two most studied in the literature of agglomeration effects: urbanization externalities and local industrial specialization. We use total market size represented in the total FUA population as indicator of urbanization externalities and a local industrial specialization index that uses the share of local labor force by industry in each city compared with the total industry share of the country as indicator of specialization. Results indicate that agglomeration effects exist and exert a positive and significant impact on labor productivity in the Ecuadorean cities. In detail, we obtain an elasticity of wages of 3.8% for urbanization externalities and 0.9% for the local specialization index. Moreover, higher educated people benefit more from the agglomeration effects. We also find evidence of congestion in wages in the two largest cities because the agglomeration effects are higher when the two largest cities are excluded from the regression. We also study the agglomeration effects on the presence of the dual labor market by dividing the labor force between formal and informal workers. We consider informal workers as workers employed in firm with less than 100 workers with no tax identification number (RUC-*Registro único de contribuyentes*). Results show that, as for urbanization externalities, informal workers are penalized with respect to formal workers. Nonetheless, informal workers benefit from these agglomeration effects. As for specialization, informal workers do not obtain benefits. We also try to obtain information of the channels through which agglomeration effects work for formal and informal workers in cities. Although the surveys are not precisely designed to follow workers over time, we built a panel subsample of the data by exploiting some characteristics of the structure of the survey. Findings show that the advantages of agglomeration for formal workers might be explained by positive sorting and better gains from job changes, while for informal workers gains arise from positive learning externalities.

In chapter four, we study the association between city size and the quality of life offered by the cities in Ecuador. In particular we lack well-sign by focusing on the incidence of slums. We define slums as informal settlements in cities, formed by households living in the dwelling, which is lacking one or more of the following conditions: access to improved water, access to improved sanitation, sufficient living area, and durability of housing. Following the definition of slums, we elaborate a proxy to measure slums characteristics of the households in the FUAs using several variables gathered from the census of the population in Ecuador during the period 1990 – 2010. We define three indexes to measure the slum conditions of the households; the slum severity index measured based on adding dummy variables (ADD) and the Principal Component Analysis where two indexes are obtained: the first component of the Principal Component analysis (PC1) is labelled as the lack of basic infrastructure coverage, while the second principal component analysis of the Principal Component Analysis (PC2) that captures bad housing physical quality index. All the built indexes have to be interpreted in an inverted way: the higher the index is, the worse the living conditions are. We find a significant and negative association with city size, meaning that larger cities offer better quality of life in Ecuador. Moreover, covariates such low education level, and more children in the households are associated with higher levels of slum severity index in cities. Education is the most important variable that affect the indexes. Besides, the role of the two largest cities is also considered in this chapter. After excluding the two largest cities, we find melt evidence of congestion as the two largest cities contribute more to the slum severity index and the bad housing physical quality index. But, on the other hand, we also find evidence that the two largest cites contribute more to covering the basic infrastructure of the households. Finally, we also analyze rapid urban growth as a determinant of the creation of slums, focusing on the particular case of Guayaquil, the largest city of Ecuador. In Guayaquil, fast urbanization seems to be associated with slums creation because the new extension of the city after 1990 are associated with a higher level of slum characteristics. In 2010, the extensions after 2001 present the highest level of slum severity index. We find small improvements in the slum severity index for these areas over the decades. Moreover, Guayaquil presents a higher level of the slum severity index compared with other cities with high urban primacy in Ecuador.

5.2 Policy implications

Relevant policy implications can be derived from this thesis. Chapter one emphasizes the importance of specific studies on urban issues in the developing world and stresses the fact that the lack of adequate data has not to shy away researchers from conducting relevant studies. On the contrary, we have learned that there is a list of opportunities for academics to expand current knowledge, which is food for thought for policy makers.

Chapter two identifies the FUAs in Ecuador. This chapter has important implications for policy makers and urban planners, because the FUAs are the best delimitation of the economic definition of cities. For instance, this chapter helps to trace the urbanization pace that is taking place in Ecuador once we know the boundaries of these urban areas. In detail, we have learned that Ecuadorean FUAs are mostly small in size, although they are steadily growing. We consider that the identification of such secondary smaller cities would deserve important attention in the future in order to reduce dependence on urban primacy and to develop policies aiming at exploiting agglomeration benefits in such smaller cities rather than ameliorating congestion costs in larger FUAs.

We also show that administrative boundaries are very heterogeneous, being very large in the Amazon and Coastal region, while they are very small and fragmented in the Andean region. Furthermore, the size of the urban cores is really small compared with the extension of the administrative boundaries. Thus, maintaining the current administrative boundaries in Ecuador has not to be an excuse to impulse local policies considering the actual dimension of the spatial economic units. As it has been found in the literature, political fragmentation of economic units, such as metropolitan areas, is a source of inefficiencies that can be avoided. The design and implementation of policy actions at the FUAs level would be useful in order to find the adequate benefit not only for the urban core but also for the spatial units that are surrounding it. A clear recommendation derived from chapter two is the need of commuting data to have a definitive definition of FUAs based in actual commuting flows.

Chapter three presents interesting findings on the type of urbanization that is taking place in developing countries. First, and in line with the economic literature, we can say that larger cities are presenting benefits for workers. However, we have found space for improvement in a list of aspects. There is a clear need of increasing the overall education level of the labor force, which is mostly composed of low educated workers. This characteristic is linked to the duality in the labor force in terms of formal versus informal economy. We have seen that larger cities seem to offer a solution for informal workers as they can find work and better wages. However, this is not an optimal solution as the agglomeration economies tend to decline at some point especially for informal workers, as we found evidence of congestion in wages in larger cities. The growth of larger cities based on low educated workers who finally end in the informal economy is a strategy that is not fully exploiting the advantages of agglomeration economies. We see a need of continuing the development of educational policies over the country and also to define spatially tailored growth strategies also for areas with smaller cities.

In chapter four, we have learned on the relationship between quality of life and city size in Ecuador. We have seen that there is a lack of official data on the rate of slums in Ecuador following the UN Habitat definition. Nevertheless, there are sufficient statistics to follow the evolution and distribution of most slums characteristics. Our results suggest that the proportion of households with slums characteristics in Ecuador is higher than the slum rate estimated by the UN-Habitat. More than half of Ecuadorean households in urban areas are facing some slum conditions. It is an important call for the attention of the government, as it a key determinant of individual well-being, what should turn into a political priority. We have focused our attention to the particular case of Guayaquil, where a rapid urban growth in the last two decades has been based on a spatial expansion in areas that are considered nowadays as slums. We have verified that these new areas are formed by households living in dwellings strongly characterized as slums, what we link with rapid urbanization. We have seen that the response of individuals to improve their housing material conditions is faster than the response of public authorities in terms of providing the adequate infrastructures. This has important implications for the reduction of poverty and inequality within cities, which is becoming the major driver of spatial inequality in countries.

5.3 Future research

In this PhD thesis, we have analyzed two aspects of agglomeration, by studying the relationship of wages and slums with respect to city size. However, there are still relevant aspects to consider about the effects of urbanization in the developing world. The costs of agglomeration such as housing prices, pollution, or crime are also very relevant aspects in this literature. Moreover, the study of cluster of industries is also relevant to assess the effects of agglomeration in the developing world. But, the limitation of data has always been a huge barrier. In that case, we propose as future research the comparison of agglomeration effects using a larger sample of cities in Latin American countries. We also propose to deeply explore the role of the informal economy in Ecuador and Latin America. It is also needed to study the determinants of being in the informal economy and to find adequate instruments to study the relationship with city size. Besides, the joint analysis of wages and quality of life is also a relevant topic to study. However, it introduces the necessity of longitudinal data. Finally, the evolution of urbanization using larger periods of analysis is also suggested as it would explain the current performance of the cities.

We aim that this thesis is a first step in a research agenda. As an example, we are currently working on developing a thorough exploration of the determinants of the slums in a list of Ecuadorean cities. We are mapping deprivation zones at small spatial units in the whole country, zones, which are composed of some one hundred households in order to have not only individual information but also a further dimension of slums: groups of houses with bad conditions. This study is expected to be completed for the 2010 census, as it allows to derive this spatial detail. As this would imply the study of a cross section, we are also developing a parallel strategy of studying housing deprivation by means of survey data, which allows to build a panel data set at the province level.

Finally, we plan to study another dimension of the costs associated to city size, such as crime. We have gathered a large data set on geo-localized emergency-calls of robberies. We expect to conduct an exploration of the spatial structure of crime in the cities of Ecuador. These data are rich in the sense of being geo-located in longitude and latitude. Thus, we want to go further, evaluating crime within cities. However, there is a lack of data at the individual level to contrast the determinants of crime at such low spatial level. Future research will focus on giving evidence of the agglomeration effect on crime rates.

References (Full list)

Adams, J. S., VanDrasek, B. J. & Phillips, E. G. (1999). Metropolitan Area Definition in the United States. Urban Geography, 20(8), 695–726.

Ahlfeldt, G. M. & Wendland, N. (2016). The spatial decay in commuting probabilities: Employment potential vs. commuting gravity. Economics Letters, 143, 125–129.

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.

Alonso, W. (1964). Location and Land Use: Towards a General Theory of Land Rent. Cambridge, MA: Harvard University Press.

Arsanjani, J. J., Helbich, M. & Mousivand, A. J. (2014). A Morphological Approach to Predicting Urban Expansion. Transactions in GIS, 18(2), 219–233.

Athanassoglou, A. (2015). Multidimensional welfare rankings underweight imprecisions: a social choice perspective. Social Choice Welfare (44), 719 – 744.

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

Ballas, D. (2013). What makes a “happy city”?. Cities, 32, 539–550.

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

Bates, Robert H. (1981). Markets and States in Tropical Africa: The Political Basis of Agricultural Policies. Berkeley, CA: University of California Press.

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.

Bretagnolle, A., Giraud, T., Guerois, M. & Mathian, H. (2010). Naming UMZ: methods and results. Technical Report, ESPON 2013 Database. ESPON Database Portal, technical report.

Brueckner, J. (2013). Slums in developing countries: New evidence for Indonesia. Journal of Housing Economics, 22(4), 178-290.

Cabrera-Barona, P., Murphy, T., Kienberger, S. & Blaschkle, T. (2015). A multi-criteria spatial deprivation index to support health inequality analyses. International Journal of Health Geographics, 14 (1), 1-14.

Camagni, R., Capello, R. & Caragliu, A. (2015). Static vs. Dynamic agglomeration economies. Spatial context and structural evolution behind urban growth. Papers in Regional Science, 95 (1), 133-159.

Camagni, R., Diappi, L. & Leonardi, G. (1989). Urban growth and decline in a hierarchical system: a supply-oriented dynamic approach. Regional Science and Urban Economics, 16, 145–160.

Carlsen, F., Rattsø, J. & Stokke, H. E. (2016). Education, experience, and urban wage premium. Journal of Regional Science and Urban Economics, 60 (1), 30-49.

Casado-Díaz, J. M., & Coombes, M. (2011). The delineation of 21st century local labour market areas: a critical review and a research agenda. Boletín de la Asociación de Geógrafos Españoles, 57, 7-32.

Castells-Quintana, D. (2016). Malthus living in a slum: Urban concentration, infrastructure and economic growth. Journal of Urban Economics, 98, 158-173.

Castells-Quintana, D. & Royuela, V. (2018). Spatially blind policies? Analysing agglomeration economies and European Investment Bank funding in European neighbouring countries. The Annals of Regional Science, 60(3), 569–589.

Census of India. (2001). Metadata and brief highlights on slum population. Available at http://www.censusindia.gov.in/2011-Common/CensusData2011.html

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.

Chen, M., Zhang, H., Liu, W. & Zhanf, W. (2014). The global pattern of urbanization and economic growth: Evidence from the last three decades. Plos One, Open access (9).

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.

Cohen, J. E., Roig, M., Reuman, D. C. & GoGwilt, C. (2008). International migration beyond gravity: a statistical model for use in population projections. Proceedings of the National Academy of Sciences of the United States of America, 105(40), 15269–15274.

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

Combes, P. P., Démuger, S. & Li, S. (2015). Migration externalities in Chinese cities. European Economic Review, 76, 152 – 167.

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. (2011). The identification of agglomeration economies. Journal of Urban Economics, 11 (2), 253 -266.

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

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica, 80(6), 2543–2594.

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

Coombes, M. (2004). Multiple Dimensions of Settlement Systems: Coping with Complexity. In the New forms of urbanization: beyond the urban-rural, Chapter 16.

Coombes, M.G., Green, A.E. & Openshaw, S. (1986). An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain. Journal of the Operational Research Society, 37, 943-953.

Cörvers, F., Hensen, M. & Bongaerts, D. (2009). Delimitation and Coherence of Functional and Administrative Regions. Regional Studies, 43(1), 19–31.

Costas, D. & Hahn, M. (2000). Power couples: changes in the locational choice of the college educated: 1940-1990. Quarterly Journal of Economics, 115 (4), 1287-1315.

D'Costa, S., and Overman, H. G. (2014). The urban wage growth premium: Sorting or learning?, Journal of Regional Science and Urban Economics, 48 (1), 168–179.

Davoudi, S. (2008). Conceptions of the City-region: A Critical Review. Urban Design and Planning, 161, 51–60.

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

Decancq, K. & Lugo, M. A. (2013). Weights in Multidimensional Indices of Wellbeing: An overview. Econometric Reviews, 32(1), 7–34.

Demographia. (2015). Demographia World Urban Areas & Population Projections. Demographia, (January), p. 132.

Departamento Nacional de Planeación (DNP). (2012). Algunos aspectos del análisis del sistema de ciudades colombiano, MISIÓN del Sistema de Ciudades, Bogotá, Colombia.

Díaz, J.P., Royuela, V. & Romaní, J. (2018). Migration Effects on Living Standards of the Left Behind. The Case of Overcrowding Levels in Ecuadorian Households, AQR working paper.

Diener, Ed. & Ryan, K. (2009). Subjective Well-Being: A General Overview. South African Journal of Psychology. 39 (4), 391-406.

Djikstra, L. & Poelman, H. (2012). Cities in Europe, the new OECD-EC definition. Regional focus 01/2012, European Commission.

Duque J.C., Royuela V. & Noreña M. (2012). A Stepwise Procedure to Determinate a Suitable Scale for the Spatial Delimitation of Urban Slums. In: Fernández Vázquez E., Rubiera Morollón F. (eds) Defining the Spatial Scale in Modern Regional Analysis. Advances in Spatial Science (The Regional Science Series). Springer, Berlin, Heidelberg, 237 – 254.

Durán, J. & Condorí, M. (2017). Deprivation index for Small Areas Based on Census Data in Argentina. Social Indicators Research, 1-33.

Duranton, G. (2015). A proposal to delineate metropolitan areas in Colombia. Desarrollo y Sociedad, (75), 223‑264.

Duque, J. C., Ramos, R. & Suriñach, J. (2007). Supervised Regionalization Methods: A Survey. International Regional Science Review, 30(3), 165-220.
Durán, J. & Condorí, M. (2017). Deprivation Index for Small Areas Based on Census Data in Argentina. Social Indicators Research, 1-33.
Duranton, G. (2009). Are Cities Engines of Growth and Prosperity for Developing Countries. In M. Spence, P. Annez, and R. Buckley, (ed) Urbanization and Growth, the Commission Growth and Development, Washington, D.C.
Duranton, G. (2015). Growing through cities in developing countries. The World Bank Research Observer, Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development, 30 (1).
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 Handbook of Regional and Urban Economics, 4 ed. J.V. Henderson and J.-F. Thisse (Amsterdam: Elsevier-North-Holland), 2063–2117.
ESPON. (2005). Potentials for polycentric development in Europe. Project report. ISBN: 91-89332-37-7.
Eurostat. (1999). Regio Database, User's Guide, Methods and Nomenclatures. Official Publication Office, Luxemburg.
Fay, M. & Opal, C. (2000). Urbanization without growth: a not-so-uncommon phenomenon. World Bank Policy Research Working Paper 2412.
Feria, J. M., Casado-Díaz, J. M. & Martínez-Bernabéu, L. (2015). Inside the metropolis: the articulation of Spanish metropolitan areas into local labor markets. Urban Geography, 3638, 1‑24.
Ferreira, J. A., Condessa, B., Castro e Almeida, J. & Pinto, P. (2010). Urban settlements delimitation in low-density areas-An application to the municipality of Tomar (Portugal). Landscape and Urban Planning, 97(3), 156–167.
Fields, G. (1990). Labour Market Modelling and the Urban Informal Sector: Theory and Evidence. In: D. Thurnham, Salome, B. and Schwarz, A. (Ed.), The Informal Sector Revisited. Paris, OECD.
Flórez-Revuelta, F., Casado-Díaz, J. M. & Martínez-Bernabéu, L. (2008). An evolutionary approach to the delineation of functional areas based on travel-to-work flows. International Journal of Automation and Computing, 5(1), 10–21.
Foster, N. & Stehrer, R. (2009). Sectoral Productivity, Density and Agglomeration in the Wider Europe. Spatial Economic Analysis, 4(4), 427-446.
Fox, K. A. & Kumar, T. K. (1965). The functional economic area: delineation and implications for economic analysis and policy. Papers and Proceedings of the Regional Science Association, 15, 57–85
Frankenhoff, C. A. (1967). Elements of an Economic Model for Slums in a Developing Economy. Economic Development and Cultural Change, Chicago 16, (1), 27.
Frey, W. H. & Speare, A. (1992). Metropolitan areas as functional communities a proposal for a new definition. Research Report 92, 245.
Frick, S. & Rodríguez-Pose, A. (2017). Big or Small Cities? On city size and economic growth. Growth and Change, 49 (1), 4–32.

Gajovic, V. (2013). Comparative analysis of different methods and obtained results for delineation of functional urban areas. Spatium, 29, 8–15.

Galiani, S., Gertler, P., Undurrage, R., Cooper,R., Martínes, S. & Ross, A. (2017). Shelter from the storm: Upgrading housing infrastructure in Latin American slums. Journal of Urban Economics (98), 187 – 213.

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

Gasparini, L. & L. Tornarolli. (2009). Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Surveys Microdata. Desarrollo y Sociedad, Primer Semestre, 13–80.

Gisbert, F. J. G. & Martí, I. C. (2014). El concepto europeo de ciudad: una aplicación para España. Investigaciones Regionales, 30, 145‑156.

Glaeser, E. L. (2011). Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier. Penguin Books. ISBN-13: 978-0143120544.

Glaeser, E. L., Kolko, J. & Saiz, A. (2001). Consumer city. Journal of Economic Geography, 1, 27-50.

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, David C. (2001). Cities and skills. Journal of Labor Economics, 19 (2), 316–342.

Glaeser, E. L. Ponzetto, G. A. M. & Zou, Y. (2016). Urban networks: connecting markets, people, and ideas. Regional Science, 95(1), 17–59.

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

Guerois, M., Bretagnolle, A., Giraud, T. & Mathian, H. (2012). A new database for the cities of Europe? Urban Morphological Zones (CLC2000) confronted to three national databases of urban agglomerations (Denmark, Sweden and France). Environment and Planning B, 39(3), 439-458.

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.

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

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

Harvard, S., Deguen, S. Bodin, J. Louis, K. & Laurent, O. (2008). A small-are index of socioeconomic deprivation to capture health inequalities in France. Social Science and Medicine, 67 (12), 2007-1016.

Herold, M., Goldstein, N. C. & Clarke, K. C. (2003). The spatiotemporal form of urban growth: Measurement, analysis and modelling. Remote Sensing of Environment, 86(3), 286–302.

Henderson, V. (2002). Urban primacy, external costs and quality of life. Resource and Energy Economics, 24, 95 – 196.

Henderson, V., Lee, T. & Lee Y. J. (2001). Scale externalities in Korea, Journal of Urban Economics, 49, 479-504.

Holmes, T. J. (2010). Structural, experimentalist, and descriptive approaches to empirical work in regional economics. Journal of Regional Science, 50, 5–22.

Holzer, M. Fry, J., Charbonneau, E. & Van Ryzin, G. (2009). Literature Review and Analysis Related to Optimal Municipal Size and Efficiency. Local Unit Alignment, Reorganization, and Consolidation Commission, SPAA papers, 44.

Jacobs, J. (1969). The Economy of Cities. Random House, New York.

Jaitman, L. (2015). Urban infrastructure in Latin America and the Caribbean: public policy priorities. Latin America Review, 24-13.

Kahn, M. E. (2010). New evidence on trend of the cost of urban agglomeration. Agglomeration Economics; Chapter of the book: 339 – 354. ISBN: 0-226-29789-6.

Karlsson, C. & Olsson, M. (2006). The identification of functional regions: theory, methods, and applications. The Annals of Regional Science, 40(1), 1–18.

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

Kim, S. (2008). Spatial Inequality and Economic Development: Theories, Facts and Policies, Working paper 16, Commission on Growth and Development.

Kim, H., Chun, Y. & Kim, K. (2013). Delimitation of Functional Regions Using a p-Regions Problem Approach. International Regional Science Review, 38(3), 235–263.

Kim, K., Chun, Y. & Kim, H. (2016). p-Functional Clusters Location Problem for Detecting Spatial Clusters with Covering Approach. Geographical Analysis, 49(1), 101–121.

Kim, K., Dean, D. J., Kim, H. & Chun, Y. (2015). Spatial optimization for regionalization problems with spatial interaction: a heuristic approach. International Journal of Geographical Information Science, 30(3), 451–473.

Klapka, P. & Tonev, P. (2013). Functional regions of the Czech Republic: comparison of simpler and more advanced methods of regional taxonomy. Geographica, 44(1), 45–57.

Koster, H. & Rouwendal, J. (2013). Agglomeration, commuting costs, and the internal structure of cities. Regional Science and Urban Economics, 43, 352 -366.

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

Lall, S.V. Shalizi, Z. & Deichmann, U. (2004). Agglomeration Economies in the Indian Industry. Journal of Development Economics, 73 (2), 643-673.

Lall, S. V. Selod, H. & Shalizi, Z. (2006). Rural–urban migration in developing countries: A survey of theoretical predictions and empirical findings. World Bank Policy Research Working Paper No. 3915. Washington, DC: The World Bank.

Liu, B.C. (1978). Variations in social quality of life indicators in medium metropolitan areas. American Journal of Economics and Sociology, 37(3), 241-260.

Lipton, M. (1977). Why Poor People Stay Poor: Urban Bias in World Development. The Scandinavian Journal of Economics, 80.
Maloney, W. F. (2004). Informality Revisited. World Development, 32(7), 1159-1178.
Marchetti. (1994). Anthropological Invariants in Travel Behaviour. Technological Forecasting and Social Change, 47, 75–88.
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.
Marshall, A. (1890). Principles of Economics. Book London: MacMillan.
Marx, B. Stoker, Stoker, T. & Suri, T. (2013). The economics of slums in the developing world. Journal of Economic Perspective, 27 (4), 187-210.
Masucci, A. P., Serras, J., Johansson, A. & Batty, M. (2013). Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. Physical Review, 88(2), 1–9.
McCann. (2001). Urban and Regional Economics. Oxford University Press.
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.
Milego, R. (2007). Urban Morphological Zones, Definition and procedural steps. Final Report, Copenhagen: European Environment Agency, ETC Terrestrial Environment.
Mitra, A. & Nagar, J. (2018). City size, deprivation and other indicators of development: Evidence from India. World Development, 106, 273–283.
Mion, G. & Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. Canadian Journal of Economics, 42 (1), 28–55.
Muth, R. F. (1969). Cities and Housing. Chicago, IL: University of Chicago Press.
Narayan, D. (2000). Voices of the Poor: Can Anyone Hear Us?. World Bank Publication, DC.
Obaco, M. & Díaz, J.P. (2018). Urbanization in Ecuador under the FUAs definition. Mimeo.
Obaco, M. & Dimitris, B. (2018). Deprivation Index for Small Areas Based on Census Data in Ecuador. Mimeo.
OECD. (2008). Handbook on Constructing Composite Indicators: Methodology and user guide. OECD Publishing: Paris (3): 162
OECD. (2012). Redefining “urban”. A New Way to Measure Metropolitan Areas. OECD Publishing, Paris.
OECD. (2013a). Definition of functional urban areas (FUA) for the OECD metropolitan database. Paris: OECD Publishing.
OECD. (2013b). How’s Life? 2013: Measuring Well-being. Paris: OECD Publishing.
OECD. (2015). OECD Urban Policy Reviews: China 2015. Paris: OECD Publishing.
OECD. (2016). Making Cities Work for All. Data and Actions for Inclusive Growth. Paris: OECD Publishing.
OECD. (2017). Employment Outlook. Annex Chapter 3.

Overman, H. & Venables, A. (2005). Cities in the Developing World. CEP Discussion Paper, 695, London Perry.

Patel, A., Koizumi, N. & Crooks, A. (2014). Measuring slum severity in Mumbai and Kolkata: A household-based approach. Habitat International, 41, 300-306.

Pham, H. M., Yamaguchi, Y. & Bui, T. Q. (2011). A case study on the relation between city planning and urban growth using remote sensing and spatial metrics. Landscape and Urban Planning, 100(3), 223–230.

Puderer, H. (2008). Defining and Measuring Metropolitan Areas: A Comparison between Canada and the United States Geography. Canada. Geography Working Paper Series Defining, 92.

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

Ravallion, M., Chen, S. & Sangraula, P. (2007). New evidence on the urbanization of global poverty. Population and Development Review, 33(4), 667–701.

Rodrigues da Silva, A. N., Manzato, G. G. & Pereira, H. T. S. (2014). Defining functional urban regions in Bahia, Brazil, using roadway coverage and population density variables. Journal of Transport Geography, (36), 79–88.

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

Rosenthal, S.S. & W.C. Strange. (2008). The Attenuation of Human Capital Spillovers. Journal of Urban Economics, 64, 373–389.

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

Royuela, V. & Ordóñez, J. (2018). Internal migration in a developing country: A panel data analysis of Ecuador (1982-2010). Papers in Regional Science, 1-23.

Royuela, V. & Vargas, M. A. (2009). Defining Housing Market Areas Using Commuting and Migration Algorithms: Catalonia (Spain) as a Case Study. Urban Studies, 46(11), 2381–2398.

Royuela, V. & Suriñach, J. (2005). Constituents of Quality of Life and Urban size. Social Indicators Research, Springer, 74, 549-572.

Royuela, V., Suriñach, J. & Reyes, M. (2003). Measuring Quality of Life in small areas and over different periods of time, analysis of the province of Barcelona. Social Indicators Research, 64, 51–74

Saisana M. (2014). Addiction Severity Index. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht.

Sanchez-Serra, D. (2016). Functional Urban Areas in Colombia. OECD Regional Development Working Papers, 2016/08, Paris: OECD Publishing.

Schmidheiny, K. & Suedekum, J. (2015). The Pan-European Population Distribution across Consistently Defined Functional Urban Areas. Economics Letters, 133, 10-13.

Shanzi, K. (2010). Agglomeration, productivity, and spatial spillovers across Chinese cities. Annals of Regional Science, 45, 157-179.

Simini, F., González, M. C., Maritan, A. & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. Nature, 484(7392), 96-100.

Smart, M. (1974). Labour market areas: Uses and definition. Progress in Planning, 2, 239–353.

Suedekum, J. (2006). Agglomeration and Regional Costs of Living. Journal of Regional Science, 46 (3), 529–543.

Tong, D. & Plane, D. A. (2014). A New Spatial Optimization Perspective on the Delineation of Metropolitan and Micropolitan Statistical Areas. Geographical Analysis, 46(3), 230–249.

Toregas, C., Swain, R., ReVelle, C. & Bergman, L. (1971). The Location of Emergency Service Facilities. Operations Research, 19(6), 1363-1373.

Toribio, J. F. (2008). Un ensayo metodológico de definición de las áreas metropolitanas en España a partir de la variable residencia-trabajo. Investigaciones Geográficas, 46, 49–68.

Turner, J. F. (1969). Uncontrolled urban settlement: problems and policies. Urbanization: development policies and planning, International Social Development Review, 1, United Nations, New York, 1968.

UN-Habitat. (2003). The challenge of slums. United Nations Human settlements program, Earthscan Publications on behalf of UN-Habitat.

UN-Habitat. (2015). Slum Almanac 2015 - 2016, UNON, Publishing Services Section, Nairobi. ISO 14001:2004-certified.

UNDP. (2005). Millennium Project—Investing in Development: A Practical Plan to Achieve the Millennium Development Goals. United Nations, London: Earthscan.

United Nations. (2014). World Economic Situation and Prospects, WESP. United Nations Publication Sales No. E.14.II.C.2, ISBN: 978-92-1-109168-7.

United Nations. (2015a). World Urbanization prospects: 2014 revision. Department of Economic and Social Affairs, Population Division, (ST/ESA/SER.A/366).

United Nations. (2015b). World Economic Situation and Prospects. United Nations Publication Sales No. E.14.II.C.2, ISBN: 978-92-1-109168-7, 198.

Van de Voorde, T., Jacquet, W. & Canters, F. (2011). Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data. Landscape and Urban Planning, 102(3), 143–155.

Veneri, P. (2016). City size distribution across the OECD: Does the definition of cities matter?, Computers, Environment and Urban Systems, 56, 86-94.

Veneri, P. (2017). Urban spatial structure in OECD cities: Is urban population decentralising or clustering?. Papers in Regional Science, forthcoming. https://doi.org/10.1111/pirs.12300.

Villacis, B. & Carrillo, D. (2012). Estadística Demográfica en el Ecuador: Diagnóstico y Propuestas. Instituto Nacional de Estadística y Censos (INEC). Quito ‑ Ecuador.

Wang, F. & Guldmann, J. M. (1996). Simulating urban population density with a gravity-based model. Socio-Economic Planning Sciences, 30(4), 245–256.

Weber, C. (2000). Urban agglomeration delimitation using remote sensing data, in: Remote sensing and urban analysis: GISDATA 9, 304.

Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. Remote Sensing of Environment, 117, 34–49.

Wheeler, A. (2001). Search, sorting, and urban agglomeration. Journal of Labor

Economics, 19 (4), 880–898.
World Bank (2011). World Development Report 2011: Conflict Security and Development. The World Bank, Washington.
Zax, J. S. (1994). When is a move a migration?. Regional Science and Urban Economics, 24(3), 341–360.