



**HAL**  
open science

# Growth agglomeration effects in spatially interdependent Latin American regions

Carolina Guevara

► **To cite this version:**

Carolina Guevara. Growth agglomeration effects in spatially interdependent Latin American regions. 2016. halshs-01281610

**HAL Id: halshs-01281610**

**<https://shs.hal.science/halshs-01281610v1>**

Preprint submitted on 2 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WP 1611 – March 2016

## **Growth agglomeration effects in spatially interdependent Latin American regions**

Carolina Guevara

### **Abstract:**

We investigate the effect of agglomeration on regional growth in Latin America, using panel data and spatial panel data techniques. By exploring the role of development in the agglomeration-growth relationship, we find evidence of the Williamson's hypothesis: agglomeration growth effects are magnified in less-developed regions. Moreover, we measure the spatial effects of agglomeration. They have a large geographical scope. International connections of Latin American regions are beneficial to obtain positive spatial effects of agglomeration. Nevertheless, spatial effects are stronger within countries. This finding points out the strong border effects in Latin America.

### **Keywords:**

growth, regional data, agglomeration economies, spatial interdependence, Latin America, urbanization, development

### **JEL codes:**

R11, O18, O54

# Growth agglomeration effects in spatially interdependent Latin American regions

Carolina Guevara

University of Lyon, Lyon F-69007, France

CNRS, GATE Lyon Saint-Etienne, Ecully, F-69130, France

Université Jean Monnet, Saint-Etienne, F-42000, France

e-mail: grace.carolina.guevara.rosero@univ-st-etienne.fr

## Abstract

We investigate the effect of agglomeration on regional growth in Latin America, using panel data and spatial panel data techniques. By exploring the role of development in the agglomeration-growth relationship, we find evidence of the Williamson's hypothesis: agglomeration growth effects are magnified in less-developed regions. Moreover, we measure the spatial effects of agglomeration. They have a large geographical scope. International connections of Latin American regions are beneficial to obtain positive spatial effects of agglomeration. Nevertheless, spatial effects are stronger within countries. This finding points out the strong border effects in Latin America.

**Keywords:** growth, regional data, agglomeration economies, spatial interdependence, Latin America, urbanization, development.

**JEL Classification:** R11, O18, O54

## 1 Introduction

The emergence of giants cities in developing countries needs attention because agglomeration in those cities enhances spatial inequalities, not only in the economic dimension but also in the social one. According to [Williamson \(1965\)](#), the level of agglomeration in these countries is indeed very high, much higher than in developed countries.

Identifying the mechanisms of agglomeration in these countries is a key element towards understanding how agglomeration impacts on their growth. From theoretical approaches based on urban economics and the link between growth theory and economic geography, there exists a strong positive correlation between agglomeration and growth. Indeed, such a link is bidirectional, determined by forward (demand) and backward (cost) linkages (Martin and Ottaviano, 1999, 2001; Baldwin and Forslid, 2000). However, the positive relationship may change due to different factors. The common explanation of reduced effects or even negative effects of agglomeration on growth is related with congestion effects. Further concentration of activities entails high housing costs, pollution, traffic and high transport and commuting costs which produces dispersion of economic agents. Another aspect, less studied in the literature, lies on the role of the level of development of countries in explaining the agglomeration-growth relationship. Little attention has been dedicated to such a factor (Venables, 2005; Duranton, 2007) most likely due to the implicit assumption that the experience of developed countries can be translated to that of developing ones. This is what Chauvin et al. (2013) called the knowledge mismatch which recognizes the importance of focusing research on developing countries. The fact that the economic conditions vary across the stages of development might matter for the effects of agglomeration. Indeed, the empirical evidence provided by Williamson (1965) shows that the degree of agglomeration is higher in early stages of development than in later stages. Accordingly, the effects of agglomeration on growth are more pronounced in countries with low levels of development (Brühlhart and Sbergami, 2009). Thus, powerful<sup>1</sup> mechanisms seem to occur in developing countries. The factors that cause the differences have not been explored.

This paper investigates how agglomeration impacts on growth and whether the mechanisms of agglomeration are related with the level of development. To answer these questions, we use sub-national data of Latin American countries.

The remainder of this paper is organized as follows. In section 2, the context of spatial disparity of Latin American countries is described. In section 3 we present the data, the strategy estimation and the results. Section 4 concludes.

---

<sup>1</sup>As Venables (2005) qualified the mechanisms of mega cities in developing countries that keep attracting more firms.

## 2 The context of inequality in Latin America

Lately, Latin America has experienced a continued economic expansion. During the last decade, the average growth rate<sup>2</sup> has increased from 2.4% in 2000, year in which several countries faced both internal and external crisis, to 5.8% in 2005 and by the end of the decade, the average growth rate has reached 6.5% in 2010, reflecting one of the best signs of recovery after the world recession in 2008. During the period 2004-2010<sup>3</sup>, these countries recorded an annual growth rate of 5.5% in average.

Within countries, the economic scenario is very heterogeneous. Few domestic regions concentrate most of the economic activity. Overall, capital regions alone or together with no more than two regions account for most of the Gross Domestic Product (GDP) of each country. For instance, the capital region of Peru, alone represented 52% of the total GDP; in Chile, the capital region accounted for 48%; in Ecuador only two regions recorded 54%; and in Colombia three main regions produced 52% of the total GDP. In general, spatial concentration at regional level is a common pattern in Latin America.

Such a spatial configuration has been the result of a sequence of historic events. It was first formed in concordance with Indian settlements (Aztec, Maya and Inca empires). Then, colonizing countries (mainly Spain and Portugal) promoted specific established settlements which become political and economic centers ([Atienza and Aroca, 2012](#); [Massiris-Cabeza et al., 2012](#)). Globalization was another key factor that reinforced the spatial concentration in regions where sources of comparative advantage of countries were located. Along time, those geographically advantaged regions have improved their urban infrastructure conditions. Thus, the process of urbanization has been different across regions within countries. The primacy of politically and economically favored regions has been enhanced. As shown in Figure 1, Panama, Paraguay, Uruguay, Argentina, Chile and Peru concentrate more than 40% of their total urban population in their primate regions. In other countries, urban population has been concentrated in the main region and secondary regions, for example Bolivia, Ecuador, Honduras, Colombia, Venezuela and Brazil.<sup>4</sup>

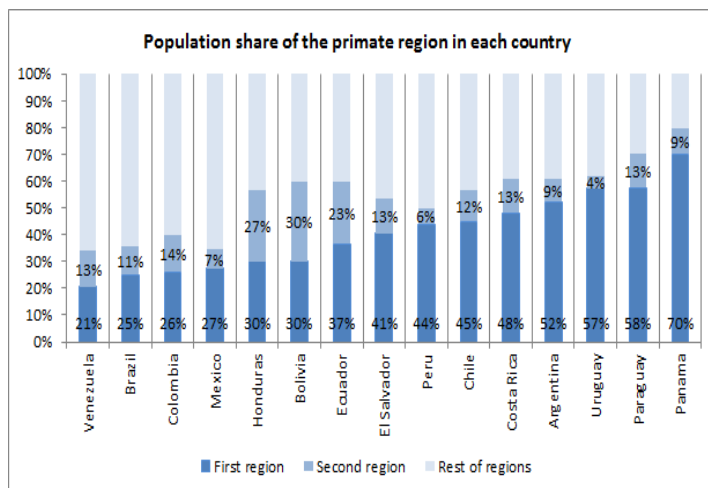
---

<sup>2</sup>These statistics were obtained using data from the Economic Commission for Latin America and the Caribbean (ECLAC). They correspond to the 10 biggest countries of South America, Mexico and Panama.

<sup>3</sup>The average on countries is taken since 2004 in order to avoid years of crisis in some of them.

<sup>4</sup>In Bolivia, there exists a system with two main regions: La Paz and Santa Cruz. Ecuador has two main urban centers: Guayas and Pichincha. Honduras also has two main regions: Francisco Morazan and Cortes. In Colombia, the two main regions are Cundinamarca and Antioquia. Venezuela has a two-region urban system with the State of Miranda whose main city is Caracas

Figure 1: Percentage of urban population in primate regions; 2000



Source: ECLAC database on National Population Censuses.

It is also argued that the trade policy of Import Substitution Industrialization<sup>5</sup> (ISI acronym is spanish) played an important role in reinforcing regional concentration and primacy within countries. The reason is that the increase of industrial activity in main cities has attracted great flows of internal migration from rural areas towards urban areas (Atienza and Aroca, 2012).

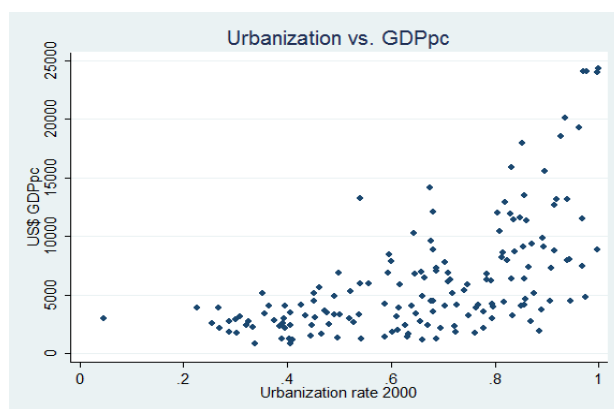
Spatial disparities have been accentuated not only in the economic aspect but also in the social aspect. The provision of services, access to health, education and technology differ across space. The gap between high income and low income regions is large. In economic terms, at a country level, the GDP per capita of the wealthiest region is 8 times as large as the poorest region. In social terms, some regions within countries have more than 60% of their population in poverty situation. Regions with high levels of urbanization, commonly the favored ones, provide high levels of wealth. As we can observe in Figure 2, the correlation between urbanization rate and GDP per capita is positive (the correlation coefficient is 0.37 and it is signifi-

and the State of Zulia whose main city is Maracaibo. Finally, Brazil was initially a urban system with three-main-regions: Sao Paulo, Rio de Janeiro and Minas-Gerais. However, Rio de Janeiro and Minas-Gerais have decreased their share of the total urban population.

<sup>5</sup>In the decades of the fifties and sixties, most of Latin American governments adopted the ISI policy which consisted in replacing foreign industrialized imports with domestic production.

cant) at regional level.

Figure 2: Urbanization and GDPpc



Source: ECLAC database on National Population Censuses and National Accounts of each country.

The analysis presented throughout this section reveals a high degree of spatial concentration within Latin American countries. The next section 3 is devoted to determine to what extent agglomeration affects growth of Latin American regions.

### 3 Estimation of the effects of agglomeration in Latin American regions' growth

Because spatial disparities are present not only in economic terms but also in social terms, the study of agglomeration effects in Latin American regions merits special attention. Nevertheless, empirical literature is scarce. The main reason is the lack of consolidated data. The present study surpasses such issue, being the first attempt to analyze the impact of agglomeration on Latin American regions' growth.

This study focuses on data at regional level for two reasons. First, the heterogeneity between regions can be taken into account. Second, the study of agglomeration economies entails geographical interactions which are more evident at a rather small spatial scale. The geographical proximity between economic agents promotes more immediate and dynamic connections. Thus, externalities of agglomeration are more

likely to be observed within regions.

### 3.1 Data

We use regional information of 166 regions of 8 countries: Argentina, Bolivia, Chile, Colombia, Ecuador, Mexico, Peru and Panama<sup>6</sup>. The inclusion of more variables reduces the sample to 87 regions of three countries: Colombia, Ecuador and Mexico. Hereinafter, we refer to the first database as *A* and to the second database as *B*.

We built a panel database for the period 2000-2009. Since the process of agglomeration takes time, we believe that the annual variation is not enough to explain growth (Gardiner et al., 2010; Bosker, 2007; Brülhart and Sbergami, 2009). Hence, we use a three year-period database: 2001-2003, 2004-2006 and 2007-2009.

The main sources of information are the National Statistical Institutes of each country. In particular, we obtained regional series for real GDP, population by age, surface area, education level of population and government expenditure.

The second data source is ECLAC (Economic Commission for Latin American countries) from which we obtained the Gross Domestic Product (GDP) at the regional and industrial level,<sup>7</sup> and urbanization rates at regional level<sup>8</sup> The geographical coordinates were collected from the tool of geo-localization GeoHack Wikimedia Toolserver Wiki.

Given that national currencies vary across Latin American countries, we standardize the measure of nominal GDP to real GDP using the Purchasing power parity (PPP) conversion factor of 2000 of the World Bank, International Compari-

---

<sup>6</sup>Data at sub-national level is limited in some Latin American countries. The countries that do not account for regional information of National Accounts are Costa Rica, El Salvador, Guatemala, Haiti, Honduras, Nicaragua and Venezuela.

<sup>7</sup>Data on industries is used to compute the industrial specialization index of each region.

<sup>8</sup>The database of urbanization contains information until 2000. We updated it with information on recent population censuses of countries. We impute an intermediate value of urban population between the last census and the recent census using an annual growth rate. The rationale behind this computation is that urbanization is an increasing process over time. This allows capturing the effect of urbanization in the dynamic economic growth rate, using three year period data.



son Database.<sup>9</sup> Another data issue that we face is the difference in the procedures<sup>10</sup> measuring urban population in each country. We consider that the variety of procedures constitutes a robustness test since if a pattern is observed despite the diversity of measures, the result would be more persuasive.

## 3.2 Estimation Strategy

Using both databases, we estimate the impact of agglomeration on regional growth of Latin America. A parsimonious model of agglomeration is estimated using sample *A* and an extended model including more variables using sample *B*. The general specification of the model is:

$$\Delta y_{i,t+k}/y_{i,t} = \alpha + \psi \ln y_{i,t} + \gamma \text{agglom}_{i,t} + X_{it}\beta + \mu_i + \xi_t + \varepsilon_{it} \quad (1)$$

where  $\Delta y_{i,t+k}/y_{i,t}$  is the growth rate<sup>11</sup> of per capita GDP of region *i* between time *t* and *t* + 3<sup>12</sup>,  $y_{i,t}$  is the initial per capita GDP of region *i* and its associated coefficient estimate  $\psi$  is generally interpreted as the conditional convergence parameter in panel models (Barro and Sala-i Martin, 1992; Arbia and Piras, 2005). If the coefficient estimate is negative, poor regions tend to grow faster than rich ones and convergence takes place<sup>13</sup>.  $\text{agglom}_{it}$  is the level of agglomeration of region *i* at time *t*. It is proxied by two indicators: urbanization rate and population density. The vector  $X_{it}$  includes additional explanatory variables, such as the level of education of region *i* at time *t* which is measured by gross school enrollment ratio in tertiary

---

<sup>9</sup>The PPP factor measures the number of units of a country's currency required to buy the same amount of goods and services in the domestic market as U.S. dollar would buy in the United States. Thus, it allows transforming the GDP in national currencies to dollars in real terms, which are comparable.

<sup>10</sup>Overall, there are three general methods to establish the urban character: i) localities with more than 2000 inhabitants, ii) definition (i) combined with the provision of public services and urban conditions, and iii) capital cities constitute urban areas; the areas outside capitals are considered rural.

<sup>11</sup>It is worth noting that we do not use GDP in levels because of two reasons. First, we would deviate from the theoretical literature that focuses on the effect of agglomeration on growth and not in economic performance. Second, the relationship between urbanization and GDP in levels is potentially endogenous.

<sup>12</sup>We use the logarithmic formula to calculate the growth rate.  $\Delta y_{i,t+k}/y_{i,t} \cong \ln y_{i,t+k} - \ln y_{it}$

<sup>13</sup>It is worth noting that only three periods are taken into account. Then, the results of convergence do not show a long-run equilibrium. It is interpreted as the convergence of regions to their own steady states.

education; the role of government in region  $i$  at time  $t$  measured by the share of public investment in regional GDP<sup>14</sup>, and industrial specialization<sup>15</sup> of regions which is computed using the formula:  $spe_i = \max_j \frac{GDP_{ij}/GDP_i}{GDP_j/GDP}$  where  $i$  refers to the region and  $j$  refers to the industry. It is calculated at each time  $t$ .  $\mu_i$  represents the non observed regional-specific effects,  $\xi_t$  represents the time-specific effects<sup>16</sup> and  $\varepsilon_{it}$  is the idiosyncratic error term which is independent and identically distributed with mean zero and constant variance. The parsimonious model estimates the parameters  $\alpha$ ,  $\psi$  and  $\gamma$ . The extended model includes the parameters in  $\beta$ .

Since the level of agglomeration,  $agglom_{i,t}$ , is likely to be correlated with regional characteristics  $\mu_i$  such as geography or historic factors of development (commonly not observed), the assumption of uncorrelated errors in Ordinary Least Squares estimation is violated. Then, the OLS estimator will be both biased and inconsistent. In order to obtain consistent and efficient estimates, a panel model with fixed effects is adequate. It allows controlling for unobserved heterogeneity.<sup>17</sup>

Another issue when estimating the effect of agglomeration on growth is the endogeneity from reverse causality. In the literature, it has been well documented that the relationship between the two is bidirectional (Martin and Ottaviano, 1999, 2001; Baldwin and Forslid, 2000). The method of instrumental variables is appropriate to treat reverse causality. It consists in finding a third set of variables  $z$  which has to be relevant and valid (Combes and Lafourcade, 2012). Geological variables are commonly used to correct the endogeneity issue of agglomeration (Combes et al., 2008). The instrument that we use is altitude<sup>18</sup>. According to the difference-

---

<sup>14</sup>It is the annual average of three years in each period.

<sup>15</sup>We compute the indicator of specialization based on seven aggregated sectors: 1. the primary sector, 2. the manufacturing sector, 3. the market services sector, 4. public services 5. Wholesale and retail trade, 6. Electricity, gas, and Water supply; and 7. Construction.

<sup>16</sup>By including time fixed effects for periods 2001-2003 and 2007-2009, we control the effects of internal and external crisis. Between 1999 and 2002, external and internal crisis have occurred: the Asian crisis in 1998, the Brazilian crisis in 1999 (Brazil is not included in our analysis), the Argentinean crisis in 2000-2001 and the Ecuadorian crisis in 1999. In 2008, world crisis took place.

<sup>17</sup>In the case where  $u_{it}$  is assumed to be fixed and the remainder disturbances are stochastic with  $\varepsilon_{it}$  independent and identically distributed with mean zero and a constant variance, the fixed effects model is an appropriate specification. The second case is to assume that  $u_{it}$  is random where the individual effect is characterized as random. In order to choose between fixed effects model or random effects model, the Hausman test will be used. (For more details about the Hausman (1978) test, see Baltagi (1995))

<sup>18</sup>We use the average altitude of the capital city in each region.

in-Sargan statistic<sup>19</sup>, we cannot reject the null hypothesis that urbanization rate is exogenous (sample A: C-statistic=0.009, p-value=0.925; sample B: C-statistic=0.10, p-value=0.747).

### 3.2.1 Spatial data issues

When using spatial data, a general issue is the interaction between spatial units. In our case, the spatial dependence is likely because one region's growth can affect neighboring regions' growth. In empirical literature, this issue has been considered (Ciccone and Hall, 1995; Ciccone, 2002; Crozet and Koenig, 2008; Bosker, 2007), however not systematically. Since the spatial dependence violates the independence assumption in the panel model, we are required to test it. We use the Cross Dependence (CD) test proposed by Pesaran (2004)<sup>20</sup> for panel data models and the test proposed by Moran (1948), the so-called Moran's I test<sup>21</sup>.

If evidence of interdependence among regions exists, we use spatial econometric tools to deal with this issue. We presume that spatial dependence occurs in the context of growth. For that reason, we consider a model specification with a spatial lag of the dependent variable as follows.

$$\Delta y_{i,t+k}/y_{i,t} = \alpha + \psi \ln y_{i,t} + \gamma \text{agglom}_{it} + \rho \sum_{j=1}^N w_{ij} \Delta y_{j,t+k}/y_{j,t} + X_{it}\beta + u_{it} \quad (2)$$

where  $\rho$  is the spatial dependence coefficient associated with the spatial lag of the dependent variable,  $w_{ij}$  are the elements of a spatial weight matrix that represents geographical relationships between regions and  $u_{it}$  is the error term.<sup>22</sup> In the

---

<sup>19</sup>Under the null hypothesis, the suspected variable can be treated as exogenous. The test statistic is distributed as chi-squared with degrees of freedom equal to the number of regressors tested.

<sup>20</sup>The Pesaran's CD test is based on average of pair-wise correlation coefficients of the OLS residuals from the individual regressions in the panel, and it is used to test for cross section dependence. Through the investigation of the small sample properties of the test, it is shown that the test is appropriate for small samples. Furthermore, the Pesaran's CD test does not require a priori specification of the spatial matrix. For details of the test statistic, see Pesaran (2004)

<sup>21</sup>The Moran's I is a global index of spatial correlation that indicates the degree of similarity between geographical units. The index ranges between -1 (dissimilar values are closer) and 1 (similar values are closer).

<sup>22</sup>In order to know whether this specification using the spatial lag of the dependent variable is appropriate, the spatial dependence after the regression has to be ruled out.

literature, many ways of specifying the spatial dependence have been considered. In our model, three specifications of the spatial weight matrix are tested: k=1 nearest neighbors-based matrix, distance weight matrix and gabriel-method-based weight matrix. These weight matrix specifications are explained in more detail later in Section 3.3.2.

The methodology used is the Maximum Likelihood Estimation because it provides minimum variance unbiased estimators. Given that our database is not large, the computational problems that this method could cause are ruled out.<sup>23</sup> A specific-to-general approach is used. We start with a non-spatial panel model and then we test whether the model needs to be extended with spatial interaction effects.

## 3.3 Results

### 3.3.1 Non-Spatial Panel models

In Table 1, we present<sup>24</sup> the fixed effects panel model estimations using two measures of agglomeration: urbanization in columns (1, 3) and population density in columns (2, 4). We show the results for both samples A and B.

The standard errors of the coefficients are robust to heteroskedasticity. According to the Hausman test (bottom of Table 1), the assumption that non-observable individual effects are not correlated with explanatory variables does not hold true and fixed effects model is preferred.

---

<sup>23</sup>NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, 15/01/2014

<sup>24</sup>Outlier observations are eliminated. They are Campeche (Mexico), Buenos Aires (Argentina), CABA (Argentina) and El Beni (Bolivia).

Table 1: Non Spatial Panel models using sample A and sample B

	(1)	(2)	(3)	(4)
	urb A	Popd A	urb B	Popd B
urb	0.634 (2.883)***		1.067 (2.456)**	
ln Population density		0.172* (1.912)		0.111 (0.882)
initial GDPpc	-0.160 (-3.772)***	-0.149 (-3.457)***	-0.262 (-3.251)***	-0.219 (-2.542)**
yearcris09	-0.104 (-15.126)***	-0.111 (-13.720)***	-0.106 (-12.419)***	-0.105 (-11.938)***
yearcris01	-0.144 (-9.571)***	-0.142 (-10.155)***	-0.137 (-5.174)***	-0.145 (-6.045)***
Constant	1.014 (3.383)***	0.782 (2.511)**	1.654 (2.981)***	1.535 (3.032)***
N observations	486	486	258	258
N regions	162	162	86	86
Region-specific effects	Y	Y	Y	Y
F	133.0	136.0	110.1	98.52
p-value F	0.000	0.000	0.000	0.000
$R^2$	0.563	0.560	0.599	0.584
Hausman test	28.99	38.67	29.77	25.08
p-value	0.000	0.000	0.000	0.000

*t* statistics in parentheses

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < 0.01$

All estimations show a positive and significant effect of agglomeration on economic growth after controlling for region-specific effects and time-specific effects. Both proxies of agglomeration yield to similar results in terms of the coefficient sign; but the effect is not significant enough when using population density. One interpretation of this result might be that urbanization rate is a better measure of agglomeration because it ensures a structure with appropriate conditions to generate agglomeration externalities.

The coefficient estimate of the GDP per capita at the initial year of each period shows the beta convergence conditional on the level of agglomeration. The speed of convergence <sup>25</sup> is about 6% annually when using urbanization and 5% when using population density. These results are larger compared to the [Serra et al. \(2006\)](#)'s estimates of convergence which range between 1% and 2%. They conditioned the convergence rate only with regional characteristics. Hence, regions reach their steady states much faster if the level of agglomeration is taken into consideration.

Finally, the assumption that economic growth outputs across regions are independent might not hold true because regions are very likely to be connected to each other, specially within countries. Then, we test the spatial dependence in the model using the Cross Dependence (CD) test proposed by [Pesaran \(2004\)](#) for panel data and the Moran's I test<sup>26</sup>. For the latter, we use three spatial matrices: 1. the  $k = 1$ -nearest neighbors (W k1), 2. the neighborhood based on Gabriel method (W g) and 3. distance weight matrix (W d). Now, we provide basic concepts of such matrices, leaving a detailed description to Section 3.3.2<sup>27</sup>. The first matrix implies that each region has only one neighbor. The second matrix implies more neighboring regions than the former. The third matrix relates all regions with each other according to the distance between them.

In Table 2, the CD Pesaran's test shows that outputs of regions are spatially correlated in both samples A and B. Likewise, the Moran's I reports global spatial correlation using any of the spatial matrices. We address this issue by estimating our specification 2 using spatial panel data techniques in Section 3.3.2.

---

<sup>25</sup>The speed of convergence is calculated as:  $s = -\ln(\psi + 1)/k$  where  $k$  is equal to 3 (3 periods) in our case. ([Bosker, 2007](#))

<sup>26</sup>The Moran's I statistic is computed as in [Anselin \(1995\)](#):  $I = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N \sum_{j=1}^N w_{ij}}$ .

<sup>27</sup>Other spatial configurations are included in the analysis of spatial correlation between Latin American regions in Section 3.3.2.

Table 2: Spatial Dependence tests

Sample A			Sample B	
Test	Statistic value	p-value	Statistic value	p-value
CD Pesaran's test	15.025	0.0000	7.023	0.0000
Moran's I (W k1)	period1=0.422	0.000	0.223	0.048
	period2= 0.438	0.000	0.288	0.016
	period3=0.437	0.000	0.348	0.005
Moran's I (W g)	period1=0.43	0.000	0.235	0.006
	period2= 0.399	0.000	0.181	0.024
	period3= 0.439	0.000	0.225	0.008
Moran's I (W d)	period1=0.149	0.000	0.053	0.030
	period2=0.146	0.000	0.043	0.056
	period3= 0.168	0.000	0.049	0.039

### 3.3.2 Spatial panel models

In the literature, only few studies analyze the spatial interdependence between Latin American countries. By examining the knowledge spillovers and their channels of diffusion in South America, [Guevara and Autant-Bernard \(2015\)](#) show that the level of productivity in one country generates indirect effects in other countries' productivity. They state that the spatial dependence is driven by both physical proximity and trade intensity between countries. [Ramírez and Loboguerrero \(2002\)](#) show that spatial dependence is significant in a worldwide regression. Notwithstanding, the results for Latin America show no significance of the spatial dependence term. [Blanco \(2011\)](#) focuses on the role of spatial interdependence in the Foreign Direct Investment in Latin America. Her results show that the spatial correlation is not significant.

In our case, it is crucial to look at the intensity of spatial interconnections between Latin American countries because they could influence the effects of agglomeration. To this end, two types of spatial configuration between Latin American regions are considered.

- A: High interaction between countries implying international connections of their domestic regions. In this configuration, the link of country  $x$  and country  $y$  immediately implies relationships between domestic regions of both countries.
- B: Low interaction between countries implying weak international connections of regions. In this configuration, the link between domestic regions in the same

country is high. But the link of domestic regions of different countries is lower.

It is worth noting that pairs of regions of the same country tend to establish a stronger connection than pairs of regions of different countries due to national border barriers. These are the so-called border effects (McCallum, 1995). Besides tariffs, another factor determining border effects in Latin America is the difference between national currencies.

To construct the weight matrices<sup>28</sup>, we use the coordinates of capitals of each region and eliminate islands (Galapagos-Ecuador and San Andres-Colombia). We apply the  $k$ -nearest neighbors method and the Gabriel method<sup>29</sup>. We use the inverse of row standardized matrices as suggested by Anselin (1988). The  $k$ -nearest neighbors criterion implies that region  $i$  is considered as neighbor of region  $j$  if their distance is equal, or less than equal, to the minimum possible distance that can be found between region  $i$  with all other regions. This definition ensures that each spatial unit has the same number  $k$  neighbors. For instance, allowing  $k = 1$  implies that each region has one neighbor; allowing  $k = 4$  implies that each region has four neighbors. The resulting graphs using such a methodology are shown in Figure 3.

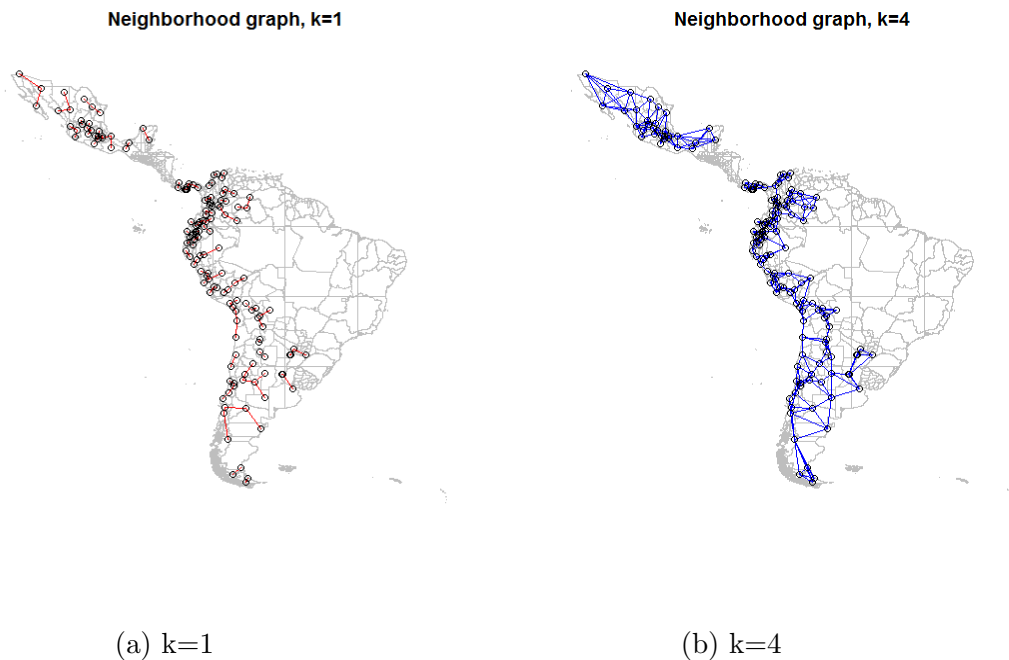
---

<sup>28</sup>We use the software R to create the spatial weight matrices.

<sup>29</sup>This method was proposed by Gabriel and Sokal (1969).



Figure 3: k-nearest neighbors configuration



In case A of high interaction, we use the spatial weight matrix setting  $k$  to the maximum. In sample  $A$ , there are 160 regions. Then, each region can have at most 159 neighbors. Then, we set  $k = 159$ . The elements  $w_{ij}$  of such a matrix represent the distance between region  $i$  and  $j$  where every region  $i$  is connected to all other regions ( $i \neq j$ ). This spatial weight matrix is noted as  $(W d)$ .

In case B of low interaction, we use the nearest neighbor-based spatial weight matrix,  $k = 1$ . It is noted as  $(W k1)$ . We also define a spatial weight matrix to illustrate pure border effects by assuming that regions from different countries do not interact at all. To do so, the connections in  $(W d)$  of regions of country  $x$  with those of country  $y$ ,  $x \neq y$  are set to 0. The resulting matrix  $(W beH)$  connects all domestic regions within countries but does not connect international regions. The high intensity of connections of domestic regions is represented by  $H$  in  $(W beH)$ .

Using the Gabriel's method, we construct a spatial weight matrix noted as  $(W g)$ . Such a matrix connects countries through their frontier regions. Thus, it also represents the case A of high interaction but to a lesser extent than the spatial matrix  $(W d)$ . The connectivity graph is shown in Figure 4. We also use this method to construct another weight matrix of pure border effects. To do so, we proceed as

in  $(W_d)$ . The resulting weight matrix  $(W_{beL})$  reflects less number of connections of domestic regions within countries than in  $(W_{beH})$ . Low intensity of domestic connections is represented by  $L$  in  $(W_{beL})$ .

Figure 4: Gabriel method graph



Elaboration: by author

To sum up, spatial weight matrices reflecting high interaction are:  $(W_d)$ , interpreted as quasi-complete regional interaction; and  $(W_g)$  interpreted as moderate regional interaction. Spatial matrices reflecting low interaction are:  $(W_{k1})$ , interpreted as the configuration of lowest regional interaction;  $(W_{beH})$ , interpreted as pure border effects with high level of domestic connections; and  $(W_{beL})$  interpreted as pure border effects with moderate level of domestic connections. All matrices illustrate the border effects definition because the intensity of connections decreases with distance between regions.

In order to deal with spatial dependence<sup>30</sup>, we estimate a Spatial Autoregressive Model by Maximum Likelihood. The spatial panel model includes a spatially-lagged

---

<sup>30</sup>After including the spatial dimension in regressions, the residuals do not present spatial dependence in most of the models according to the Moran's I test shown in Appendix A.

dependent variable capturing interaction effects between regions. Region-specific effects and time-specific effects are also included.

Before presenting the results, it is worth noting that the presence of the spatial autoregressive parameter in our model produces indirect effects that come from geographical proximity. Thus, the total effect of the changes in the explanatory variables has two components: the direct effect or local effect and the indirect effect or spatial effect. These effects come from the matrix of elasticities  $\Xi$ . In the case of the estimate coefficient of agglomeration, we have the following matrix.

$$\Xi_Y = \hat{\gamma}(I - \hat{\rho}W)^{-1}$$

The direct effect of agglomeration in region  $i$  on its economic growth is represented by the  $i,i$ th element from the diagonal of matrix  $\Xi$ . The indirect effect of agglomeration of region  $j$  on economic growth of region  $i$  is represented by the  $i,j$ th element of such a matrix.

Table 3 shows the results using the aforementioned spatial weight matrices; in column (1), the distance matrix ( $W_d$ ) representing the configuration of high regional interaction, in column (2), the matrix based on the Gabriel method ( $W_g$ ) representing the moderate regional interaction, in column (3), the matrix based on  $k$  nearest neighbor with  $k=1$  ( $W_{k1}$ ) representing the configuration of low regional interaction. The results regarding border effects with high intensity of intra-connection ( $W_{beH}$ ) are shown in column (a) and border effects with low intensity of intra-connection ( $W_{beL}$ ) in column (b). Both direct and indirect effects are presented.

Table 3: SAR Model with Parsimonious specification-Database A

	(1)	(2)	(3)	(a)	(b)
	(W d)	(W g)	(W k1)	(W beH)	(W beL)
<b>Direct</b>					
urb	0.598 (3.332)***	0.594 (3.274)***	0.641 (3.508)***	0.580 (3.224)***	0.572 (3.159)***
initial GDPpc	-0.167 (-5.779)***	-0.161 (-5.510)***	-0.165 (-5.621)***	-0.165 (-5.690)***	-0.159 (-5.486)***
yearcris01	-0.120 (-12.389)***	-0.124 (-13.035)***	-0.129 (-13.635)***	-0.121 (-12.761)***	-0.123 (-12.965)***
yearcris09	-0.0818 (-10.546)***	-0.0865 (-11.476)***	-0.0934 (-12.656)***	-0.0763 (-9.521)***	-0.0845 (-11.200)***
<b>Spatial</b>					
$\rho$	0.403 (6.054)***	0.225 (5.797)***	0.134 (4.677)***	0.610 (6.478)***	0.245 (6.324)***
<b>Variance</b>					
$\sigma_e^2$	0.00173 (15.481)***	0.00172 (15.383)***	0.00176 (15.404)***	0.00171 (15.448)***	0.00169 (15.365)***
<b>Indirect</b>					
urb	0.446 (2.573)**	0.166 (2.845)***	0.0975 (2.892)***	0.261 (2.708)***	0.162 (2.832)***
initial GDPpc	-0.125 (-3.256)***	-0.0448 (-3.989)***	-0.0251 (-3.817)***	-0.0741 (-3.693)***	-0.0451 (-4.145)***
<b>Total</b>					
urb	1.043 (3.170)***	0.760 (3.262)***	0.738 (3.496)***	0.840 (3.181)***	0.734 (3.155)***
initial GDPpc	-0.292 (-4.935)***	-0.205 (-5.457)***	-0.190 (-5.591)***	-0.239 (-5.396)***	-0.204 (-5.450)***
N observations	480	480	480	480	480
N regions	160	160	160	160	160
Region-specific effects	Y	Y	Y	Y	Y
$R^2$	0.585	0.584	0.581	0.587	0.585
Hausman test	16.40	16.67	17.88	17.74	17.45
p-value	0.00580	0.00516	0.00310	0.00330	0.00372

*t* statistics in parentheses

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < 0.01$

Firstly, we observe that Latin American regions have a positive and highly significant effect of urbanization in their economic growth after controlling for spatial dependence. The coefficient estimate of the direct effect is stable across all spatial configurations. It is slightly lower than the coefficient estimate of the panel model (see Table 1). The reason could be that a part of the total effect is attributed to spatial interactions between regions.

The spatial effects of urbanization are significant. It indicates that urbanization in one region brings benefits for growth of its neighbors. It could be indirectly related to knowledge spillovers among regions. Urban conditions in one region would allow knowledge diffusion towards neighboring urbanized regions thank to geographical proximity. Thus, economic growth of close regions is positively affected.

The spatial effect of urbanization decreases as the level of regional interaction is lower. Thus, the highest indirect effect of urbanization is observed when using the distance weight matrix ( $W_d$ ), and the lowest effect when using the  $k=1$  nearest neighbors-based matrix ( $W_{k1}$ ). Regarding spatial configurations with border effects, the indirect effects of urbanization are larger in the framework of high intensity of internal connections ( $W_{beH}$ ) than in the framework of low intensity of internal connections. The more interconnected the regions, the higher the spatial growth effects of urbanization.

The total effect of the coefficient of convergence is close to the panel model's estimate. The contribution of indirect effects coming from geographical proximity between regions to convergence is small but significant. Besides, those spatial effects vary depending on the spatial configuration used. The proximity between regions matters for the process of convergence as stated by [Serra et al. \(2006\)](#).

The spatial autocorrelation coefficient is statistically positive and significant for all models. It assess the extent to which one region's growth affects the economic growth of other regions. The degree of spatial autocorrelation increases with the level of connections that spatial weight matrices reflect. For instance, the value of  $\rho$  is high when using the distance weight matrix ( $W_d$ ), which exhibits a quasi-complete regional interaction and it is low when using the  $k = 1$  nearest neighbor matrix ( $W_{k1}$ ), which exhibits low regional interaction between Latin American regions. We consider that the model using ( $W_g$ ) spatial weight matrix which links South America with Mexico is appropriate because in trade terms Mexico is one of the ten largest export destination countries of South America accounting for 3% of

total exports; and it is one of the ten largest import origin countries representing 3.8% in total imports (Guevara and Jarrin, 2011).

Regarding the analysis of border effects spatial configurations, interesting insights are drawn. By eliminating connections of regions from different countries and keeping the internal connections (W beH), the spatial correlation coefficient  $\rho$  is large, even larger than the coefficient estimate of (W d) where all regions are connected. It means that the interdependence between regions within one country is much higher than between regions of different countries. Similarly, the spatial autocorrelation coefficient of (W beL) is higher than that of (W g).<sup>31</sup> Based on these results, we confirm that border effects are significant in Latin America. The aspects that could explain this result are related to the diversity of national currencies across countries and the scarce infrastructure across frontiers.<sup>32</sup>

### 3.3.3 Spatial Panel Model, Williamson’s hypothesis

Now, we turn to the analysis regarding the role of the level of development of regions in the agglomeration-growth relationship. Based on the literature (Williamson, 1965; Henderson, 2000; Davis and Henderson, 2003; Brülhart and Sbergami, 2009)(Henderson, 2003), we presume that the effect of urbanization on growth might be different between regions according to their level of development. The intuition is the following. Since transport and communication infrastructure is scarce at low levels of development, activities concentrate in the region that provides better conditions (Brülhart and Sbergami, 2009). Conversely, at high levels of development, better transport connection would promote dispersion. Therefore, agglomeration effects could vanish in such a dispersed environment.

In order to test the Williamson’s hypothesis, the econometric strategy is to run various simulations<sup>33</sup> using sub-samples built out of cutoffs according to the level of

---

<sup>31</sup>Recall that (W beL) is based on the spatial configuration of (W g). For that reason, we compare the results between these two matrices.

<sup>32</sup>Only recently, UNASUR has proposed an Initiative for the Integration of South America regarding infrastructure (IIRSA). The planning of such a project started in 2000 and its implementation in 2005. Hence, the outcomes of said initiative are not tangible yet. In this regard, one direction of future research is the assessment of the implementation of infrastructure projects proposed by IIRSA in modifying the strength of border effects. To this end, spatial weight matrices using actual flows of trade or road density would be appropriate instead of spatial weight matrices based on geographical distance used here.

<sup>33</sup>An alternative econometric strategy would be to introduce an interaction term between urbanization and the level of development. However, such an interaction term implies high collinearity

development (the initial per capita GDP). Put differently, we look at the changes of the effect of urbanization on growth at different levels of development. Through this exercise, the threshold at which the effects of urbanization on growth reach their maximum could be identified. Such a threshold is US\$5700 of per capita income.<sup>34</sup> Likewise, we attempt to determine the threshold at which urbanization effects become negative. According to simulations, negative effects seem to appear at US\$10,500 of per capita income. This value is close to the threshold identified by [Brühlhart and Sbergami \(2009\)](#) at US\$10,000 in a worldwide analysis. The results of simulations are shown in Table 7 of Appendix B.

Based on our threshold of US\$5700, we split the sample A into more developed regions and less developed regions. The group of high-developed regions contain 29 regions from Mexico (total 31); 12 from Chile (total 13); 12 from Argentina (total 22), 8 from Colombia (total 33), 1 from Peru (total 24) and 2 from Panama (total 9). The rest of regions of those countries and Bolivian (total 8) and Ecuadorian (total 21) regions are classified into the group of less developed regions.<sup>35</sup>

Table 4 presents the results of Spatial Autoregressive models for each group of regions: low-developed regions (column 1) and high-developed regions (column 2).

---

with urbanization rate and the estimates are no longer efficient. In order to avoid multicollinearity, the database is divided in sub-samples according to the initial per capita GDP of regions. Then, we run the regression using a sub-sample of regions that have an income level less/higher than a certain value and we look at the effect of urbanization for those regions.

<sup>34</sup>It is noteworthy that [Henderson \(2000\)](#) found that primacy generates negative effects on growth. The annual growth rate losses from excessive primacy increases until a level of income of US\$4900.

<sup>35</sup>The total of regions corresponds to the regions of study without outliers.

Table 4: Spatial panel models for low-developed and high-developed regions

	(1)	(2)
	(W g) Low-dev	(W g) High-dev
<hr/>		
Direct		
urb	0.682 (2.462)**	0.350 (1.286)
initial GDPpc	-0.177 (-4.955)***	-0.102 (-3.034)***
yearcris01	-0.142 (-10.574)***	-0.0815 (-8.180)***
yearcris09	-0.0849 (-6.136)***	-0.0810 (-10.002)***
<hr/>		
Spatial		
$\rho$	0.235 (4.536)***	0.287 (6.074)***
<hr/>		
Variance		
$\sigma_e^2$	0.00214 (11.904)***	0.000766 (9.712)***
<hr/>		
Indirect		
urb	0.201 (2.359)**	0.143 (1.412)
initial GDPpc	-0.0513 (-3.124)***	-0.0406 (-2.407)**
<hr/>		
Total		
urb	0.897 (2.845)***	0.504 (1.487)
initial GDPpc	-0.230 (-4.512)***	-0.145 (-2.687)***
<hr/>		
N observations	288	192
N regions	96	64
Region-specific effects	Y	Y
$R^2$	0.552	0.753
Hausman test	12.69	7.583
p-value	0.0264	0.181

*t* statistics in parentheses

\*  $p < .10$ , \*\*  $p < .05$ , \*\*\*  $p < 0.01$



The results confirm the Williamson's hypothesis claiming that positive agglomeration effects on economic growth increase up to a certain threshold of development level. The total effect of urbanization for low-developed regions is significant at low stages of development whereas at later stages, agglomeration does not impact on economic growth. Hence, most of Bolivian, Ecuadorian, Colombian, Peruvian and Panamanian regions enjoy increasing benefits of urbanization. Conversely, most of Mexican, Chilean and Argentinean regions face decreasing benefits of agglomeration, still positive but not significant.

The conditions of development might produce specific mechanisms that provoke such differences. One of those conditions could be the urbanization rate itself. At low income levels, the pace of urbanization is rapid (Henderson, 2003). In our sample, low-developed regions record an average growth rate of urbanization of 5.8% between 2001 and 2007,<sup>36</sup> whereas high-developed regions have a lower average growth rate of urbanization of 2.3% in the same period. The difference between those means is statistically significant (t statistic= 13.64, p-value=0.000).

We also analyze other variables, omitted in the model, that account for conditions of development of each country. Such omitted variables are most likely to be part of fixed-specific effects in the models as they represent characteristics of development that are not assumed to abruptly change in a short period of time. According to our threshold, most of regions of Mexico, Chile and Argentina are classified in the group of high-developed regions of Latin America. In terms of education, Mexico and Chile stand as countries with the highest shares of working population with master's degree.<sup>37</sup> The countries with the highest ratios of research and development expenditure over GDP are Argentina (0.48%) and Mexico (0.43%) followed by Chile (0.43%) and Ecuador (0.39%). Regarding transportation infrastructure, Panama (34.6%), Mexico (32.8%) and Argentina (29.4%) record the highest percentages of road paved.<sup>38</sup>

In short, Mexico, Chile and Argentina are commonly the countries with the best indicators in terms of development. The high values of those macroeconomic

---

<sup>36</sup>These statistics corresponds to sample A divided into groups of developed and developing regions.

<sup>37</sup>The indicators are 0.087% for Mexico and 0.11% for Chile in 2009. The information is obtained from the database of Network of Indicators of Science and Technology (RICYT acronym in spanish).

<sup>38</sup>The information corresponds to 1999 and the source of the data is the World Development Indicators of the World Bank.

variables implicitly reflect the enormous development of their internal markets. A process of rapid agglomeration surely took place in those markets which, in turn, induced further agglomeration. The progress is such that one can reasonably ask whether the benefits of agglomeration might be reaching their limits. According to our results, most of regions of these countries face non-significant effects. Conversely, it seems that the other countries with low levels of development are in a certain stage in which their embryony local infrastructure, markets, human capital and other development conditions promote an ongoing process of agglomeration which produces positive effects for growth.

Furthermore, there are indications that negative effects might emerge at higher levels of income than US\$10,500. Notwithstanding, the negative effect of urbanization is weak and not significant. Therefore, Latin American regions do not yet face strong negative effects.

The spatial correlation parameter is higher in high-developed regions than in low-developed regions. However, the difference between both coefficient estimates is not statistically significant (t-statistic = -0.744391226, p-value > 0.05).

### 3.3.4 Extended Spatial Panel Model

Until now, this study has introduced key variables of our specification 2 but some explanatory variables are still missed in the model. The following extended model includes a set of observable control variables ( $X_i$  in our specification 2), using sample B. To recall, sample B contains information on 85 regions<sup>39</sup> of Colombia, Ecuador and Mexico over the period 2001-2009.

In concordance with growth theory, we include the enrollment ratio in tertiary education, public investment and level of industrial specialization. The first variable measuring the mass of educated population could entail collinearity with urbanization since access to education would improve with urbanization. The willingness of families to invest in education would increase in urban areas where schools and universities facilities are located. Indeed, the correlation coefficient between education and urbanization is high and significant (0.6320). Thus, we might think that tertiary education encloses the effects of agglomeration that bears benefits for growth.

---

<sup>39</sup>Sample B initially has 87 regions but 2 island regions (Galapagos from Ecuador and San Andres from Colombia) and 1 outlier region (Campeche from Mexico) are eliminated.

In order to test this hypothesis, we conduct an estimation in two stages. In the first one, we regress education on the level of urbanization and in the second one, the resulting predicted values are introduced in the extended model. In Table 5 the estimations with both variables urbanization and tertiary education are presented in column 1 and the estimation in two stages is presented in column 2.

The urbanization rate is no longer significant after controlling for more variables and spatial autocorrelation (see column 1 of Table 5). By contrast, the second model shows that urbanization drives growth through education. This means that education constitutes a channel of agglomeration economies in these countries. The intuition is that urban areas host university facilities so population move there to invest in human capital. Thus, urbanization generates positive effects on education, which in turn, drives economic growth.

Moreover, the effects of educated labor force spill across regional borders. The mass of educated workers in one region impacts on its neighbors' growth. Those spatial effects represent the so-called knowledge spillovers. The intuition regarding this positive growth effect of education on neighboring regions is the following. When a new product, resulting from research and development investment, is created in one region, neighboring regions would have access to it thank to geographical proximity. Thus, the economic growth of neighboring regions is positively affected.

The public investment has a positive impact on economic growth of regions through two main aspects: the improvement of life quality of population and the development of better transport and communications infrastructure. Public investment in basic services, such as education and health improves living conditions which allow working-age population to access to the labor market. Specialized workers would increase productivity and economic growth. Moreover, investment in communication and transportation infrastructure induces higher growth by increasing productivity.

Regarding spatial effects of public investment, they are significant and account for almost 45% of the total effect. Given that the indivisible facilities funded by public investment cannot be used in more than one region, we presume that significant spatial effects come mainly from social investment such as education and health.

The concentration of production in few sectors seems to be detrimental for regional growth. The negative effect could be due to the nature of the sector of specialization. If the sector in which one region specializes does not entail potential technological progress, its effect on growth would be negligible. Besides, if the main sector is volatile, external shocks can significantly reduce its performance. In this

sample, many regions specialize in the primary sector, manufacturing and public services and very few regions specialize in service industry. Then, the negative average effect may come from the first three sectors. We presume that the primary sector is very likely to bear potential negative effects due to its high volatility. In order to test this hypothesis, we distinguish the effect of different sectors. According to the results shown in the Appendix C, the regional share in primary sector brings negative effects for growth. By contrast, the effect of the sector of construction is statistically positive.

By bringing together the effects of human capital and public investment, the speed of convergence increases. Now, the regions of the three countries converge at rates of 16% when using skilled labor force and 9% when using urbanization. The factor that could explain such a change is the presence of spatial effects between regions. As mentioned before, a region that spills over its knowledge will affect its neighbor's growth. Likewise, there exist positive spatial effects of public investment. Thus, regions would converge more rapidly to their steady states.

Table 5: Spatial Autoregressive Extended Models

	(1)	(2)
	Extended (W g)	Extended (W g)
<b>Direct</b>		
urb	0.343 (1.272)	
y_hat_educ		0.667 (4.725)***
initial GDPpc	-0.375 (-7.065)***	-0.323 (-7.120)***
[1em] inv_GDP	0.193 (3.368)**	0.216 (3.843)**
tertiary_educ	0.506 (3.831)**	
IER	-0.0453 (-3.120)**	-0.0489 (-3.499)**
<b>Spatial</b>		
$\rho$	0.304 (6.161)***	0.295 (5.918)***
<b>Variance</b>		
$\sigma_e^2$	0.00116 (11.077)***	0.00119 (11.083)***
<b>Indirect</b>		
urb	0.139 (1.165)	
y_hat_educ		0.263 (3.241)**
initial GDPpc	-0.152 (-3.736)**	-0.127 (-3.833)**
inv_GDP	0.0771 (2.799)**	0.0842 (3.180)**
tertiary_educ	0.205 (2.810)**	
IER	-0.0184 (-2.586)**	-0.0192 (-2.945)**
<b>Total</b>		
urb	0.482 (1.252)	
y_hat_educ		0.930 (4.568)***
initial GDPpc	-0.526 (-6.541)***	-0.450 (-6.617)***
inv_GDP	0.270 (3.412)**	0.300 (3.904)***
tertiary_educ	0.711 (3.712)**	
IER	-0.0637 (-3.095)**	-0.0681 (-3.498)**
N observations	252	252
N regions	84	84
Region specific effects	Y	Y
Time specific effects	Y	Y
$R^2$	0.692	0.691
Hausman test	26.78	20.01
p-value	0.000771	0.00556

*t* statistics in parentheses

\*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < 0.0001$

## 4 Conclusions

The current study responds to the claim of extending the knowledge base about the effects of agglomeration for developing countries by looking at sub-national regions in Latin America. Therein, we have been able to distinguish the specific effect of regions within countries instead of attributing a general effect to a whole country.

One of the conclusions is the clear evidence that agglomeration is of significant importance for the economic growth of Latin American regions. However, the effects vary across regions. As [Williamson \(1965\)](#) argued, the key element of the differences is the level of development. According to our results, at low stages of development less than 5,700 dollars per capita income, the effects of urbanization are magnified. Then, the effect of urbanization decreases until a threshold of 10,500 dollars per capita income, at which negative effects might appear. Notwithstanding, those negative effects are weak and not significant, suggesting that Latin American regions do not yet face strong negative effects.

Additionally, Latin American regions are spatially interdependent. The spatial effects have a large geographical scope. They are low in the immediate neighborhood and high at a larger distance. The intensity of spatial effects is high within countries which points out the strong border effects in Latin America. Lastly, we could identify a channel of agglomeration economies: education. Urbanization impacts growth through human capital. Besides, such an effect diffuses across space. Likewise, public investment produces positive direct and indirect spatial effects. Those elements together make regions to converge more rapidly to their steady states.

Careful interpretation of the results is requested since the temporal dimension of the sample is rather small. In addition, the definition of agglomeration using urbanization does not take into account the industrial composition of regional economies. The consideration of such an element is a complementary topic to be investigated. Lastly, regional integration defined by geographical distance is limited. Although spatial weight matrices based on distance illustrate general spatial configurations, they cannot shed light on the evolution of spatial patterns. Using spatial weight matrices based on trade flows or road density would provide promising results regarding the evolution of economic integration. However, using such matrices entails potential endogeneity in the model. This issue could be addressed following [Qu and Lee \(2015\)](#) who provide an adequate method for estimating spatial models with endogenous spatial weight matrices.

## References

- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer, Dordrecht.
- Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2):93–115.
- Arbia, G. and Piras, G. (2005). Convergence in per-capita GDP across european regions using panel data models extended to spatial autocorrelation effects. Technical report, Istituti di Studi e Analisi Economica, Working paper 51.
- Atienza, M. and Aroca, P. (2012). Es la concentración un problema para el crecimiento de América Latina? (In english: Is the concentration an issue for Latin American growth?). Technical report, Universidad Católica del Norte Chile.
- Baldwin, R. E. and Forslid, R. (2000). The core-periphery model and endogenous growth: Stabilizing and destabilizing integration. *Economica*, 67(267):307–324.
- Baltagi, B. (1995). *Econometric Analysis of Panel Data*. John Wiley Sons, 1st edition.
- Barro, J. R. and Sala-i Martin, X. (1992). Convergence. *Journal of Political Economy*, 100(2):223–251.
- Blanco, L. (2011). The impact of spatial interdependence on FDI in latin america. Technical report, Pepperdine University, School of Public Policy Working Papers, Paper 27.
- Bosker, M. (2007). Growth, agglomeration and convergence: A space-time analysis for european regions. *Spatial Economic Analysis*, 2(1):91–110.
- Brühlhart, M. and Sbergami, F. (2009). Agglomeration and growth: Cross-country evidence. *Journal of Urban Economics*, 65(1):48–63.
- Chauvin, J.-P., Glaeser, E., and Tobio, K. (2013). Urban economics in the US and India. Work in progress, Harvard University.
- Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review*, 46(2):213–227.

- Ciccone, A. and Hall, R. E. (1995). Productivity and the density of economic activity. Economics Working Paper 120, Department of Economics and Business, Universitat Pompeu Fabra.
- Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2008). Estimating agglomeration economies with history, geology and worker effects. Technical report, Série des Documents de Travail du CREST, Institut National de la Statistique et des Etudes Economiques.
- Combes, P.-P. and Lafourcade, M. (2012). Revue de la littérature académique quantifiant les effets d’agglomération sur la productivité et l’emploi. (In english: Literature review about agglomeration effects in productivity and employment.). Report, Societe de Grand Paris.
- Crozet, M. and Koenig, P. (2008). Arbitrage cohesion-croissance: Une analyse des régions européennes (1980-2000) (In english: The cohesion vs growth tradeoff: Evidence from EU regions). *Revue économique*, 59(2):331–340.
- Davis, J. C. and Henderson, J. V. (2003). Evidence on the political economy of the urbanization process. *Journal of Urban Economics*, 53(1):98–125.
- Duranton, G. (2007). From cities to productivity and growth in developing countries. Working Paper tecipa-306, University of Toronto, Department of Economics.
- Gardiner, B., Martin, R., and Tyler, P. (2010). Does spatial agglomeration increase national growth? some evidence from europe. *Journal of Economic Geography*, pages 1–28.
- Guevara, C. and Autant-Bernard, C. (2015). Technological interdependence between south american countries: A spatial panel data growth model. *Revue d’Economie Régionale et Urbaine*, 2(1):181–209.
- Guevara, C. and Jarrin, S. (2011). *Modelo Gravitacional de Comercio para America del Sur. (In english: Gravity Model for South America)*. Editorial Academica Espanola.
- Henderson, J. V. (2000). The effects of urban concentration on economic growth. Technical report, National Bureau of Economic Research Working Paper 7503.
- Henderson, V. (2003). Urbanization and economic development. *Annals of Economics and Finance*, 4(2):275–341.



- Martin, P. and Ottaviano, G. (1999). Growing locations: Industry location in a model of endogenous growth. *European Economic Review*, 43:281–302.
- Martin, P. and Ottaviano, G. I. P. (2001). Growth and agglomeration. *International Economic Review*, 42(4):947–968.
- Massiris-Cabeza, A., Espinoza-Rico, M. A., Ramirez-Castaneda, T., Rincón-Avellaneda, P., and Sanabria-Artunduaga, T. (2012). *Procesos de ordenamiento en América Latina y Colombia (In english: Planning processes in Latin America and Colombia)*. Universidad Nacional de Colombia.
- McCallum, J. (1995). National borders matter: Canada-U.S. regional trade patterns. *American Economic Association*, 85(3):615–623.
- Moran, P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, series B*, 37(1):243–251.
- Pesaran, H. M. (2004). General diagnostic tests for cross section dependence in panels. Technical report, University of Cambridge, Working Papers 0435.
- Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(1):209–232.
- Ramírez, M. T. and Loboguerrero, A. M. (2002). Spatial dependence and economic growth: Evidence from a panel of countries. Technical report, Banco de la República Colombia.
- Serra, M.-I., Pazmino, M.-F., Lindow, G., Sutton, B., and Ramirez, G. (2006). Regional convergence in Latin America. Technical report, International Monetary Fund.
- Venables, A. J. (2005). Spatial disparities in developing countries: cities, regions and international trade. *Journal of Economy Geography*, 5(1):3–21.
- Williamson, J. G. (1965). Regional inequality and the process of national development: A description of the patterns. *Economic Development & Cultural Change*, page 3.

# Appendices

## A Spatial dependence test

In Table 6, we present the test of spatial dependence for estimations shown in section 3.3.2. According to the Moran's I test, the null hypothesis of no spatial correlation is not rejected for most of periods  $t$ . However, when using the spatial weight matrix (W k1), spatial correlation remains in periods p1 and p3. In the case of matrices (W g) and (W beL), spatial correlation has been considerably reduced. There is absence of spatial dependence in the residuals after the estimations using the distance weight matrix (W d).

Table 6: Moran's I test for SAR models Sample A

	(W d)	(W k1)	(W g)	(W beH)	(W beL)
	I (p-value)	I (p-value)	I (p-value)	I (p-value)	I (p-value)
p1	-0.014 (0.337)	-0.195 (0.027)	-0.069 (0.183)	-0.023 (0.339)	-0.056 (0.248)
p2	-0.003 (0.421)	-0.078 (0.233)	-0.008 (0.488)	0.007 (0.369)	-0.005 (0.492)
p3	-0.016 (0.303)	-0.266 (0.004)	-0.175 ( 0.008)	-0.051 (0.132)	-0.175 (0.011)

## B Definition of the threshold of development

In order to determine the threshold of the level of development at which the effects of urbanization are maximized, we run some simulations. We split the database by income levels. For instance, at an income level of US\$4000, we keep regions under such a value and we estimate the model. We look at the parameter estimates and so on. According to the results, the maximum effects of urbanization on growth are reached between US\$5500 and US\$6000 of per capita income. Table 7 shows the results when setting the parameter of the level of development lower than US\$5000, lower than US\$5700, lower than \$6000 and lower than US\$7500. The maximum coefficient is reached at US\$5700 of per capita income. Likewise, we identify the threshold at which the coefficient estimate of urbanization becomes negative.

Table 7: Simulations, level of development threshold

	(1)	(2)	(3)	(4)	(5)
	< 5000	< 5700	< 6000	< 7500	> 10500
urb	0.640 (2.132)*	0.683 (2.365)*	0.671 (2.366)*	0.648 (2.438)*	-0.0471 (-0.084)
initial GDPpc	-0.170 (-3.091)**	-0.174 (-3.274)**	-0.171 (-3.261)**	-0.161 (-3.255)**	-0.280 (-5.715)**
yearcris01	-0.176 (-7.804)**	-0.169 (-7.815)**	-0.169 (-7.849)**	-0.160 (-8.106)**	-0.107 (-4.409)**
yearcris09	-0.0915 (-7.451)**	-0.0921 (-8.187)**	-0.0937 (-8.664)**	-0.0979 (-10.308)**	-0.1000 (-7.409)**
Constant	1.074 (2.960)**	1.087 (3.111)**	1.070 (3.124)**	1.001 (3.035)**	2.794 (3.674)**
N observations	270	288	300	354	75
N regions	90	96	100	118	25
F	41.31	45.24	48.47	73.82	94.00
p-value F	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$R^2$	0.529	0.521	0.522	0.525	0.776

$t$  statistics in parentheses

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < 0.0001$

## C Specialization in different sectors

In Table 5, the effect of industrial specialization is negative. In table 8, we estimate the specific effect of each sector on growth. The results show that specialized regions in primary sector have a significant and negative effect on their growth. On the contrary, specialization in construction brings positive effects on growth which indicates that investment in capital is beneficial. The coefficient estimates of other sectors are not significant at 5% level.

Table 8: The effect of specialization in different sectors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Primary	Manuf.	Serv.	Pub.Adm.	Retail	Elect. water	Constr.
urb_p	0.584 (1.085)	0.628 (1.215)	0.635 (1.237)	0.589 (1.075)	0.538 (0.953)	0.654 (1.249)	0.790 (1.668)*
Primary	-0.0302 (-1.692)*						
Manufacturing		0.0200 (0.465)					
Services			-0.0200 (-0.299)				
Public Adm.				0.0244 (0.655)			
Retail					-0.0480 (-0.747)		
Elect, water supply						-0.0157 (-1.278)	
Construction							0.0499 (1.738)*
initial GDPpc	-0.469 (-3.459)**	-0.486 (-3.498)**	-0.487 (-3.446)**	-0.481 (-3.551)**	-0.490 (-3.482)**	-0.494 (-3.538)**	-0.485 (-3.832)**
inv_GDP	0.409 (2.508)**	0.408 (2.438)**	0.412 (2.400)**	0.427 (2.305)**	0.443 (2.255)**	0.405 (2.424)**	0.398 (2.949)**
tertiary educ	0.716 (2.526)**	0.724 (2.525)**	0.721 (2.535)**	0.715 (2.542)**	0.732 (2.553)**	0.731 (2.546)**	0.683 (2.973)**
$R^2$	0.689	0.683	0.683	0.684	0.685	0.686	0.705

*t* statistics in parentheses

\*  $p < .1$ , \*\*  $p < .05$ , \*\*\*  $p < 0.0001$