PAPER

The effect of colonization on land concentration: Evidence from Colombian municipalities

Germán Tabares Pozos (MSc) gtabare2@eafit.edu.co
Thomas Goda (PhD) tgoda@eafit.edu.co

Escuela de economía y finanzas

Universidad EAFIT (Medellín, Colombia)

Subject area: *Population movements, inequality, poverty and inclusive growth*

Abstract: *This paper aims to identify the existence of a causal effect of the recent colonization process on the increase of rural land concentration in Colombian municipalities between 1984 and 2016. For this purpose, new calculations of the extent of large estate holdings in the municipalities of Colombia for the period 1984-2016 are presented. In a second step, municipalities that experienced a process of recent colonization are identified on the basis of a multivariate hierarchical clusters analysis. Finally, the existence of a causal relationship between recent colonization and the observed increase in land concentration is verified by means of Propensity Score Matching (PSM). The obtained results suggest that the recent colonization process in Colombia increased land concentration by approximately 8 percentage points in those municipalities that were affected.*

Keywords: land concentration, colonization, Propensity Score Matching, Colombia
JEL codes: J11, O18, Q15.

1. Introduction

Colombia has one of the highest levels of rural land concentration in the world (IGAC-CEDE, 2012). This fact is relevant considering that the country still has a relatively high level of rurality (PNUD, 2011), and that empirical evidence suggest that in developing countries land concentration tends to increase poverty rates by limiting the poor's access to means of production (Deininger and Olinto, 2000; Deininger and Squire, 1998), lower agricultural productivity (Berry and Clain, 1973; Feder, 1985), and spur violence by stimulating armed insurrections in regions where property rights are precarious (Binswanger et *al.*, 1995).

Taking into account these potential negative effects, the topic of land concentration has occupied a preponderant place in public discussions in Colombia and stimulated studies about its extent (IGAC-CEDE, 2012). However, one important limitation of these studies is that they exclusively focus on the Gini coefficient as concentration measure. While this widely used inequality measures has many benefits, it is not the most adequate to capture the share of large estates (so called *latifundios*) in total land holdings. Moreover, due to data constraints, Colombian land Gini coefficient estimates are only available for the post-2000 period on a municipal level.

A second limitation of the existing literature is that only few studies exist on the potential causes of the high degree of land in Colombia. Most of these studies argue that the colonization of new geographic areas plays an important role to explain the progressive land concentration in some specific Colombian regions (e.g., Jaramillo, 1986; Fajardo, 1994). However, despite the wide reproduction of this theory and its influence on public debates, there is still a lack of empirical evidence to quantitatively identify the existence and magnitude of the link between recent colonization and land concentration in Colombia.

This paper aims to contribute to the research on the impact of colonization on land concentration with three novel contributions: First, new calculations of the extent of land concentration in the municipalities of Colombia for the period 1984-2016 are presented, using as concentration measure the share of large land holdings (≥500 hectares) in total land holdings. Second, a multivariate hierarchical clusters analysis is used to identify those municipalities that experienced a process of recent colonization. Third, to establish if a causal effect between the recent colonization process and the observed increase in land concentration in Colombian municipalities exist a Propensity Score Matching (PSM) method is applied.

As for the first contribution, the calculations point to the existence of marked regional differences in the level and the variation of land concentration in Colombia, with a general increasing tendency. As for the second contribution, two groups with 119 recent colonized municipalities are identified that have a greater greographically scope than previously suggested. These share the common characteristics that their founding date is relatively recent, they have a long terrestrial distance to the nearest large city, and their population and agricultural boundary growth are greater than the national average. Finally, the results point to a causal effect of the colonization process on the variation of land concentration during 1984-2016, with a magnitude of approximately 8 percentage points.

The rest of the document is divided into five sections. The second section gives a brief description of the theoretical mechanism that can explain a potential link between recent colonization and variations in land concentration in Colombian municipalities. The third, fourth and fifth sections present the background, methodology and results of the three different contributions. Section Three handles land concentration, section Four newly colonized municipalities, and section Five the estimates of the effect of recient colonization on the variation of the concentration of land in Colombian municpalities. Section Six concludes.

2. The theoretical link between recent colonization and land concentration

Several theories hypothesize potential mechanism that explain the potential impact that the process of colonization has on the concentration of rural property. Among these are institutional ones developed by Engerman and Sokoloff (1997, 2000) and Acemoglu, Johnson and Robinson (2001), which argue that the access, strength and legitimity of property rights are of upmost importance to understand diverging land concentration patterns among colonized countries. While in North America solid property titles were given to new settlers, in the territories of Spanish and Portuguese dominion much of the land was divided in large estates that were worked by slaves and indigenous people and essentially remained under ecclesiastical or crown ownership. As a result, enduring power structures emerged that entrenchened the appropriaton of large holdings by the elite after independence.

The study of the relationship between the configuration of the land tenure structure and the concentration of land in *latifundios* in Colombia often is connected with the violence that the country has suffered from historically. This vision has been strongly influenced by the work of Legrand (1986, 1988, 1989) who identifies the colonization

of new geographical areas as a generator of violence that first leads to a more equitable distribution but afterwards to the expropriation and concentration of land.

To be more precise, according to this theory, the expansion of the agricultural frontier initially occurs under a relatively equitable distribution of land among settlers, without a clear definition of property rights. At this stage, the value of land is low and investment in the area is minimal. However, with the arrival of new settlers the occupation process intensifies, which increases the demand for land increases and thus its value. The increase in land value, in turn, leads to the growing need to define property rights, the purchase of land, and also to violent and legal conflicts over land (Simmons, 2005). Given that large property owners posess more money and local power, over time the land becomes concentrated in few large estates, which eventually leads to the displacement of those peasants that do not any longer have access to the land (initiating a new process of colonization to yet unsettled areas).

This vision of colonization as a structuring element of land tenure and land concentration has been widely reproduced in studies aimed at investigating the socio-economic problems of rural areas in Colombia (see, Machado, 1998; Kalmanovitz, 2004).

3. Land concentration in Colombia (1984-2016)

a. Background

The prominence of the agrarian question in Colombia has given rise to several studies aimed at calculating the extent of land concentration. The first of these studies were undertaken in the 1960s, using the information of the first colombian statistics about the rural sector from the *Muestras Nacionales Agropecuarias* and the *Censo Nacional Agropecuario*. According to these studies, the share of holdings with extensions greater than 500 hectares increased substantially from about approximately 27% to 41% between 1956 and 1970 (Alameda, 1964, CIDA, 1966, Barraclough and Dumike, 1966; Berry, 1973).

Throughout the 1980s, 1990s and beginning of the 2000s, the research on the magnitude of the concentration of rural property had a national and departmental scope, using the Gini coefficient as measure to study its possible connections with different socio-economic issues. For example, Machado (1998) results suggest that the national land Gini coefficient was 0.83 in 1984 and 0.88 in 1997, while Rincón's (1997) results suggest a slightly lower degree of inequality with Gini coefficients of 0.81 in 1984 and 0.83 in 1997. Castaño (1998), on the other hand, estimates a national land Gini

coefficient of 0.84 for 1997. The World Bank (2004) estimates for 1997 are similar and suggest that a slight inequality increase to 0.85 in 2002. In the latter year, the departments with the lowest land inequality were Caquetá and Cesar, as well as the far eastern departments of the country, with a Gini below 0.65, whereas the departments of Cauca, Quindío, Meta, Casanare and Valle del Cacua show the highest degree of inequality with registered Gini above 0.80.

More recent estimates increasingly are focusing also on land inquality figures within Colombian municipalities, taking advantage from better data availability thanks to the census data of the Agustín Codazzi Geographic Institute (IGAC), which provides comprehensive national, departmental and municipal level data from 2000 onwards. Considering these data, Ibáñez and Muñoz (2010) and IGAC-CEDE (2012) highlight the persistence of a high level of concentration of rural property in Colombia with a national land Gini equal to 0.85 in 2009 but also the existence of large variations between municipalities, with Gini coefficients ranging from 0.14 (San José del Palmar, Chocó) to 0.98 (Chiscas, Boyaca).

b. Methodology

As discussed above, during the last three decades the Gini coefficient has been used as measure of choice to estimate the degree of land inquality on a national and municipal level. While the Gini coeffificient is the most widely used inequality measure, it presents two important drawbacks. First, it restricts the time period that can be studied (given that only after the year 2000 reliable data about the overall distribution of land is available). This is an important limitation for the present investigation on the ground that the colonization process is a long-term phenomenon. Second, the Gini coefficient is not the most adequate indicator when considering changes in the concentration of land, given that it accounts for the whole distribution. It is well documented that this measure is largely influenced by changes close to the mean value of the distribution and that it shows little variation when it reaches high inequality values. Moreover, municipalities with the same Gini coefficient can have very different levels of land concentration in large real estates. Hence, other measures seem preferable if one wants to capture above all changes in the upper tail of the distribution (Cowell, 2000; Atkinson and Piketty, 2010, Alvaredo, 2011, Osberg, 2016)[1].

[1] This is the reason why recent papers prefer top income and wealth shares as measure to show changes in the concentration of income and wealth (e.g., Piketty, 2005; Piketty and Saez, 2006; Alvaredo *et al.*, 2013).

Following Lorente et al. (1986), and taking into account the available data, this study proposes the use of the participation of *latifundios* in the total rural cadastral area as measure of land concentration. According to the Colombian Institute of Agrarian Reform (INCORA), *latifundios* are real estates with an extent equal or larger than 500 hectares. Although small in number, these estates represent a considerable part of the rurual cadastral area and are a focal point of agrarian debates in the country (Lorente *et al.*, 1986). Hence, the concentration measure that is used in this paper is the following:

$$C_i = \frac{SC \geq 500_i}{SCT_i} \quad (1)$$

where C_i is the land concentration index in the *i-th* municpality, $SC \geq 500_i$ is the occupied cadastral area by properties with an extension greater than or equal to 500 hectares in the *i-th* municipality, and SCT_i is the total rural cadastral area total of the *i-th* municipality.

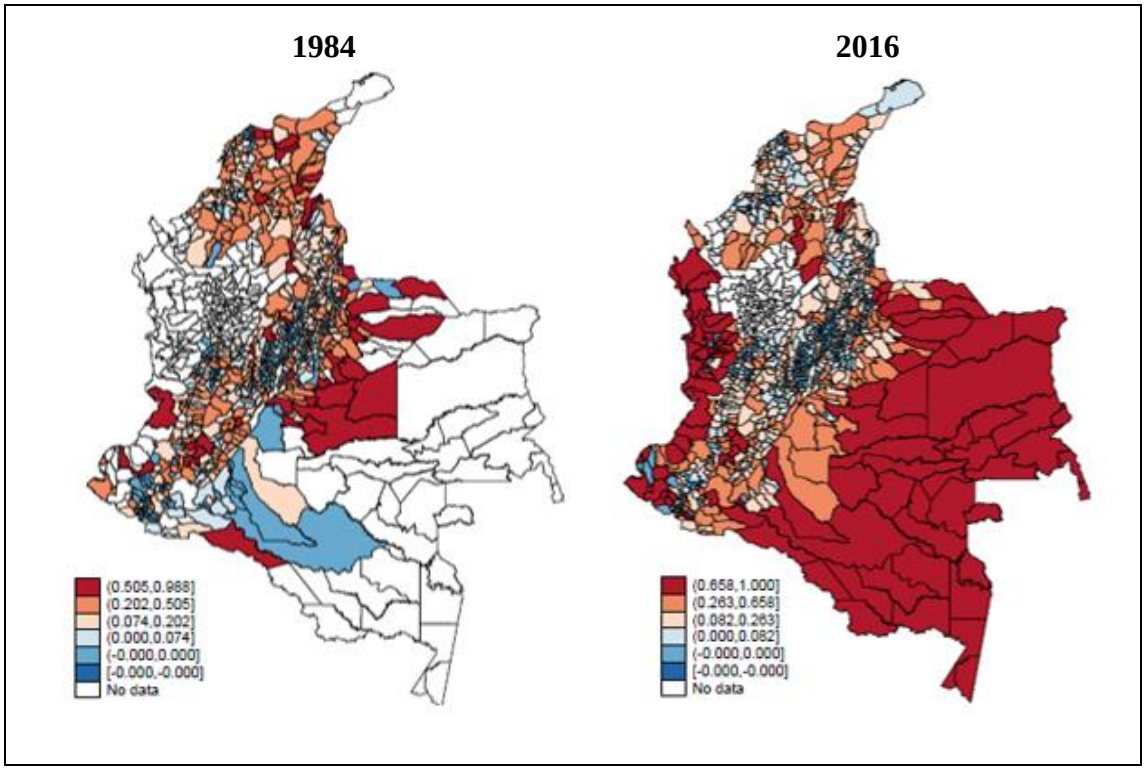
C_i takes values between zero and one, zero being the complete absence of *latifundios* and one the total concentration of land in large estates. For the year 1984 the index data are taken from Lorente *et al.* (1986), and for the year 2016 they were kindly provided by IGAC.[2] Please note that through time for a growing number of municipalities information is available. In the year 1984 only data for 804 municiplaties exist, while in the year 2016 this number rises to 916 (out of 1.122). One reason for the missing data is that neither Lorente *et al.* nor IGAC provide municipality data for the department of Antioquia, which is the sixth largest department of the country.

c. Results

Figure 1 presents the degree of land concentration for the years 1984 and 2015 for all Colombian municipalities for which data are available. The data show a general increase in land concentration and the heterogeneity between municipalities. In 1984 the average index value was 0.13 (with a standard deviation of 0.18), while it rises to 0.20 (std.dev. 0.28) in 2016. In other words, the results show that currently estates with a size of ≥500 hectars comprise 20% of all cadastral land in Colombia. The two maps also suggest the existence of regional patterns, being especially prominent the results observed in the Amazon region and the Nariñense Pacific.

[2] To make the concentration estimates comparable over time, the political-administrative division of the year 2016 was modified to reflect that of the year 1984 (80 new municipalities were created after 1984, and were treated as if they still were part of the area they belonged to previously).

Figure 1: The degree of land concentration in Colombian municipalities



Note: This two maps show the cartographic representation of land concentration in 1984 and 2016 at the municipal level. The used index takes values between zero and one, where one is total concentration.

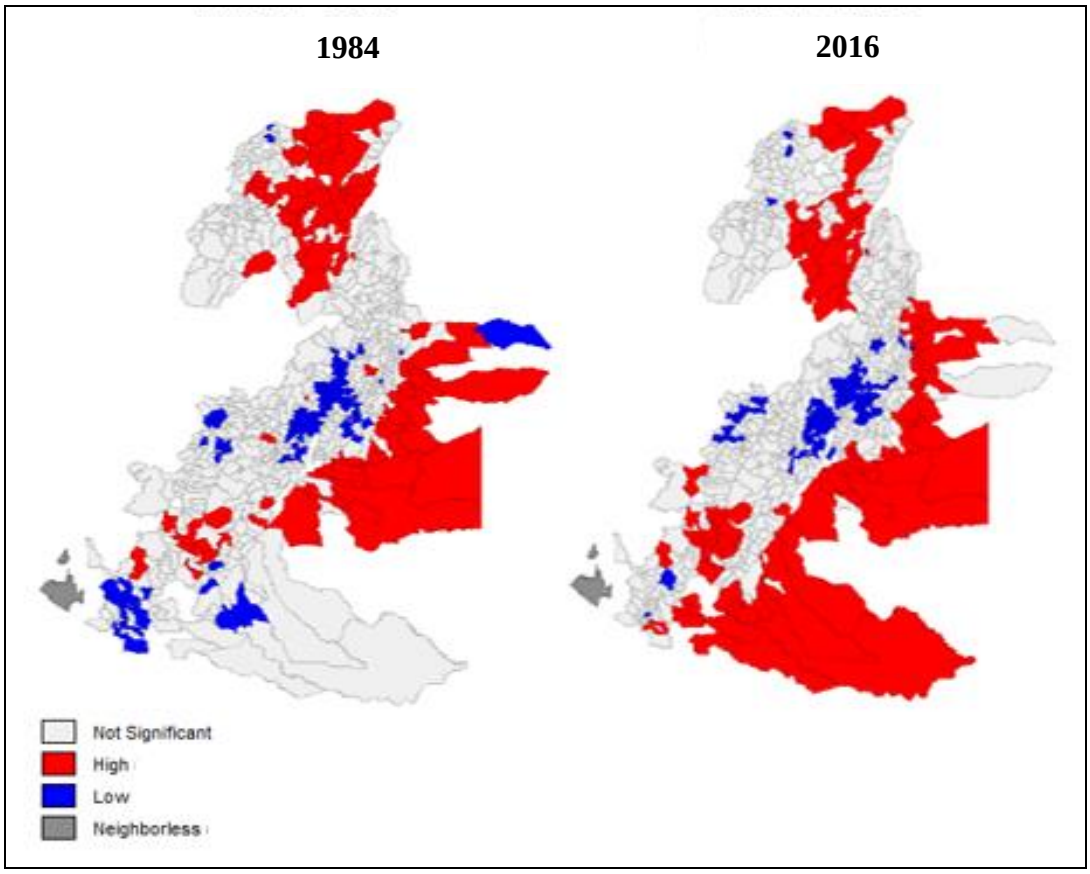
Given the existence of regional patterns, we also perform a spatial analysis with the Moran's I Index, to provide a more complete exploration of these results and to capture the potential existence of neighborhood effects.[3] The results of this spatial analysis suggests the spatial dependence of the concentration of the land, the formation of clusters, and evidences land concentration in Colombia is a regional issues. Accordingly, the Local Space Association Indicators (LISA) analysis (Anselin, 1995) is used to identify high and low concentration clusters. LISA also allows the detection of spatial autocorrelation patterns in small areas of the territory that is being analyzed globally.

The results of the LISA analysis for the year 1984 suggest clusters of high land concentration in the regions of the Eastern Plains (Meta, Casanare and Arauca), South Bolivar, Eastern Atlantic Coast (around the foothills of the Sierra Nevada de Santa Marta), and South Tolima. On the contrary, low concentration clusters are identified in

[3] Moran's I index incorporates in the Pearson correlation coefficient the location in space by including a matrix of spatial weights W_{ij}, obtaining $I = \frac{N}{S_0}\frac{\sum_i \sum_j W_{ij} z_i z_j}{\sum_i z_i^2}$. Where I is Moran's I Index, $Z_i = X_i - \bar{X}$, where X_i is the value of the rural property concentration of the *i-th* municipality, $S_0 = \sum_i \sum_j W_{ij}$ and $W_{ij} Z_i$ is the spatial lag of Z. The index for the years 1984 and 2016 is positive and statistically significant at the 1%-level, with values of 0.37 and 0.43 respectively.

the Cundiboyacense Altiplano, around Bogota, Eastern Arauca, Nariñense Altiplano, and Caquetá in the Amazonas (see Figure 2).

Figure 2: Clusters of high and low land concentration in Colombia

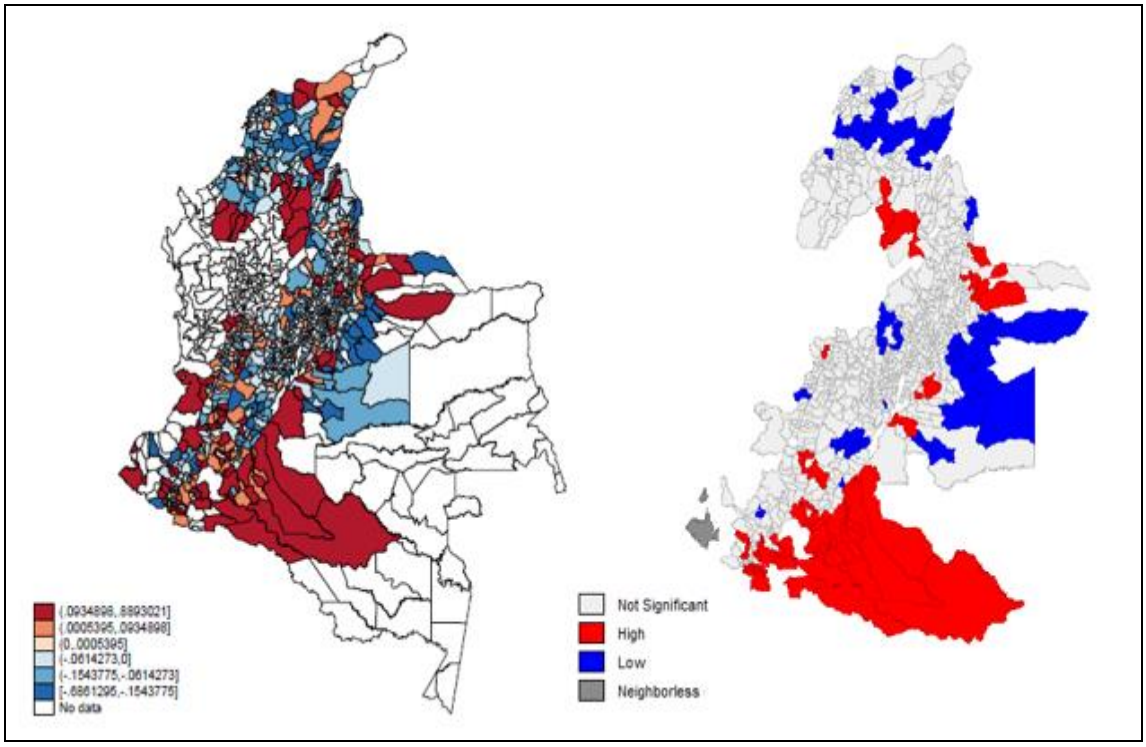


Note: These two maps show the results of the spatial LISA analysis of high and low land concentration for ther years 1984 and 2016.

In 2016, the clusters of high concentration cover the great majority of Amazonian territory (Arauca, Casanare, Meta, Caquetá, and Putumayo), South Bolivar, La Mojana Sucrense, the foothills of the Sierra Nevada de Santa Marta, and the zone comprised by the South of Tolima, the West of Huila and the North of Cauca. The clusters of low concentration, on the other hand, are in the Altiplano Cundiboyacense, and some zones in Nariño and the main coffee region of the country (Eje Cafetero).

Figure 3 shows both the absolute change in the concentration of land during period of study as well as the clusters of increasing and decreasing land concentration. Land concentration increased in the Amazonian regions of Caquetá and Putumayo, Arauca, South Tolima, and South Bolivar, whereas it decreased in the Center of the Atlantic Coast, in the Eastern Plains (Meta and Casanare), and some small parts of the Andean region.

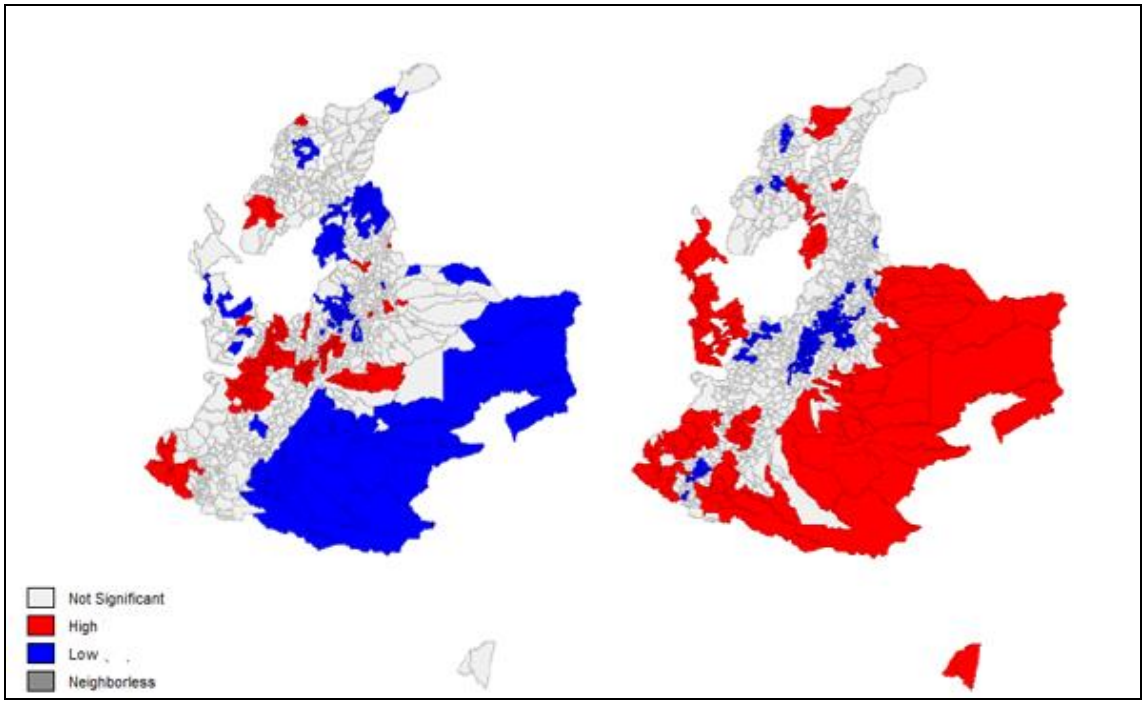
Figure 3: Variation in the concentration of land in Colombia during 1984-2016



Note: The left map shows the absolute change in the land concentration index for the period 1984-2016, while the right map shows spatial clusters of increase and decrease according to LISA.

Finally, Figure 4 compares the spatial clusters for the year 2016 that are obtained when using the Gini index and when using our alternative measure of concentration, and shows that the results obtained with the concentriation index are opposite of those from the Gini index. Most of the regions with low Gini coefficients are those that exhibit high levels of concentration. To be more precise, the concentration index is high and the Gini index is low in large part of the Amazon, the region of Orinoquía, South Bolivar, and some zones of Chocó, whereas in some areas of the Andina Region (Eje Cafetero and West Cundinamarca) the opposite is the case. Both measures only coincide in identifying the Pacific Coast of Nariño and the Northwest of Meta as high concentration clusters.

Figure 4: The Gini coefficient and the land concentration index in comparision



Note: This figure compares high and low concentration spatial clusters obtained by using the Gini coefficient and the concentration index used in this paper.

The marked differences between these two measures show the importance of differentiating between inequality and concentration, and can be seen as an important contribution to the study of land inequality in Colombia. Moreover, given that we are interested to capture the concentration of the land at the top of the distribution (as pointed in Section Two, the colonization process is expected to favor the formation and expansion of large estates), the proposed concentration index seems to be more adequate than the Gini coefficient for our research purpose.

4. Recently colonized municipalities in Colombia

a. Background

Existing literature identifies Colombian territories that experienced a recient colonization process on the basis of population growth data (Jimeno, 1983), changes in land titles and available land (IGAC, 1986), and the spontaneous settlements registered between 1970-1985 (Ministry of Agriculture, 1985; Lorente *et al.*, 1986). The results of this identification strategy have been widely reproduced in the socio-economic literature about Colombia and is still used in recent literature (see e.g. Albertus and Kaplan, 2013; Caicedo Vargas, 2015). Figure 5 shows the cartographic representation of the such identified regions, which are Guaviare, Caquetá, Caguán, Magdalena Medio, Arauca, Urabá, South Córdoba, South Bolívar and Catatumbo as zones of recent colonization (in vertical lines). All of these regions are characterized for being territories that have a

relatively low altitude and that are located relatively far away from the largest population centers.

Figure 5: Areas of Recent Colonization by IGAC (1986)

Note: This map shows the areas of recent colonization (vertical lines) according to IGAC (1986) (presented in Legrand (1989)), as well as the territories of influence of the Antioqueno colonization (horizontal lines) and the Sumapaz region.

It is important to note that the above mentioned identification strategy has several shortcomings: (i) it stems from the mid-1980s and thus does not incorporate the dynamics that have occurred during the last three decades, (ii) it does not consider some other important elements besides titled areas, and population growth, and (iii) it lacks a statistical methodology that allows a more rigorous identification and grouping of the municipalities, taking into account multiple variables.

b. Methodology

In order to identify the municipalities that have experienced a process of recent colonization in Colombia[4], a multivariate hierarchical cluster analysis is used, where municipalities with similar conditions are grouped together according to a set of variables (X):

$$Groups = f(X_1, \dots, X_k) \quad (2)$$

[4] The administrative political division used in this section is the one corresponding to 1984, in order to have results that are comparable with those obtained in terms of land concentration.

The variables used for the grouping process are proxies of potential elements that explain the recient colonization process in Colombia, i.e. the migration to previous uninhabited areas that leads to the geographic expansion of agricultural and extractive activities (Fajardo *et al.*., 1997). To capture the conditions of geographic periphery, population growth, the change of land use, and the insertion of the area into the political administrative system, the following four variables are considered for the grouping: (i) time of overland travel to the nearest major city (geographic periphery)[5], (ii) deviation of rural population growth in terms of the national average for the period 1964-2015[6], (iii) deviation of the agricultural frontier growth with respect to the national average during the period 1960-2014 (land use)[7], and (iv) age of the municipality in the year 2015 (insertion into the administrative political system).

The descriptive statistics of these four variables are presented in Table 1.

Table 1: Identification variables of the colonizing process

Variable	Source	Mean	Std. Dev	Min.	Max.
Age of municipality in 2015	CEDE	154	106.9	19	490
travel time to nearest large city	Google Maps	5.12	4.5	0.28	21.74
Deviation of rural population growth from the national aggregate during the period 1964-2015	*Censo General de Población 1964* / CEDE	0.71	8.8	-1.20	252.26
Deviation of agricultural frontier growth from the national aggregate during the period 1960-2014	*Censos Nacionales Agropecuarios* 1960 – 2014	0.14	2.1	-1.16	52.04

Note: This table shows the sources and the descriptive statistics of the variables used to group the Colombian municipalities according to their characteristics.

Since the four variables that are used for the grouping process are continuous three clustering algorithms would be suitable (Everit et al., 2011; Mooi and Sarstedt, 2011): The nearest neighbor, the most distant neighbor, and Ward's (1963) method. Since in

[5] Colombia's five main cities are Bogotá, Medellín, Cali, Barranquilla and Bucaramanga.

[6] Between 1964 and 1984 108 municipalities were created in Colombia, so that the population data provided by the General Population Census of 1964 do not correspond to those observed in 1984. Hence, newly created municiplaities are treated according to the administrative political division of 1964.

[7] Please note that some interpolations had to be made in order to make the figures for area under agricultural exploitation comparable in 1960 and 2014.

our case there exist no objective selection criterion, the decision which of these algorithms should be used depends mainly on the ease of the final results (Tezi, 2012; Everit et al, 2011). Considering that Ward's method permits to minimize the loss of information during the grouping process (Everit *et al.*, 2011), and that it is generally used in studies about development taxonomies (see e.g., Tezano, 2012, Tezano and Quiñones, 2012, Tezano and Sumner, 2013), Ward's method is employed.

To this end the squared sum of the Euclidean distance of each unit with respect to the centroid of the group to which it belongs is quantified, and the distances corresponding to all the units are summed. The formal expression of Ward's distance is presented in equation (3).

$$W = \sum_g \sum_{i \in g} (x_{ig} - \overline{x_g})'(x_{ig} - \overline{x_g}) \quad (3)$$

where, $\overline{x_g}$ is the average of the group g and i is the *i-th* municipality that is part of the group.

One of the most important points in the cluster analysis is the identification of the number of groups in which the the analyzed units (i.e. municipalities) will be clustered. To identify the number of groups the decision rule proposed by Duda *et al.* (2001) is applied, which seeks to identify the extent to which it is justified to continue the division process of the set of units analyzed. This rule is based on the hypothesis that, if all *n* samples ar normally distributed, the formation of new clusters beyond a specific number is the result of spurious separation. It is expected that the latter is the case, if by increasing by one unit the number of groups the sum of the square errors of the samples $J_{c(2)}$ is less than the sum of the square errors of the number of previous clusters $J_{c(1)}$. Consequently, it is only justified to increase the number of groups up until the point where $J_{c(2)} \geq J_{c(1)}$.

c. Results

Duda *et al.*'s (2001) decision rule indicates that in our case the optimal number of municipality groups is 25. To select the groups that are formed by recently colonized municipalities, the following selection criteria is used: all groups with a mean age below the average, longer than average displacement times, and a standard deviation of population and agricultural frontier growth that is higher than the mean (see Table 1 for the reference data). Under this selection criterion, two groups (group 22 and 23) that comprise 119 municipalities can be regarded as recently colonized (Table 2 for the descriptive data of these two groups).

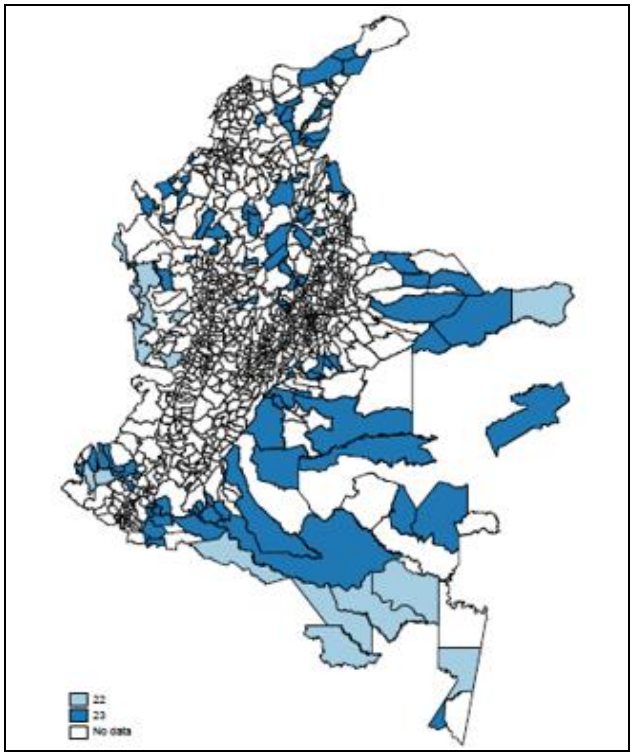
Table 2: Groups of newly colonized municipalities

Group	No. of Municipalities	Age	Travel time to nearest large city	Population growth	Agricultutal frontier growth
22	18	63	21	2.37	0.98
23	101	35	8.3	1.83	1.51

Note: This table shows the characteristics (averages) of recently colonized municipalities in Colombia.

The cartographic representation of these 119 municipalities is shown in Figure 6. In the north of the country they are located in the north-east of the Atlantic Coast, the south of the Sierra Nevada of Santa Marta, South Bolivar, the center of the Magdalena region, the south-east of Cordoba, Urabá and Catatumbo. In the pacific region the Chocoano coast and the north of the Nariñense Coast can be considered as recently colonized, and in the Orinoquía region Arauca, North Vichada, Casanare, and the center and south of Meta and Guaviare. Finally, the majority of the municipalities in the Amazonian territory are part of Group 22 and 23.

Figure 6: Recently colonized municipalities in Colombia



Note: This figure shows the geographical location of those municipalities that are identified as recently colonized municipalities.

When comparing Figure 5 with Figure 6, it can be observed that our identification strategy comprises more municipalities than previously identified. One reason arguably

is that we use newer information but anther reason is that, in contrast to the other studies, we use multiple elements that can be seen characterize the process of recient colonization. Hence, the findings of this section on their own constitute an important contribution to the existing research and debastes about the phenomenon of recent colonization in Colombia.

5. The effect of colonization in land concentration in Colombia

a. Background

Various studies aim to understand the link between recient colonization and the structure of rural land ownership in Colombia. For example, Reyes (1987) stresses the importance of the historic and recent colonization and migration processes to explain the formation of today's agriculture boundaries in the departmens of Antioquia and Santander. In the same vein, Molano (1987, 1988), Jaramillo (1986), and Fajardo (1994, 2002) study the effect of colonization on the agrarian issues that are present in the mountain range of Macarena and Guaviare.

Legrand (1989), on the other hand, gives an exhaustive overview of the different approaches to explain the socio-economic relations that are present in Colombia's border regions, exploring the possible link between the colonisation process and waves of violence in Colombia. According to her, the precarious state presence in these areas, the absence of land titles, and the increasing profitability of land ownership have sparked disputes over land that goes along with increasing land concentration.

However, to support the argument that recent colonization is vital to explain observed changes in land concentration in Colombia descriptive statistics about demographic behavior, changes in land use, and cadastral information are used (Legrand, 1986; Reyes, 1987), as well as case reports on different socio-economic aspects ranging from the characteristics of political changes to agricultural dynamics (Molano, 1986; Fajardo, 1994). In other words, the existing studies on the topic neither have a national character, nor do they explore empirically the existence of a causal relation between recient colonization and land concentration neither its magnitude.

b. Methodology

In theory, to establish and quantify causal effects of a phenomenon over another the same unit should be observed in an experiment with and without the phenomenon. Unfortunately, to study simultaneously the effect of the occurrence and non-occurrence of something at the same time is not possible in practice. Thus, to approximate the existence of a causal relation, experimental studies typically choose randomly units

where some are exposed to the phenomenon (treatment) and others not. However, such an experimental section is not viable in our case, given that their specific characteristics define whether municipalities are subject to being colonized or not. Therefore, following Rosenbaum and Rubin (1985), a Propensity Score Matching (PSM) approach is used to observe causality in a quasi-experimental setting (Steiner and Cook, 2011). To be more precise, to estimate the possible causal effect of recient colonization on land concentration the information that are available for both recient colonized and not recently colonized municipalities is used.

More formally, each municipality has two potential outcomes, a control outcome Y_i^0, under the condition of not being recently colonized $(Z_i = 0)$, and the treatment outcome Y_i^1, under the condition of being recently colonized$(Z_i = 1)$. Both scenarios are potential results to the extent they are not yet known (it is assumed that the process of recent colonization has not taken place yet) but fixed before receiving treatment. Once the treatment is given one of the two potential outcomes is revealed.

Given the pair of results (Y_i^0, Y_i^1) one could quantitatively estimate the causal effect effect of the colonization process on the variation of land concentration:

$$\tau_T = E(Y_i^1 - Y_i^0 | Z = 1) = E(Y_i^1 \,|Z = 1) - E(Y_i^0 \,|Z = 1) \tag{4}$$

However, in practice both potential results cannot be observed simultaneously. At the group level only the result of colonization on those municipalities that are recently colonized can be observed $E(Y_i \,|Z = 1) = E(Y_i^1 \,|Z = 1)$, and the control result for municipalities not recently colonized $E(Y_i \,|Z = 0) = E(Y_i^1 \,|Z = 0)$. The difference between the two terms is $E(Y_I^0)$ and $E(Y_I^1)$, given the difference in the selection of reciently and not reciently colonized municipalities. Therefore, the average difference between both groups $\hat{\tau} = \frac{1}{N_T}\sum_{t\in T} Y_i - \frac{1}{N_C}\sum_{c\in C} Y_i$ is a biased estimator (Steiner and Cook, 2011).

This estimator would only be unbiased if recently colonized municipalities would have been selected to experience this random process. However, this is impossible in our context because the selection that the colonization process in Colombia did not occur randomly. To overcome this problem a set of covariates (X) is employed to capture the selection process of those territories that were colonized in the 1960s, so that the potential outcome (Y_i^0, Y_i^1) is independent of selection process:

$$(Y_i^0, Y_i^1) \perp Z \,|\, \mathbf{X} \tag{5}$$

This condition, known as unconditional dependence, allows an unbiased estimator because it allows to identify within the group of municipalities not recently colonized those that are likely to have experienced this process $0 < P(Z = 1|X) < 1$, thus building a control group whose changes in land concentration during the period 1984-2016 serve as counterfactual for those municipalities that did experience a recient colonization process. Therefore, the estimator of the effect of recient colonization on the variation of land concentration in reciently colonized municipalities is:

$$\hat{\tau}_T = E_{P(X)|Z=1}\{E[Y(1)|\, Z = 1,\, P(X)] - E[Y(0)|\, Z = 0,\, P(X)]\}$$

It is the estimator of the causal effect obtained by matching strategy known as PSM, whose basic idea is to approach the analysis of counterfactual scenarios (non-experimental) observed data as a problem of imputation of missing data from the comparison of treatment groups and control, obtained by estimating a propensity index given a set of covariates which overcome the problem of selection bias.

It is considered that there are (at least) three conditions that determine the likelihood of a municipality to have experienced a recent colonization process, which at the same time also affect the variation of land concentration: (i) the condition of the geographic periphery regarding territory and settled areas, (ii) the availability of space for the migrant population and agricultural activities to develop, and (iii) weather and land conditions, since it is expected that the territories more suitable for human settlement and agriculture were occupied in previous colonization stages.

The variables used to try to capture these conditions are altitude and distance to the main market (peripheral geographical condition), population density (available space), and precipitation and soil suitability (climatic and ecological conditions). These five variables also are likely to have an impact on changes in land concentration in Colombia (see Machado, 1998; Ibáñez and Querubín, 2004, BM, 2004). Table 3 presents the descriptive statistics of these variables.

Table 3: Covariates

Variable	Source	Mean	Std. Dev	Min.	Max.
Altitude	CEDE	1217.03	1169.39	2	3087
Distance to the nearest large agricultural market	CEDE	119.77	97.80	0	926.46
Population density 1964	CEDE / CGP 1964	0.66	1.96	0	40.05
Soil quality	CEDE	2.68	1.20	0	8
Rain	CEDE	153.73	91.22	20.50	793.18

Note: The altitude variable is measured in meters above sea level; The linear distance in km; Density is expressed as the ratio of inhabitants per hectare; The suitability of the soil takes values between 0 and 8 being 0 a poor suitability of the soil for agricultural use and 8 the existence of perfect conditions for the development of this activity, these calculations obey to the methodology used by the IGAC for the categorization of the soil; And the rainfall variable is measured as the average rainfall cm3 received annually in the territory.

c. Results

The results of the econometric model used to capture probability of a municipality to experience a recent colonization process shows that all the variables have the expected sign and only soil suitability is not statistically significant (see table A2). By doing the estimate of the effect of the colonization process over the variation of the concentration of land in the municipalities recently colonized, there is evidence that this turns out to be positive and statistically significant the short use of nonparametric algorithm (Kernel) for pairing (matching) between the municipalities in the treatment group and control (table 4). The magnitude of this effect is 0.080, which means that within 1984-2016 in the municipalities recently colonized, the land concentration increased about eight percent higher than in municipalities that did not live this process.

Table 4: PSM results

	Observed Coef	Bootstrap Std. Error	Z	P>|z|	Normal-based (95% Conf. Interval)	
$\hat{\tau}_T$	0.0800	0.3837	2.09	0.037	0.0048	0.1552

The quality of the resulting matching algorithm using Kernel was tested by incorporating propensity score in probabilistic regression, bringing with it the loss of

significance of covariates on the probability of having experienced the colonization process suffers, revealing the quality of pairing (see Table A1). Meanwhile, tests balance of covariates, both individually and overall were found to be favorable (see Table A2), making group control identified suitable for the counterfactual analysis. As an exercise robustness of the causal effect estimates they were conducted under different matching algorithms including parametric algorithms such as nearest neighbor (nn), as shown in table 5. It shows how with different estimates the effect of colonization on the variation of land concentration in the municipalities recently colonized turns out to be positive and statistically significant. The magnitude of the effect varies between 7.9 and 8.8 percentage points.

Table 5: Robustness

Algorithm	$\hat{t}_T$	**S.E**	**Z**	**Off**	**B**	**R**
NN(5)	0.079	0.036	2.20	4	15.4	0.97
NN(5) C(0.05)	0.088	0.033	2.62	4	20.1	1.13
Radius (0.03)	0.083	0.033	2.48	4	16	1.11
Kernel	0.080	0.038	2.09	4	15.8	0.61

Note: the estimates for each type of algorithm were obtained through a Probit model. The expression C (0.05) alludes to the imposition of a caliper of 0.05 in the algorithm of five nearest neighbors. The standard error for the estimates using the Kernel and Radius Algorithms is obtained by using Boostrap. For the case of the closest Neighbors algorithm was obtained following Abadie and Inbens (2006). The Off column indicates the number of municipalities treated outside the common support area. The variables B and R constitute measures of global balance of the matching proposed by Rubin (2001); If B takes a value less than 25 and R between 0.5 and 2, the treatment and control groups are considered to be globally balanced.

6. Conclusions

This research offers three novel contributions to the existing literature: (i) the degree and variation of land concentration in Colombia's municpalities is shown for a longer time period than previous done (1984-2016), (ii) a new strategy is used to define reciently colonized municipalities, covering more elements than previously, and (iii) this is the first paper that tests for the causal effect of recient colonization on land concentration in Colombia, and that offeres estimates about its magnitude.

We show that marked regional differences exist regarding the degree and variation of land concentration in Colombia. In gernal terms, land concentration increased during the period 1984-2016 (from 13% to 20%), with the largest increases recoreded the

Amazon, Arauca, Southern Tolima, and Magdalena Medio. Moreover, we find that for the municipalities of Colombia the Gini coefficient is not the most adequate measure to depict the concentration of land holdings in large real estates (so called *latifundios*).

The second contribution of the paper is to identify and group those municipalities that experienced a process of recient colonization according to four criterias (geographical pheriperhy, population growth, land use, insertion into the country's political administrative system). According to the presented findings, more municipalities than previously considered should be regarded as recently colonized; more specifically, municipalities in the north of Nariño´s Pacific Coast, Chocó, Lower Guajira, the southern Sierra Nevada de Santa Marta, the west of Cesar and Vichada, and most parts of the Colombia's Amazon.

Finally, taking into account the findings of the first two contributions, in a quasi-experimental setting it is observed that the process of recent colonization indeed has had a significant causal effect on the progressive concentration of land in come of Colombia's municipalities. During 1984-2016 those municipalities that can be regarded as recently colonized experienced an increase in land concentration by the considerable magnitude of approximately 8 percentage points. This finding provides empirical evidence in favour of multiple existing socio-economic studies that content that such a relationship exists.

In light of the presented results, interesting and fruitful avenues for future research could be to study in more detail the mechanism that explain the interaction between colonization and land concentration. This is especially true for the role that violence has played in the colonization process (viceversa).

References

Abadie, A., and Imbens, G. (2006). *Large Sample Properties of Matching Estimators for Average Treatment Effects.* Econometrica, 74 (1), 235–267.

Acemoglu, D; Johnson, S; and Robinson, J. (2001). *The Colonial Origins of Comparative Development: An Empirical Investigation.* American Economic Review, 91(5), 1369-1401.

Alameda, R. (1964*). La tenencia de la tierra y la reforma agraria en Colombia*. Economía colombiana. Vol. 20, No 58 (Feb.). Bogotá.

Albertus, M., & Kaplan, O. (2013). *Land Reform as a Countrerinsurgency Policy: Evidence from Colombia*. Journal of Conflict Resolution, 57(2): 198-231.

Alvaredo, F. (2011*). A note on the relationship between top income share and the Gini coefficient*. Economic Letters 110, 274-277.

Alvaredo, F., Atkinson, A., Piketty, T., & Saez, E. (2013*). The top 1 percent in international and historical perspective*. Journal of Economic Perspective. Journal of Economic Perspectives, 27(3), 3-20.

Anselin, L. (1995*). Local Indicators of Spatial Association*. Geographical Analysis, 27(2), 93-115.

Atkinson, A.B., Piketty, T., 2010. *Top Incomes over the Twentieth Century: A Global Perspective*. Oxford University Press, Oxford.

Banco Mundial. (2004). *Colombia, una política de tierras en transición*. Banco Mundial – Documento CEDE 2004-29.

Barraclough, S y Dumike, A. (1966*). La estructura agraria en siete países de América latina*. El trimestre económico No. 13º, p. 230-302. México.

Berry, A. (1973*). Land Distribution, Income Distribution and the Productive Efficiency of Colombian Agriculture.* Food Research Institute Studies. 12(3): 199-232.

Berry, A y Cline, W. (1979). *Agrarian Structure and Productivity in Developing Countries*. Baltimore: Johns Hopkins University.

Bonet, J., & Meisel, A. (2006*). Polarización del ingreso per cápita departamental en Colombia, 1975-2000*. En Documento de Trabajo, 76. Cartagena: Centro de Estudios Económicos Regionales, Banco de la República.

Binswanger, H., Deininger y Feder, G. (1995*). Power, Distortions, Revolt and Reform in Agricultural Land Relations*. In J. Behrmann & T. N. Srinivasan (Eds.), Handbook of Development Economics (Vol. III): Elsevier Science.

Caicedo Vargas, S.A. (2015). Apropiación y uso del territorio en la colonización campesina en el suroccidente del departamento de Cundinamarca en la segunda mitad del siglo XIX. Revista Grafía, 12(1): 142-163.

Cardenas, M., Eslava, M., & Ramirez, S. (2016). *Why internal conflict deteriorates state capacity? Evidence from colombian municipalities*. Defence and Peace Economics.

Castaño, L. (1999). *La Distribución de la Tierra Rural en Colombia y su Relación con el Crecimiento y la Violencia*. Master's thesis, Facultad de Economía. Universidad de los Andes.

CIDA. (1966). *Tenencia de la tierra y desarrollo socio-económico del sector agrícola: Colombia*. Unión Panamericana, Washington.

Córtez, D., & Vargas, J. (2012). *Inequidad Regional en Colombia*. Documento CEDE, 127.

Cowell, F.A. (2000*). Measuring Income Inequality*. London: Harvester Wheatsheaf.

Deininger, K., y P. Olinto. (2000). *Asset Distribution, Inequality and Growth*. World Bank Policy Research Working Paper 2375. Washington, D.C.

Deininger, Klaus and Lyn Squire (1996). *A New Data Set Measuring Income Inequality*. The World Bank Economic Review, 10(3):565-591.

Diaz, A., & Sánchez, F. (2004). *Geografía de los Cultivos Ilicitos y el Conflicto Armado en Colombia.* Documento CEDE, 18.

Duda, R., Hart, P., & Stork, D. (2001). *Pattern Classification and Scene Analysis, 2nd ed*. New York: Wiley.

Engerman, S and Sokoloff, K (1997). *Factor Endowments, Institutions, and Differential Paths of Growth among New World Economies.* in S.H. Haber ed. How Latin America Fell Behind, Stanford University Press, Stanford CA.

Engerman, S and Sokoloff, K (2000). *Institutions, Factor Endowments, and Paths of Development in the New World.* Journal of Economic Perspectives, 3, 217-232.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. Chichester: John Wiley & Sons.

Fajardo, M., Mondragon, D., & Moreno, O. (1997). *Colonización y estrategias de desarrollo*. IICA, Bogotá.

Fajardo, D. (2002). *Para sembrar la paz, hay que aflojar la tierra*. Instituto de Estudios Ambientales IDEA, Universidad Nacional de Colombia. Bogotá.

Feder, G. (1985). *The relation between farm size and farm productivity: The role of family labor, supervision and credit constraints.* Journal of Development Economics, 18(2-3), 297-313.

Ibáñez, A y Muñoz, J. (2010). The Persistence of Land Concentration in Colombia: What Happened Between 2000 and 2009?. In Bergsmo, M., Rodríguez-Garavito, C., Kalmanovitz, P., and Saffon, M.P. (eds.): *Distributive Justice in Transitions*. Oslo: Peace Research Institute (PRIO), Ch. 9.

Ibáñez, A., & Querubín, P. (2004*). Acceso a tierras y desplazamiento forzado en Colombia*. Documento CEDE, 23.

IGAC. (1986). *Situación y análisis del proceso colonizador en Colombia*. Subdirección de Investigación y Divulgación Geográfica, Instituto Geográfico Agustín Codazzi.

IGAC-CEDE. (2012). *Atlas de la distribución de la propiedad rural en Colombia*. Bogotá.

Jaramillo, J., Mora, L., & Cubides, F. (1989*). Colonización, coca y guerrilla*. Alianza editorial colombiana. Bogotá.

Jimeno, M. (1983). *La descomposición de la colonización campesina en Colombia*. Colombia Amazónica. Vol. 6, № 1.

Kalmanovitz, S. (2006). *La Agricultura Colombiana en el Siglo XX*, Banco de la República - Fondo de Cultura Económica, Bogotá.

LeGrand, C. (1986). *Frontier expansion and peasant frontier in Colombia 1830-1936*, U. of New Mexico Presa. Albuquerque.

_____. (1988). *Frontier Expanssion and Peasant Frontier in Colombia 1830-1836*. University of New México Press. Albuquerque.

_____. (1989). *Colonization and Violence in Colombia: Perspectives and Debates.* Canadian Journal of Latin American Studies 14(28): 5-29.

Llorente, L; Salazar, A y Gallo, A. (1984). *Distribución de la Propiedad Rural en Colombia 1960-1984*. Bogotá: Ministerio de Agricultura; CEGA (Centro de Estudios Ganaderos y Agrícolas).

Machado, A. (1998). *La Cuestión Agraria en Colombia a fines del Milenio*. Bogotá: El Áncora Editores.

Meisel, A. (2014). *La no reversión de la fortuna en el largo plazo: geografía y persistencia de la prosperidad en Colombia 1500-2005*. Cuadernos de Historia y Economía Empresarial, 35.

Molano, A. (1987). *Selva adentro. Una historia oral de la colonización del Guaviare*. El Áncora Editores. Bogotá.

_____. (1988). *Violencia y colonización*. En Revista Foro. 6: 25-37. Bogotá

Mooi, E., & Sarstedt, M. (2011*). A concise guide to market research*. Berlin: Springer-Verlag.

Offstein, N. (2005). *National, departmental and municipal rural agricultural land distribution in Colombia: analyzing the web of inequality, poverty and violence*. Documento CEDE 2005-27.

Osberg, L. (2016*). On the limitations of some current usages of the Gini Index*. Review of Income and Wealth, early view, doi: 10.1111/roiw.12256.

Piketty, T. (2005). Top income shares in the long run: An overview. European Economic Association.

Piketty, T., & Saez, E. (2006). *The evolution of top incomes: A historical and international perspective*. NBER Working Paper, 11955.

Pizarro, E. (2016). *Una lectura multiple y pluralista de la historia. Comisión de Historia del Conflicto y sus Víctimas.* disponible en http://www.altocomisionadoparalapaz.gov.co/mesadeconversaciones/PDF/una-lectura-multiple-y-pluralista-de-la-historia-1447178719-1460381905.pdf

PNUD. (2011). *Informe Nacional de Desarrollo Humano. Colombia rural razones para la esperanza*. En http://escuelapnud.org/biblioteca/documentos/abiertos/06_indh2011co.pdf.

Reyes A. (1987). *La violencia y el problema agrario en Colombia*, in Análisis Político n° 2, Instituto de Estudios Políticos y relaciones Internacionales, Universidad Nacional de Colombia.

Rincón, C. (1997). *Estructura de la Propiedad Rural y Mercado de Tierras*. Thesis, Facultad de Economía, Universidad Nacional, Bogotá.

Rosenbaum, P., & Rubin, D. (1985). *Constructing a control group using multivariate*

matched sampling methods that incorporate the propensity score. The American Statistician, 39, 33-38.

Rubin, D. B. (2001*). Using propensity scores to help design observational studies: Application to the tobacco litigation. Health Services and Outcomes.* Research Methodology, 2, 169-188.

Simmons, C. (2004). *The political economy of land conflicto in the Eastern of Brazilian Amazon*. Annals of the Association of American Geographers, 94(1), 183-206.

Steiner, P., & Cook, D. (2011). *Matching and Propensity Score*. The Oxford Handbook of Quantitative Methods.

Tezanos, S. (2012). *Conglomerados de desarrollo en América Latina y el Caribe: una aplicación al análisis de la distribución de la asistencia oficial para el desarrollo*. Serie Financiamiento para el Desarrollo, 241. CEPAL.

Tezanos, S., & Quiñones, A. (2012). *¿Países de renta media? Una taxonomía alternativa del desarrollo de América Latina y el Caribe*. Revista Iberoamericana de Estudios de Desarrollo, 2, 4–27.

Tezano, S., & Sumner, A. (2013). *Revisting the meaning of development: a multidimensional taxonomy of developing countries.* Journal of Development Studies, 49, 1728-1745.

Ward, J. H. (1963). *Hierarchical groupings to optimize an objective function*. Journal of the American Statistical Association, 58, 36–244.

APPENDIX

Table A1: Quality of the Matching

Variable	1 P(Z=1)	2 P(Z=1)
Height	-0.000122 (-0.91)	-0.000249** (-3.20)
Distance to the nearest large agricultural market	0.00189 (1.00)	0.00377*** (6.02)
Population density	-0.237 (-1.33)	-0.424** (-2.67)
Soil quality	-0.0427 (-0.59)	-0.0214 (-0.41)
Rain	0.0013 (0.70)	0.00303*** (5.49)
Propensity Score	3.108 (1.66)	
Constant	-1.837*** (-5.72)	-1.772*** (-6.96)
N	793	984

t statistics in parentheses
*p<0,05, **p<0,01, ***p<0,001

Table A2: Quality of the Matching

Variable	Mean		t-test			VT/VC
	Tratados	Controles	%bias	t	p>t	
Altura	766.73	795.57	-2.8	(-)0.19	0.853	0.66
Distancia al mercado	153.27	152.89	0.5	0.03	0.978	0.97
Densidad Poblacional (1964)	0.24	0.35	-9.6	(-)0.87	0.387	0.63
Aptitud del suelo	2.80	2.77	2.3	0.13	0.896	0.97
Pluviosidad	174.2	170.43	4.4	0.22	0.824	1.27

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.006	1.05	0.958	3.9	2.8	15.8	0.61	0