ARTÍCULO

THE SPATIAL AGGLOMERATION OF EDUCATED PEOPLE IN COLOMBIA

Juan Tomás Sayago Gómez

Sayago Gómez, J. T. (2014). The spatial agglomeration of educated people in Colombia. *Cuadernos de Economía*, 33(62), 297-317.

Over the past few decades, Colombia's education system has been growing in terms of access and coverage. However, this development has taken place mainly in bigger cities, and displays an agglomeration of graduates in tertiary education. The purpose of this article is to test this hypothesis of agglomeration and attempt to find out which factors are associated to this phenomena: quality of life, the effects of income, political safety, and supply of education. Using empirical evidence derived from real data obtained by DANE in the 2005 Census and a variable of violence from the IEPRI, spatial econometric models are set out to understand its dynamics, to stop and reverse this agglomeration, and to create benefits for smaller municipalities.

Keywords: Education, education maps, spatial analysis, Moran's index, spatial econometric models.
JEL: I2, I29, C21.

J. T. Sayago Gómez
West Virginia University and Regional Research Institute. E-mail: jsayago@mix.wvu.edu.

I want to thank comments from Roger Bivand, Roberta Capello, Hernan Enriquez, Jorge Aponte, two anonymous referees and the assistants to the RSAI of Americas conference and the assistants to the 3rd conference of the World Spatial Econometrics Association. All remaining errors are my own.

Este artículo fue recibido el 30 de junio de 2012, ajustado el 4 de septiembre de 2013 y su publicación aprobada el 16 de septiembre de 2013.

Sayago Gómez, J. T. (2014). La aglomeración espacial de la población educada en Colombia. *Cuadernos de Economía*, 33(62), 297-317.

El sistema educativo colombiano ha crecido en las últimas décadas en acceso y cobertura, pero este desarrollo se ha centrado en las principales ciudades y se presenta una aglomeración en estas de graduados en educación terciaria. Este artículo pretende comprobar esta hipótesis y examinar cuáles factores están relacionados con este fenómeno: calidad de vida, ingreso, seguridad política y oferta educativa. Utilizando información del DANE y del Iepri se plantean modelos de econometría espacial para aproximarse al problema, entender sus causas y formular políticas que puedan contener esta dinámica y generar beneficios para las pequeñas poblaciones en el país.

Palabras clave: educación, mapas de educación, análisis espacial, índice de Moran, modelos de dependencia espacial.
JEL: I2, I29, C21.

Sayago Gómez, J. T. (2014). La concentration de la population éduquée en Colombie. *Cuadernos de Economía*, 33(62), 297-317.

Le système éducatif colombien s'est développé au cours des dernières dizaines d'années en possibilités d'accès et en étendue, mais ce développement s'est centré sur les villes principales dans lesquelles on constate une concentration de ces diplômés en éducation tertiaire. Cet article cherche à vérifier cette hypothèse et à examiner les facteurs liés à ce phénomène : qualité de vie, revenus, sécurité politique, et offre éducative. En utilisant l'information du DANE [Département Administratif National de Statistique] et de l'IEPRI [Institut d'Études Politiques et Relations Internationales] sont proposés des modèles d'économétrie spatiale pour approcher le problème, comprendre ses causes et formuler des politiques qui puissent contenir cette dynamique et générer des bénéfices pour les petites villes du pays.

Mots-clés : éducation, cartes de l'éducation, analyse spatiale, indice de Moran, modèles de dépendance spatiale.
JEL : I2, I29, C21.

Sayago Gómez, J. T. (2014). A aglomeração espacial da população educada na Colômbia. *Cuadernos de Economía*, 33(62), 297-317.

O sistema educativo colombiano tem crescido nas últimas décadas em acesso e cobertura, mas este desenvolvimento se concentrou nas principais cidades; nestas, há uma grande quantidade de formados no ensino terciário. Este artigo pretende comprovar esta hipótese e examinar que fatores estão relacionados com este fenômeno: Qualidade de vida, renda, segurança política e oferta educativa. Utilizando informações do DANE e do IEPRI, são propostos modelos de econometria espacial para se aproximar do problema, entender as suas causas e formular políticas que possam conter esta dinâmica e gerar benefícios para as pequenas populações no país.

Palavras-chave: Educação, mapas de educação, análise espacial, índice de Moran, modelos de dependência espacial.
JEL: I2, I29, C21.

"There are ten policemen for each student and one student for one thousand ignorants."[1] Over the past forty years in Colombia illiteracy rates have been reduced and access to education has increased. This constitutes a major breakthrough, but despite these great accomplishments and advancements, clusters of educated people are developing in Colombia's cities. Tertiary education graduates decide to remain in big cities even though the access to education, health and quality of life are improving in smaller cities. However, these graduates do not want to leave the big cities or return to their cities of origin for a number of reasons: there may not be enough opportunities to get a well paying job according to their skills, it does not seem safe enough, or conditions and better opportunities for their children are perceived to be insufficient. Thus, small cities have fewer educated people to implement new technologies or projects that could increase a city's opportunity to develop.

In order to compute and observe these clusters of highly educated people, data from the 2005 CENSUS undertaken by DANE[2] will be used to compute the rate of educated people residing in every municipality that data is available for[3]**.** Furthermore, the information will show that there is another cluster related to lower rates of graduates in tertiary education.

The factors to be considered are political, social, and economic: violence, a lack of public services, and poor career opportunities in relation to where graduates decide to live. The relationship between these concepts will be explored and verified in order to see whether there is a statistical relationship between them.

The two main theses of this paper follow, first, a theoretical assumption on increasing returns in endogenous economic growth**,** observed as the agglomeration of skilled people in the main cities (Easterly, 2002). And secondly, a Harris and Todaro approach regarding the concept of people who tend to migrate based on their income expectations (Harris & Todaro, 1970).

This paper is divided into 5 further sections. The second section reviews the literature regarding education and the spatial analysis in Colombia. The third section examines education in Colombian municipalities to test the agglomeration hypothesis. The fourth section analyzes the factors related to this agglomeration. The last section concludes the paper.

ADVANCES OVER THE LAST YEARS

Although the body of education-related literature has grown over the past 15 years, the analysis has been restricted to case studies, reviews of the development, and policy recommendations. However, in recent years, authors have tried to imple-

[1] Popular protest song from the 1970's in Colombia. Translation by the author.

[2] DANE: Departamento Administrativo Nacional de Estadística. National Administrative Department of Statistics.

[3] The methods and the data used to calculate these rates are described in appendix A.

ment new ways to analyze the education sector, as will be shown in this section with an overview of recently published documents regarding the education sector in Colombia.

The implementation of spatial analysis has only been applied once for the education sector. Jaime Bonet georeferenced data related to the provision of education in the departments, finding different patterns of agglomeration, which highlight regional inequality. The author concluded that the allocation of education in Colombia is distributed in such a way that the departments with poor provision are surrounded by departments that offer the same conditions and the ones with higher levels are surrounded by prosperous ones also (Bonet, 2005, p.36)[4]. Bonet also highlights the importance of using information by municipalities.

Galvis and Bonilla find evidence of this regional inequality in the education level of teachers in municipalities and types of schools (Galvis & Bonilla, 2012). Their analysis focused on distinguishing the factors related to this inequality by municipality and by school.

Mora and Muro measure the wage gains associated to having a secondary and university degree in Colombia (Mora & Muro, 2008) . Their approach is applied for the main cities. The authors used a pseudo panel and found a statistically significant increase in the hourly wages for those with a secondary education and a university diploma, confirming the presence of a sheepskin effect.

Another article with a revealing conclusion related to tertiary education was written by Mora and Ceballos, who found that the topics taught in technical and technological education are not adequate to fit the needs of the region where they are taught (Mora & Ceballos, 2006). The authors consider this one of the main difficulties for some people to remain in or go back to their cities of origin. It could be said that this education is not accomplishing its raison d'etre.

Regarding other developments in the education sector, it is important to review the works presented by Ramírez and Téllez and Ramírez and Salazar who review the history of Colombia's education sector. They provide a good historical account of education in the 20th Century underlining the importance of different policies and their effects (Ramírez & Tellez, 2006, Ramírez & Salazar, 2007). It is important to highlight two of the facts that they present: first, the recent and fast development in the sector that has taken place since the 1950s at "an unseen velocity", and even after one slow period, a subsequent new expansion; and secondly, that from the end of the 19th Century the sector has been showing other important advances.

The development of the education sector in Colombia can be said to reflect the findings of Ramirez and Tellez and is explained by an increase in the access to education also highlighted by Iregui, Melo and Ramos. The authors explain that the main problem concerning education in Colombia is related more to efficiency and quality, these being the areas which have not shown constant growth (Iregui, Melo

[4] Author's translation.

& Ramos, 2006). The same authors have published another article also with an analysis of the efficiency frontier in Colombian schools. Using information from the ICFES[5] which analyses the level of the schools with different factors, they conclude that "private schools are more efficient due to a more favorable environment" (Iregui, Melo & Ramos, 2007, p.21).

It is also worthwhile to review the results of the 2005 CENSUS published by DANE, which show a stable level of literacy in Colombia. It must be said that the percentage of people attending school has increased greatly, and illiteracy rates have lowered over the last 40 years.

EDUCATION AGGLOMERATED IN COLOMBIAN CITIES

The number of educated people in Colombia has increased greatly over the past 40 years, when it has been easier to access education, and special financial aid programs have been created for students with low or no income to pay for this education. Both of these factors have incresed the rates of education. Also, new universities and technical training centers have been created all over the country, raising the number of people educated at tertiary level. However, the proportion of tertiary level graduates is still low in almost all municipalities.

The number of graduates in tertiary education is calculated as the percentage of people with superior education (see equation (6) in the appendix). This rate appears to be skewed to the left side (close to lower levels) as can be seen in Figure 1 (Kernel density plot). The higher rates belong to the more populated cities, for which the mean rate is 15.7%[6]. But because this kernel is made using the municipalities' rates of tertiary graduates as the unit of measure, this gives way to the observed skewness which is very similar to Quah's convergence clubs (Quah, 1997).

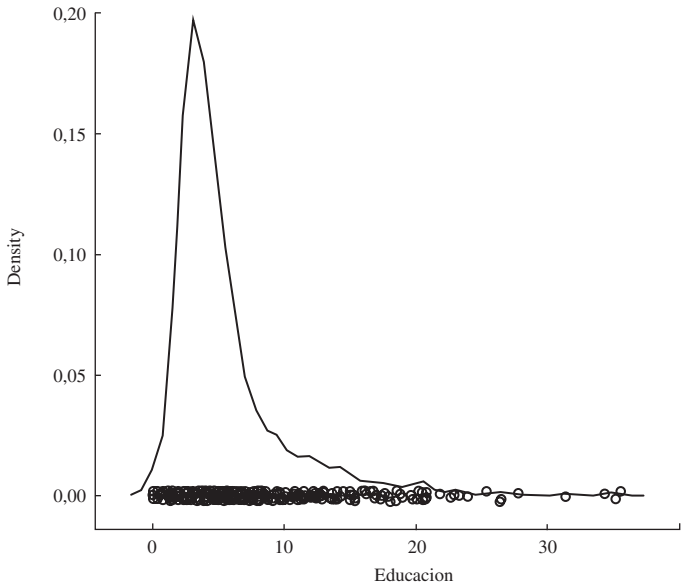
The existence of a spatial relationship between the observations and their neighbors has been widely specified as the first law of geography "Everything is related to everything else, but near things are more related than distant things" stated by Waldo Tobler (1970).

The concentration of educated people can be explained by economic activities and short-distance migration phenomena, like commuting. Another explanation is related to the access to places where tertiary education is required by economic activities, such as the industrial and services sectors. The supply of this education is very limited in Colombia, and can cause the agglomeration of educated people in large Colombian cities.

[5] ICFES is a standardized test used for university applications

[6] Calculations undertaken using data from CENSO 2005 and it is important to highlight that the data is population based and not by municipality.

FIGURE 1.
DENSITY KERNEL OF THE PERCENTAGES OF TERTIARY EDUCATION GRADUATES IN COLOMBIA BY MUNICIPALITIES



Source: Own calculations with data from DANE (2005).

The spatial concentration of people with tertiary education can be seen in Figure 2. There appears to be a higher percentage of educated people in the municipalities located in the most important cities in the country. Furthermore, these cities are the richest[7] or have the best opportunities in schooling and public services for their inhabitants. Also, political conflict appears to affect these places less than smaller municipalities in the country.

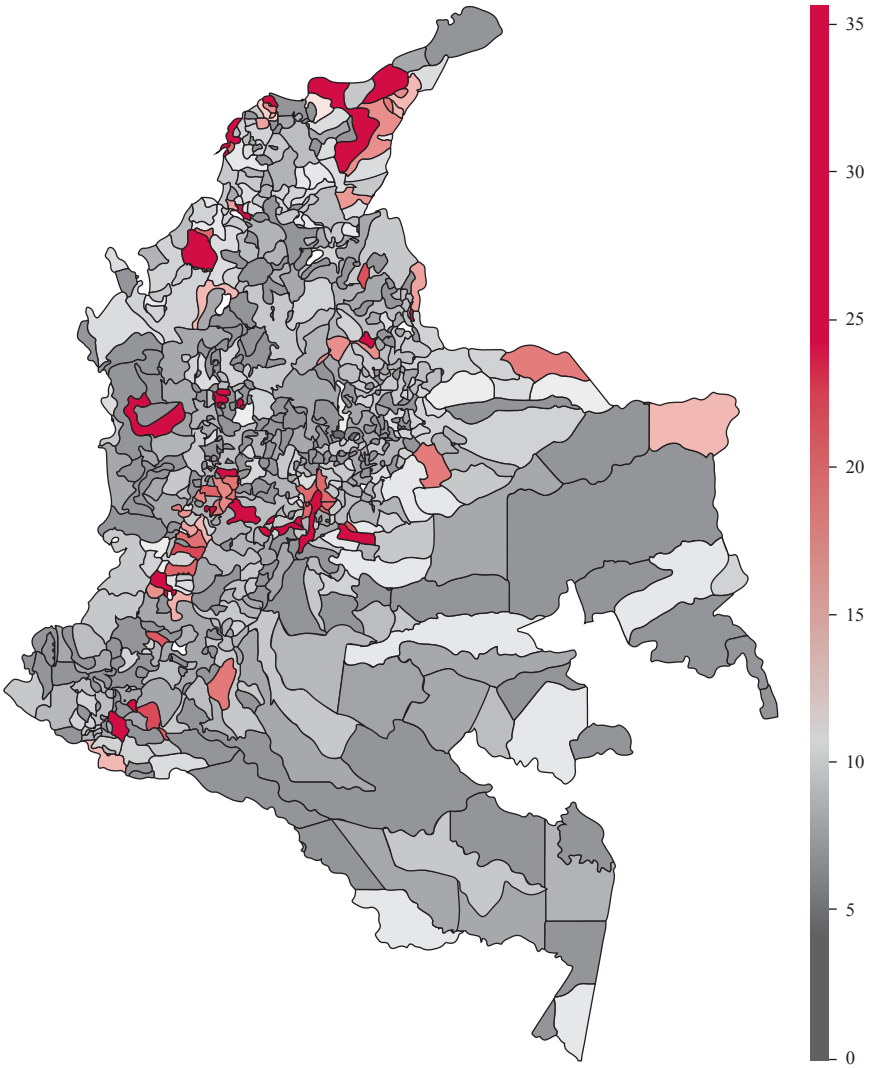
The different tests to check spatial concentration such as Moran's Index, Moran's scatterplot, and the Local Indicators for Spatial Association (LISA) will be applied to test this hypothesis, Anselin (1993, 1995, 1996). The Moran Index is a regression of the rate of educated people in the municipalities against the rate of the neighboring municipalities. A positive coefficient will show that the municipalities are located close to other municipalities in similar conditions. The Moran scatter plot will arrange the features by the municipalities against their neighbors to classify their type of relationship[8]. Based on this type of analysis, LISA maps are

[7] These cities are classified as categories special and 1.

[8] The types of relationships are: high-high, low-low, high-low and low-high. Where the first argument is the argument of the municipality that is examined and the second argument describes the municipality's neighbors, as their average.

developed to classify the municipalities and test the spatial concentration statistically. This is important because this analysis allows the graphic display of results in a map.

FIGURE 2.
MAP OF THE PERCENTAGE OF TERTIARY EDUCATION GRADUATES IN COLOMBIA BY MUNICIPALITIES



Source: DANE (2005).

TABLE 1.
MUNICIPAL CATEGORIES

Categories	Municipalities	Percentage
1	17	1.5%
2	16	1.4%
3	19	1.7%
4	20	1.8%
5	31	2.8%
6	991	88.6%
E(S)	5	0.4%
NC	20	1.8%
TOTAL	1,119	100%

Source: DANE (2005)

The agglomeration of tertiary education graduates in Colombia, as stated earlier, can be seen in Figure 2. The location of the bigger cities and their surrounding cities seem to have more graduates than other areas in the country. To analyze this aggolmeration, it is important to analyze whether there is spatial autocorrelation in the proportion of graduates in the municipalities in Colombia, and to be able to highlight the presence of this phenomenon. It is also important to see under what conditions the phenomenon is present.

The results from the Moran Index presented in Table 2 show us that there is strong spatial correlation with first order neighbors[9]. The results observed on the Moran Scatterplot (Figure 3) indicate a cluster of low-low values and the center core to be very concentrated by the nature of the index and the values displayed.

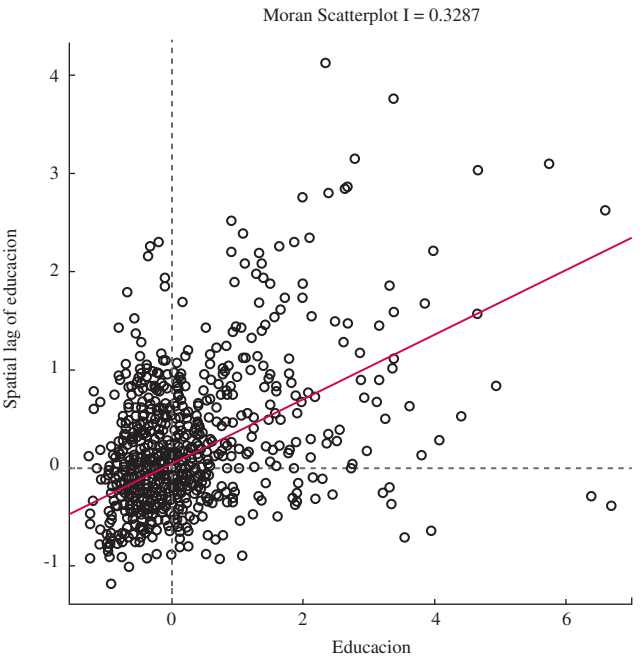
TABLE 2.
MONTECARLO SIMULATIONS FOR THE MORAN'S INDEX

Moran's I	0.3293
p-value	0.000
Permutations	1000

Source: Own calculations with data from DANE (2005)

[9] The use of first order contiguity is not the best option for this analysis because it cannot distinguish when municipalities are close but don't have any contact, but in plain regions it can work well.

FIGURE 3.
MORAN SCATTERPLOT OF TERTIARY EDUCATION GRADUATES IN COLOMBIA



Source: DANE (2005).

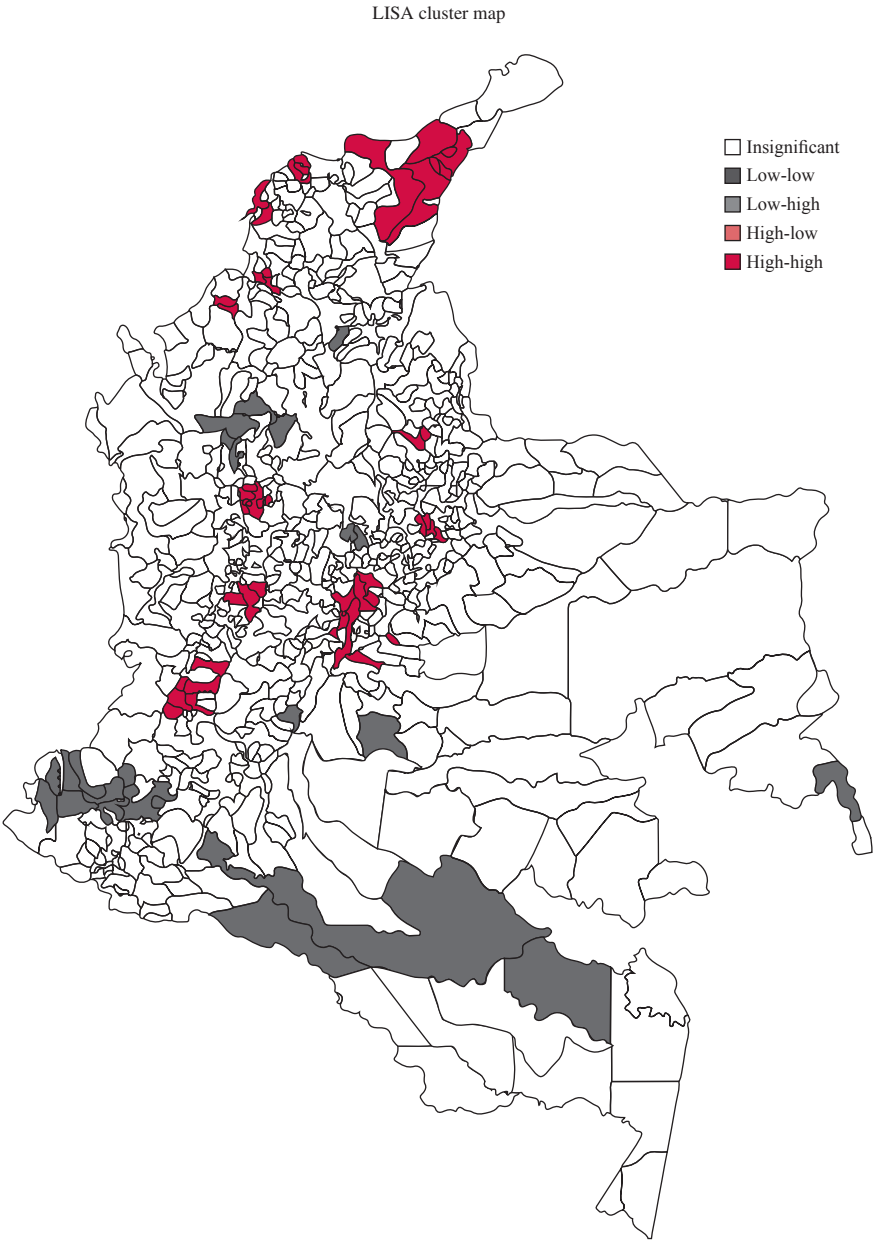
The LISA in Figure 4 show the presence of small clusters around the main capitals, specifically the ones with categories special and 1[10]. It also highlights the presence of low-high rate clusters, which confirms that the municipalities farther from these capitals are the least attractive for these graduates.

From this analysis a cluster of highly educated individuals is found in Antioquia near Medellin. In the **Eastern** region, high-high clusters can be found near Tunja, Bogota, and Bucaramanga. In the pacific region, one cluster is found in the Valle del Cauca Department around Cali. From the Atlantic region high-high clusters are observed near the main capitals in the coastal area around the main cities of the region. Finally, for the Amazon and Orinoquia region a small low-high cluster is observed, but it is also debatable due to transportation difficulties within the region.

In the next section the factors associated with this phenomenon are examined in greater detail.

[10]Municipal categories created by DANE and DNP. See Table 1.

FIGURE 4.
LISA MAP OF TERTIARY EDUCATION GRADUATES IN COLOMBIA BY MUNICIPALITIES.
ON THE LEFT THE LISA FOR THE QUEEN MATRIX LEVEL 1 AND ON THE RIGHT, LEVEL 2. THIS APPLIES TO ALL THE LISA ANALYSIS



Source: DANE (2005).

FACTORS EXPLAINING THE AGGLOMERATION OF EDUCATED PEOPLE IN COLOMBIA

As mentioned earlier, the factors possibly associated with the agglomeration of educated people in Colombia can be social, economic, and political. Economic conditions matter when it comes to choosing where to live, but also it is the social and political factors that affect that choice for tertiary graduates, as these factors also affect how desirable a municipality is. The economic variables used in the model account for the demand for educated people or the labor market space where they can enter the job market.

In this way, a proxy of the number of companies from each sector weighted by the population in the municipality will be used as an economic measure and then to test if the agglomeration is based mainly on a sectoral development. This proxy is calculated for agricultural, commercial and industrial sectors, it does not include governmental activities. More explicitly, one of the factors that affect this choice is the fact that the industrial sector in Colombia is not advanced and the qualifications required to enter it are low. On the other hand, the education and training of graduates, as stated by Mora and Ceballos, does not focus on the needs of the region (Mora & Ceballos, 2006).

The presence of sectoral firms weighted by population is proxy for an economic attractor because of job availability, in this sense a municipality with many firms per population will have more work available, than another with less firms. Although it is not possible to capture firm size or job requirements, the number of firms will be a proxy of the number of work positions, because although tertiary education graduates are also likely to set up firms, they will be hired workers[11]. The reverse causation could happen when an agglomeration is already formed and more jobs will be created because of residing tertiary education graduates.

Other proxy variables such as the Gini coefficient could not be used because the inequality measured by the index cannot capture the economic attractiveness of the municipalities, and variables such as unemployment by sectors could be used as factors but they do not exist. The perfect variable for a municipality's economic appeal could be the employment by sector or by education level; however for estimation purposes the variable would reveal simultaneity issues and measurement problems given by frictional unemployment.

Social related factors such as access to public services, can be related to economic variables. However, this sector in Colombia is not particularly correlated to economic development because of the inefficiency of public policy and the corruption of government employees and politicians who might have other interests. This variable is mostly related to effective social planning and the presence of public service amenities in cities. Deficiencies in these areas are mainly related to access to

[11]It is a case of the chicken and the egg, but the literature has proven that jobs attract people and then large concentrations of people will attract more jobs.

the water supply, and similar social indicators are correlated to it. Therefore, the access to the water supply variable will be used to account for social factors and institutions, and to consider how attractive the city is.

Political factors must be considered when analyzing Colombia's development given that the conflict and displacement problem that has plagued Colombia in the 1990s and 2000s gave way to many changes in terms of the structure of cities. The history and consequences related to the internal conflict can be read in Sanchez, Peñaranda, and Hurtado (2007). The proxy used to measure this is derived from an estimation of political homicide made by Gutierrez, because tertiary education graduates will prefer safer locations to live in and possibly work in (Gutierrez, 2006).

The supply of tertiary education is another important factor given that education at this level is only available in about 80 of the country's municipalities. The factor defines accessibility to tertiary education, and the locations where tertiary educated populations form will establish the point of origin of future tertiary education graduates.

The model was set up to include the factors that could attract educated people. The model estimated from equation (1) is used as a base to test for spatial correlation, as the results from the previous section highlight as neccessary:

$$PTG_i = \beta_0 + \beta_1 Economic_i + \beta_2 Social_i + \beta_3 Political_i + \beta_4 Supplyeduc_i + \varepsilon_i \quad (1)$$

The variables included are[12]:

- PTG_i : The Percentage of University graduates for city *i*.
- $Economic_i$: The variable considered for economic purposes is the proportion of economic activity, in thousands of people for city *i* and has been defined as *Industry* for the industrial sector, *Commerce* for the commercial sector, and *Service* for the service sector.
- $Social_i$: The variable *Wsupply*, considers the coverage of water supply in the city *i*.
- $Political_i$: *Violence* is the variable accounting for the number of political homicides for the 2000 to 2005 period, divided by the population in the municipality and multiplied by 1000[13].
- $Supplyeduc_i$: The variable *Universities* accounts for the number of universities in the municipality.

[12]Variables are more explicitly described in the appendix.

[13]I am thankful to Francisco Gutierrez for allowing the use of this information. For more details see the appendix 6.

The OLS model obtains the expected values in terms of parameters for the model since water coverage has a positive coefficient and violence a negative one. The coefficient obtained by economic variables can be explained by the type of workers required by the sectors included in the analysis. The university programs with the highest number of graduates in Colombia are related mainly to the services sector and not so much to the industrial sector. Also, the commercial sector does not require any type of special training, so it is a measure of economic activity that handles less technology. The residuals of the model in Figure 5 exhibit concentration of similarities which means that OLS is not the right way to estimate the model.

The tests also show that it is important to include a spatial parameter from the results of the Moran Test applied to the residuals in the OLS model. The results for the analysis using the spatial lag of the variable and the error lag are presented in Table 4, where we can see that the Lagrange multiplier test shows that either the error model or the lag model would be significant. These results can support the use of the SAR, SEM or SARAR models[14]. It must be highlighted that all the variables considered are not significant but, nevertheless, they should not be omitted from the model.

TABLE 3.
OLS ANALYSIS TO THE MODEL

Variable	Estimate (Std. Error)
Intercept	0.287416
	(0.350685)
WSupply	7.131214**
	(0.517399)
Violence	-0.894425**
	(0.212092)
Industry	-0.009368
	(0.018421)
Comerce	-0.054579**
	(0.017408)
Services	0.193980**
	(0.031342)
Universities	0.321844**
	(0.032496)
R-squared:	0.3167

** means significant at 5%

Source: Own calculations with data from DANE (2005) and Gutierrez (2006).

[14]The model selection of the appropriate model should be theoretically justified and explained. For a complete review on the methods applied see Anselin (1988); Kelejian & Prucha (1998, 1999); Arbia (2006); LeSage & Pace (2009).

TABLE 4.
MORAN'S I TEST TO THE RESIDUALS AND LM TESTS FOR ERROR AND LAG MODELS

Test	Statistic	p-value
Moran's I	0.29552	0.0000
LMerr	264.7295	0.00000
LMlag	210.0664	0.0000
RLMerr	58.618	0.0000
RLMlag	3.9557	0.004
SARMA	268.6852	0.0000

Source: Own calculations with data from DANE(2005) and Gutierrez (2006).

Anselin stated that standard econometrics analysis cannot be used with spatially correlated data (Anselin, 1988). I use the SARAR model because of the advantages from including a lag variable, which can account for the access to a commuting network for educated people, and the error lag is included to account for spatially correlated missing variables in the model, such as the wage differentials that can attract more people and the possibilities of obtaining a better quality of life.

Equations (2) and (3) show the spatial model that accounts for the Spatial Autorregresive with spatially Autoregressive Errors (also know as SARAR or SAC or the Kelejian-Prucha model), in the model the parameter ρ captures the effect of the spatial lag and the λ represents the spatial error component captured by the model:

$$Y_i = \rho WY_i + \beta X_i + u_i \tag{2}$$

where

$$u_i = \lambda Wu_i + \varepsilon_i \tag{3}$$

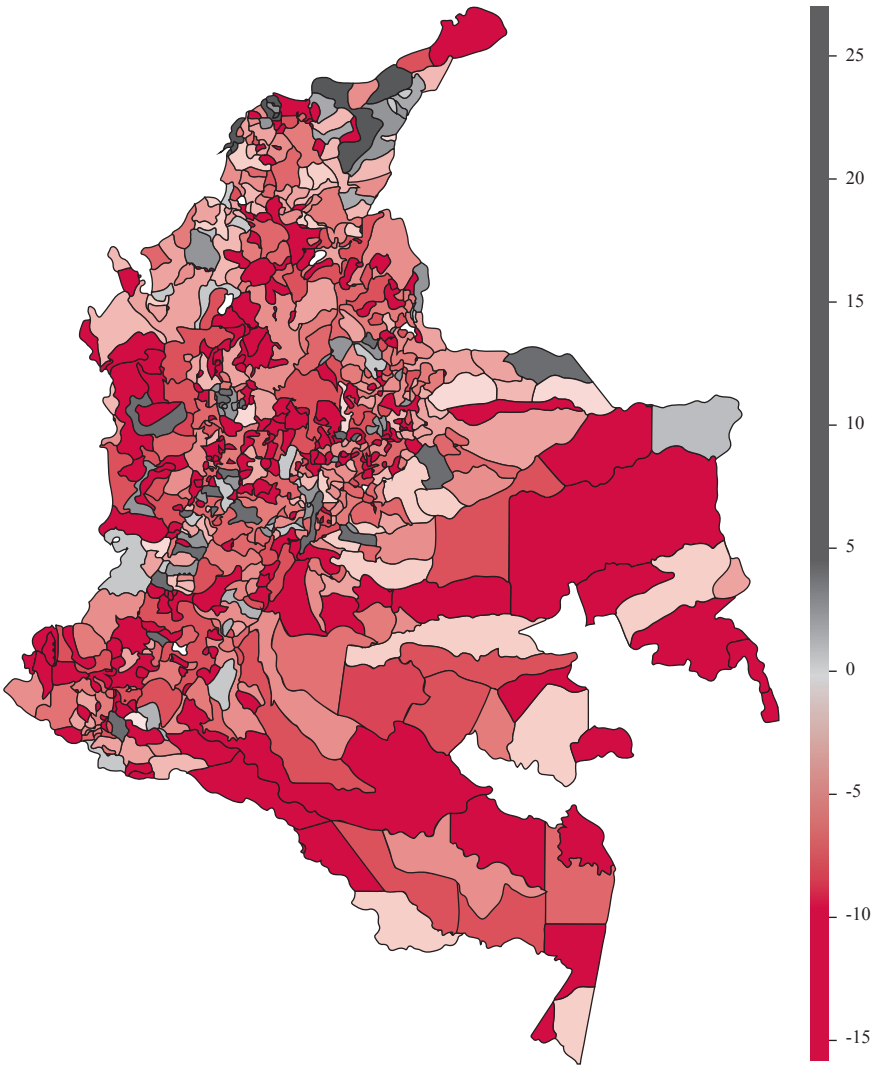
The model's equations for the tertiary education applied in the analysis is presented in equations (4) and (5):

$$PTG_i = \rho WPTG_i + \beta_0 + \beta_1 Economic_i + \beta_2 Social_i + \beta_3 Political_i + \beta_4 Supplyeduc_i + u_i \tag{4}$$

where

$$u_i = \lambda Wu_i + \varepsilon \tag{5}$$

FIGURE 5.
MAP OF THE RESIDUALS FROM THE OLS MODEL



Source: Own calculations with data from DANE (2005) and Gutierrez (2006).

The results from the model in 5 show the expected correlation between the variables included in the model as well as positive spatial parameters, but the interpretation is not as straightforward as a linear regression where the parameters represent the marginal effect of the variable. Lesage and Pace explained that results for a model with spatial lag of the dependent variable must include the effects estima-

ted to understand the parameters correctly (LeSage & Pace, 2009). The effects estimated in Table 6 are more consistent with the expected results; the quality of life in the municipality has a positive direct effect on the municipality and a negative indirect effect, and it must be noted that the presence of the industrial sector has a lower effect but is still positive, and is not significant. While the services sector and university presence obtain a higher and positive direct effect, the indirect effect is negative, which means that the commuting hypothesis is only supported in cases where the violence variable is high enough.

TABLE 5.
SARAR MODEL FOR PTG_i IN COLOMBIA. THE VALUE ON TOP IS THE COEFFICIENT AND THE VALUE IN PARENTHESIS CORRESPONDS TO THE THE STANDARD ERROR

Variable	Coefficients
(Intercept)	–0.9870
	(0.6786)
Wsupply	6.9585**
	(0.6342)
Violence	–0.7703**
	(0.1720)
Industry	–0.0017
	(0.0140)
Comerce	–0.0519**
	(0.0186)
Services	0.2047**
	(0.0297)
Universities	0.2735**
	(0.0938)
ρ	0.1975*
	(0.1098)
λ	0.3880**
	(0.1040)

** means significant at 5 % and * means significant at 10 %.
Source: Own calculations with data from DANE (2005) and Gutierrez (2006).

This identifies two more issues to be addressed: firstly, Colombian industries do not require advanced human capital or skilled workers, and secondly, as men-

tioned by Mora and Ceballos, tertiary education in Colombia does not focus on regions' needs (Mora & Ceballos, 2006).

The model with cross section data estimated the correlations specified from theoretically proposed equation. Since this model is limited to only having one moment, it shows correlations and highlights statistical regularities between observations but without other observations in time the model cannot explain or proove causality. Instead a correctly specified model can shed light on the correlates that are related or are proxy to what explains the behavior observed, and this result can improve our understanding of why certain processes occur in society.

TABLE 6.
DIRECT, INDIRECT, AND TOTAL EFFECTS OF THE SARAR MODEL.
P-VALUES ARE REPORTED IN PARENTHESIS

Variable	Direct	Indirect	Total
Wsupply	7.009006539	1.658670399	8.667676938
	(2.22*E* – 16)	(0.097345)	(1.7282*E* – 10)
Violence	-0.775866545	-0.183607601	-0.959474146
	(1.4549*E* – 06)	(0.102987)	(0.000011154)
Industry	-0.001689042	-0.000399709	-0.002088752
	(0.9685606)	(0.948947)	(0.96321)
Comerce	-0.052297497	-0.012376121	-0.064673617
	(0.0013197)	(0.148299)	(0.004014)
Services	0.206172125	0.04879031	0.254962435
	(3.0953*E* – 13)	(0.10849)	(1.0345*E* – 06)
Universities	0.275470771	0.065189726	0.340660497
	(0.00562)	(0.204256)	(0.019596)

Source: Own calculations with data from DANE (2005) and Gutierrez (2006).

CONCLUSIONS

A review of the data proposed in this article should demonstrate the importance of the agglomeration of educated people as a significant determining factor in the development of Colombian cities and the regions in which they are located. It is important not only to consider the distance between the different municipalities, but also the specific location to see the real spatial effect and the real geographic association. The first conclusion to be accounted for confirms findings put forward by Bonet: if you have a prosperous neighbor you will be wealthy or at least more developed, and if you have poor neighbors, it is probable that you will be poor, and the same goes for education (Bonet, 2005).

It is important to consider welfare relationships through access to public services in order to draw conclusions because these relationships have been vital in creating agglomerations. Furthermore, this access is also agglomerated and therefore access to the water supply as a necessity to live is crucial for this case.

Violence variables must be accounted for as well. Regardless of other factors, people tend to be strongly discouraged from moving into or staying in municipalities known for being violent. It is important to consider that violent experiences tend to drive people away from where that violence occurs. For future studies, it may perhaps be significant to use a proxy that accounts for, in greater historical depth, violence variables, accounting for people's memory.

Economic variables are indeed related. However, it can be argued that a two-way and mutually reinforcing relationship exists between educated people and developed or wealthy areas. On the one hand, educated people seek better wages, typically found in wealthy areas, while at the same time it is educated people stimulate economic growth and development[15]. This relationship may cast doubts on the results obtained in this paper, but should be considered in a comprehensive evaluation of these phenomena. Although, as stated by Easterly, "The magnitude of the relationship between initial schooling and subsequent growth is more consistent with the story of growth causing schooling rather than schooling causing growth" (Easterly, 2002).

Since spatial factors are considered, this type of analysis gives results that the same information and standard econometrics cannot. The spatial association of the data can be seen as Tobler stated in the first law of geography. Furthermore, availability of tertiary education in the bigger cities must also be considered as an important factor analyzing this agglomeration. It should be clarified that not only are the rates agglomerated, but also the population in the cities, as was observed by Galvis in "Economic Topography in Colombia" (Galvis, 2001).

It is also necessary to identify that this analysis was undertaken using the municipalities as observation units. This result keeps hidden any heterogeneity within the municipality, which is also very important in urban areas not visible in this study.

Colombia is no longer the country of one student for one thousand ignorants, at least for its larger cities. However, in order to progress, this has to happen countrywide. Therefore it is not only a problem of quality.

[15]This relationship is considered a stylized fact in economic growth.

REFERENCES

1. Anselin, L. (1988). *Spatial econometrics*. Kluwer: 32 Norwell, MA.
2. Anselin, L. (1993). Exploratory spatial data analysis and geographic information systems (Research Paper 9329). Regional Research Institute West Virginia University.
3. Anselin, L. (1995). Local indicators of spatial association-lisa. *Geographical Analysis, 27*(2), 93-115.
4. Anselin, L. (1996). The moran scatterplot as an esda tool to assess local instability in spatial association. In M. Fischer & J. Scholtenand (Eds.), *Spatial analytical perspectives on GIS* (pp. 111-125). London: Taylor and Francis.
5. Arbia, G. (2006*). Spatial econometrics: Statistical foundations and applications to regional convergence.* Berlin: Springer-Verlag.
6. Bonet, J. (2005). Inequidad espacial en la dotación educativa regional en Colombia (Documentos de Trabajo sobre Economía Regional 56). Banco de la República-Economía Regional.
7. DANE. (2005). Censo Nacional de Población 2005. Bogotá, D. C.: DANE.
8. Easterly, W. (2002). The elusive quest for growth: Economists' adventures and misadventures in the tropics. MIT Press.
9. Galvis, L. A. (2001). La topografía económica de Colombia (Documentos de trabajo sobre Economía Regional 22). Banco de la República-Economía Regional.
10. Galvis, L. A., & Bonilla, L. (2012). Desigualdades regionales en el nivel educativo de los profesores en Colombia. *Revista de Economía Institucional*, *14*(26),223-240.
11. Gutierrez, F. (2006). Tendencias del homicidio político en Colombia 1975-2004: una discusión preliminar. In Wills, M., and Sanchez, G. (Eds.), *Nuestra guerra sin nombre transformaciones del conflicto en Colombia.* Bogotá: Norma.
12. Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review, 60*(1), 126-142.
13. Iregui, A. M., Melo, L., & Ramos, J. (2006). La educación en Colombia: análisis del marco normativo y de los indicadores sectoriales. *Revista de Economía del Rosario, 9*(2), 175-238.
14. Iregui, A. M., Melo, L., & Ramos, J. (2007). Análisis de eficiencia de la educación en Colombia. *Revista de Economía del Rosario, 10*(1), 21-41.
15. Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics, 17*(1), 99-121.

16. Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review, 40*(2), 509-533.
17. LeSage, J., & Pace, R. (2009). *Introduction to spatial econometrics*. CRC Press.
18. Mora, J. J., & Ceballos, F. (2006). Pertinencia de la educación técnica y tecnológica en el mercado laboral de cali, 1994 -2005. Reportes observatorio de políticas públicas. Cali: Universidad ICESI.
19. Mora, J. J., & Muro, J. (2008). Sheepskin effects by cohorts in Colombia. *International Journal of Manpower, 29*(2), 111-121.
20. Quah, D. (1997). Empirics for growth and distribution: Stratification, polarization, and convergence clubs (Discussion Papers 1586). CEPR.
21. Ramírez, M. T., & Salazar, I. (2007). Surgimiento de la educación en la república de Colombia, en que fallamos? (Borradores de Economía 454). Banco de la República de Colombia.
22. Ramírez, M. T., & Téllez, J. P. (2006). La educación primaria y secundaria en Colombia en el siglo xx. (Borradores de Economía 379). Banco de la República.
23. Sanchez, G., Peñaranda, R., & Hurtado, C. (2007). *Pasado y presente de la violencia en Colombia.* Bogotá: La Carreta Editores, Universidad Nacional de Colombia and Instituto de Estudios Políticos y Relaciones Internacionales.
24. Tobler, W. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography, 46*(2), 234-240.

APPENDIX

A. Data description

The data used for the recent calculations was taken mostly from the CENSO 2005. The information calculated with the data from DANE is the percentage of people who have a degree in tertiary level of education, water supply coverage in the city, and economic variables. The construction of the data for use is presented below:

- Percentage of Tertiary Graduates (PTG): The percentage of graduates from tertiary level education in the city is calculated as follows:

$$PTG = \frac{Pgl_i}{(TotPop_i) - (Peoples15)} * 100 \quad (6)$$

  where the data comes from a survey undertaken in the CENSO and Pgl_i corresponds to people that graduated at tertiary level in city *i*, $TotPop_i$ is the total population from the survey in city *i*, and *Peoples*15 is the population in the survey under 15 years of age[16].

- Water supply coverage in the city (Wsupply): Percentage of people that have access to water services[17]. This data is also calculated by the CENSUS with a survey also.
- The economic variables are calculated using the number of economic units and the specific sector where they belong: the Commercial, Industrial and Service Sector. The variables account for the number of economic units for every thousand people.
- The supply of universities accounts for the number of universities in the municipality. This variable comes from the National System of Superior Education Institutions (SNIES) that belongs to the ministry of education.
- The violence variable was calculated by Francisco Gutierrez, who measured political homicide this variable is divided by the population of the municipality and multiplied by one thousand (Gutierrez 2006).

[16]The 15 years old range is used because if the 20 years old range were used it would mean that some people considered in the numerator are not considered in the denominator.

[17]This is taken into account because water is an important necessity.