CEDE

Centro de Estudios sobre Desarrollo Económico

Let the Rebels Rule? Evidence on the Economic Effects of Rebel Governance in Colombia

Santiago Pérez-Cardona

Documento CEDE

Estudiantes

#3

Enero de 2022

Universidad de los Andes
Colombia

Facultad de Economía

Serie Documentos Cede, 2022-3 ISSN 1657-7191 Edición electrónica. Enero de 2022

infocede@uniandes.edu.co
http://economia.uniandes.edu.co

Impreso en Colombia – Printed in Colombia

CEDE
Centro de Estudios sobre Desarrollo Económico

Documento CEDE

Estudiantes

Descripción: los documentos CEDE-Estudiantes contienen los resultados de las tesis de maestría (PEG o MEcA) de los estudiantes de la Facultad que se consideran merecedoras de este honor. Son seleccionados extraordinariamente por postulación de al menos un profesor de planta de la Facultad; pueden ser tenidos en cuenta también trabajos de estudiantes de pregrado.

Universidad de los Andes Colombia | Facultad de Economía

Let the Rebels Rule?
Evidence on the Economic Effects of Rebel Governance in Colombia*

Santiago Pérez-Cardona†

Abstract

I study the impact of rebel governance on economic development in rural Colombia. In 1998 the Colombian government created a 42,000 square km demilitarized zone (DMZ) to negotiate with FARC, Colombia's largest and oldest rebel group. Using a spatial regression discontinuity design, I exploit the DMZ's border defined by municipalities' pre-existing administrative boundaries to examine the causal effects of rebel based social order on education, living conditions, and agricultural production. I show that rebel governance increased the years of education by 0.1 standard deviations, access to aqueduct systems by 11 percentage points, and agricultural yield by 16 percent. These findings appear to be driven by public goods provision and less exposure to violence during rebels rule. However, I find that the positive gains from rebel governance did not translate into better living standards.

Keywords: Rebel Rule, Violence, Civil War, Development.
JEL Classification: O15, N46, D74.

*I am incredibly grateful to Juan Sebastian Galán for his support and guidance. For detailed comments and insights, I am especially grateful to Leopoldo Fergusson, Luis Martínez, Felipe Valencia-Caicedo, Germán Orbegozo, Adriana Camacho, Nicolás de Roux, Santiago Torres, Rafael Torres, Lucas Marín, and Natalia De Vivero. For valuable comments and suggestions, I am grateful to Gabriel Lombo, Andrés Dávila, and Jorge Caputo. I thank Margarita Gáfaro and the Center for Production and Sectorial Trade Studies (CEPCO) at the Central Bank of Colombia for generously sharing the data on agricultural products prices. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. All remaining errors are my own.

†Universidad de los Andes, s.perez1@uniandes.edu.co.

¿Dejar que los rebeldes gobiernen? Evidencia sobre los efectos económicos de la gobernanza rebelde en Colombia*

Santiago Pérez Cardona†

Resumen

Este artículo estudia el efecto la gobernanza rebelde sobre el desarrollo económico rural en Colombia. En 1998, el gobierno colombiano creó una zona desmilitarizada (DMZ) de 42.000 kilómetros cuadrados para negociar con las FARC, el grupo rebelde más grande y antiguo del país. Utilizando una metodología de regresión discontinua espacial se explota la frontera del DMZ, definida por los límites administrativos pre-existentes de los municipios, para examinar los efectos causales del orden social rebelde sobre la educación, las condiciones de vida y la producción agrícola. Lo resultados muestran que el gobierno rebelde aumentó los años de educación en 0.1 desviaciones estándar, el acceso a los sistemas de acueducto en 11 puntos porcentuales y los rendimientos agrícolas en un 16 %. Estos resultados parecen deberse a la provisión de bienes públicos y a menor exposición a la violencia durante el gobierno de los rebeldes. Sin embargo, se encuentra que las ganancias positivas de la gobernanza rebelde no se tradujeron en mejores condiciones de vida, medidas por la calidad de las viviendas.

Palabras clave: Gobernanza rebelde, violencia, conflicto civil, desarrollo.
Códigos JEL: O15, N46, D74.

*Estoy increíblemente agradecido con Juan Sebastián Galán por su apoyo y orientación. Por sus detallados comentarios y opiniones, estoy especialmente agradecido con Leopoldo Fergusson, Luis Martínez, Felipe Valencia-Caicedo, Germán Orbegozo, Adriana Camacho, Nicolás de Roux, Santiago Torres, Rafael Torres, Lucas Marín y Natalia De Vivero. Por sus valiosos comentarios y sugerencias, agradezco a Gabriel Lombo, Andrés Dávila, y Jorge Caputo. Agradezco a Margarita Gáfaro y al Centro de Estudios sobre Producción y Comercio Sectorial (CEPCO) del Banco de la República por compartir generosamente los datos sobre precios de productos agrícolas. Esta investigación no recibió ninguna financiación específica de ningún organismo público, comercial o sin ánimo de lucro. Todos los errores son de mi entera responsabilidad.

†Universidad de los Andes, s.perez1@uniandes.edu.co.

1 Introduction

Internal civil wars have been the most common type of war since the 1950s (Kalyvas and Balcells, 2010); by 1990, nearly one of every four countries in the world had an active civil conflict (Blattman and Miguel, 2010). However, far beyond the direct implications of violent conflict, armed groups have shaped the development path of local communities under their territorial control, especially in developing countries. Furthermore, when imposing social order, armed groups typically permeate economic activities, public goods provision, collective action, and institutions (Arjona, 2016). Yet, little is known about how armed actors' governance affects economic development at the micro-level.

The effect of rebel governance on local development has been largely overlooked. The available literature on the economic consequences of civil conflict is mainly concentrated on the negative impacts of violent shocks (Blattman and Miguel, 2010). While armed actors deploy various violent and non-violent strategies during civil conflict, destructive events (e.g., murders, kidnapping, terrorist attacks) are easier to measure and, thus, disproportionately account for most empirical evidence. However, the emerging literature on rebel governance, especially in political science, has documented how armed actors often provide social order in a wide variety of settings, from Latin America to Africa or the Middle East (Arjona, 2016).

Yet, whether rebel-based social order fosters economic development is still an open question. On the one hand, enforcing property rights and providing security are challenging tasks for governments in the developing world. Thus, armed actors could increase welfare by providing basic social order (i.e., safety and taxation) (Sanchez de la Sierra, 2020). Furthermore, armed actors frequently permeate social life beyond security and taxation. For example, ethnographic evidence suggests that The Eritrean People's Liberation Front (EPLF) promoted health care, education, and land reforms (Pool, 2001), and the Shining Path in Peru provided policing and organized recreational events (La Serna, 2012). Altogether, these interventionist social order might also affect local development.

On the other hand, there are multiple channels by which rebel governance could neglect economic development. First, rebel governance often implies violent disputes over territorial control, which has proven adverse effects on economic growth (Abadie and Gardeazabal, 2003). Second, armed actors' permanent control of areas might prevent communities from receiving public investment. Third, evidence from gang territorial control in El Salvador suggests the armed groups' rule of law decreases welfare because of substantial limitations to labor mobility (Melnikov et al., 2020). Furthermore, unlike the relationship between the civilians and the government under modern democratic regimes, armed groups' social order is typically imposed on civilians. Therefore, the absence of a political accountability mechanism could promote rent-seeking institutions and worsen economic development (Acemoglu et al., 2001).

In this paper, I leverage a historical event where rebels governed within a well-defined area in rural Colombia to study its persistent effects on local development. Three features of this study are noteworthy. First, I'm able to study a clear and explicit setting of rebel governance. This contrast with the existing literature that concentrate mostly on the direct impact of violence, considering that armed actors provision of social order is often hard to identify and measure. Second, granular data allows me to provide micro-level evidence on how rebel governance affects rural households, as most available evidence concentrates on aggregated data at the district or country level. Finally, by measuring outcomes twelve years after the exposure to rebel governance, I'm able to shed light on persistent mechanisms, which are particularly relevant for local development in the long run.

The historical events at the center of my study are the 1998 peace negotiations between the Central Government of Colombia and the Revolutionary Armed Forces of Colombia (*Fuerzas Armadas Revolucionarias de Colombia* - FARC). As part of the peace talks, the government demilitarized nearly 42,000 km of Colombia's territory - an area close to the size of Switzerland (see Figure 1). Without any presence of police and military personnel, FARC entered the demilitarized zone (from now on, the DMZ or *El Caguan*) in early 1999 and imposed their own social order until 2002, when peace talks ended without an agreement and conflict was resumed. I argue that FARC actively governed inside the DMZ since they were the only actor allowed to remain armed. During the DMZ, they restricted mobility, imposed taxes, and provided public safety, infrastructure, and justice (Espinosa and Ruiz, 2001; Espinosa, 2010; Reyes, 2012; CNMH, 2017b).

To study the causal effect of FARC rule on local development, I use granular data from Colombia's 2014 Rural Census (CNA). The CNA provides detailed information on living conditions and agricultural production for Colombia's entire rural population, which was about 5 million people. Using the CNA, I can combine nearly 1.5 million household records with restricted information of their exact geographic location. Since I can determine each household's exact distance to the boundary of the FARC rule, I can identify the effect of exposure to rebel rule using a Spatial Regression Discontinuity Design. To define the DMZ, the government used the pre-existing administrative division of Colombia's territory; the DMZ border perfectly matched the boundaries of the existing municipalities. Therefore, I argue that, when delimiting *El Caguan*, the government arbitrarily allocated some peasants to live inside the DMZ -where FARC ruled- and others to live just outside. Furthermore, I show that areas designated as part of *El Caguan* are similar from sites just outside of the boundary on geographic and basic demographic characteristics.

I study the DMZ's effect on three dimensions of local development. First, I examine human capital by computing the impact on literacy, years of schooling, and access to public health. Second, I use dwelling quality and access to public conveniences to study households' material welfare. Third, I consider variables related to agricultural production, such as yield, revenues, and crop specialization.

Contrary to the overall adverse effects of violent conflict on development, my results do not suggest that FARC governance negatively affected development in any of the dimensions I study. Instead, I find that rebel rule had modest positive effects on education, access to public services, and agricultural production. In particular, my results suggest that the DMZ increased the years of schooling by 0.4, which is about 0.1 standard deviations (SD). Furthermore, dwelling characteristics indicate that the DMZ had no impact on the quality of the household's walls and floor materials. Yet, when studying access to public conveniences, I find that the DMZ increased access to sewerage and aqueduct systems by 2 percentage points (p.p.) and 11 p.p., respectively. Finally, results on agricultural production show that the DMZ increased farmers' evenue per hectare by 16%, total revenue by 35%, and made farmers more likely to specialize in perennial crops rather than transitory crops. I then verify that these results are robust to an extensive battery of RD specifications, measurement error, inference assumptions, and the presence of selective migration.

I use both ethnographic evidence and additional empirical exercises to understand these results better and assess theoretical mechanisms. First, I follow Arjona (2016)'s typology on *rebelocracy* and historical records from the DMZ to understand why FARC decided to permeate social order with an interventionist strategy, i.e., providing public goods, regulating day-to-day activities, handling civilian complaints and misdemeanors. I suggest that they use this strategy to control the territory better and as a learning environment to share alternative forms of rebel governance within military fronts. Hence, by adopting the functions of an interventionist state, FARC was able to construct roads, bridges (El Tiempo, 2003), water tanks, and schools (CNMH, 2017a). I hypothesize that these infrastructure improvements might explain the positive results on access to public services and agricultural yields.

Second, I consider violence reduction as an alternative mechanism to explain my results. There is extensive evidence that shows how violence negatively affects welfare (Arias et al., 2014; Abadie and Gardeazabal, 2003). Since FARC had total control of the territory (i.e., *de facto* and *de jure*), I argue that the lack of armed competition made violence unnecessary. Thus, lower levels of violence might explain the positive effects on education and agricultural production. These results are consistent with empirical evidence on the positive effects of violence decrease on education (Prem et al., 2021) and agricultural investment (de Roux and Martínez, 2021). Accordingly, I show that the positive impact on schooling is only driven by age cohorts exposed to the DMZ during their schooling age, which is consistent with seminal work on the high returns of human capital investments during childhood. (Heckman, 2006; Heckman and Kautz, 2012).

Third, I use variables on property rights, agricultural practices, and collective action to study FARC's agrarian agenda as a complementary mechanism to explain the effect on agricultural production. I don't find evidence that FARC significantly affected property rights or agricultural practices. Yet, additional results show modest positive effects on capital investment (i.e., machinery and infrastructure), which might explain positive yields. Furthermore, I find that exposed farms are less likely to be vertically integrated within producers associations while more likely to use collective work. These results are consistent with historical evidence suggesting that FARC promoted collective action at the local level between peasants and shed light on the persistent effects of informal institutions from rebel governance.

Finally, I use fiscal outcomes at the municipal-year level to test for a disproportional increase of government investment in the former DMZ's municipalities as an alternative explanation. I also study educational inputs, such as pupil-teacher ratio and schools construction, to study diferential educational policies as an alternative explanation. However, I do not find evidence supporting the differential public investment of educational policy hypothesis after the DMZ ended.

Overall, it is worth noting that the modern democratic social order is not the relevant counterfactual to understand the results. Instead, I hypothesize that in the absence of the DMZ, households would have been exposed to higher levels of violence in the context of limited State presence. For instance, while I find positive effects on education and agricultural revenues, they do not translate into better living standards (i.e., dwelling characteristics). Thus, far from suggesting that rebel governance fosters economic development, my results indicate that it can mitigate the adverse effects of violent conflict.

This paper contributes to the existing literature on civil conflict for several reasons. First, while many studies have developed theoretical frameworks to understand insurgent governance (Berman et al., 2011; Bueno de Mesquita, 2013; Arjona et al., 2015; Arjona, 2016; Peñaranda Currie et al., 2021), causal evidence on their impacts has been difficult to obtain. I add to this literature by casually examining the consequences of rebel rule on local development and empirically testing the existing theoretical and ethnographic evidence.

Furthermore, this paper also contributes to the growing empirical evidence on criminal governance's effect on development, where consensus remains absent. For example, Melnikov et al. (2020) finds that El Salvador's gang rule negatively affects education, material well-being, and income by restricting mobility. In contrast, Sanchez de la Sierra (2020) shows how armed actors create social order with low uncertainty and violence and, therefore, foster positive effects on the welfare in Eastern Congo. I further expand this literature by presenting additional evidence to support the claim that criminal governance could positively affect development. Moreover, whereas the existing literature has focused on the short-term effects of criminal governance, this paper sheds light on its persistence effect over the long run.

My results also speak to a broader literature studying civil conflict and its effects on development. There is extensive evidence that shows how civil wars cause persistent adverse impacts on relevant factors for development: education (Fergusson et al., 2018; León, 2012), health (Bundervoet et al., 2009), agricultural production (Arias et al., 2019; de Roux and Martínez, 2021), collective action (Orbegozo, 2021), and growth (Abadie and Gardeazabal, 2003). Yet, most available work studies conflict where armed actors actively engage in a violent confrontation over political, economic, and social control. I complement this literature by assessing the impact on education, material welfare, and agricultural yields of non-violent settings within a civil conflict.

Finally, I also take the first steps towards empirically studying the developmental effect of a historical event of significant magnitude. This itself is an important contribution. *El Caguan* was the longest-lasting DMZ in Colombian history and significantly impacted how the civil conflict unfolded during the 2000s. Moreover, while a vast literature of qualitative and ethnographic evidence has documented *El Caguan*'s DMZ (e.g. Espinosa and Ruiz, 2001; Cadena, 2004; Espinosa, 2010; Gonzalez, 2012; Reyes, 2012; CNMH, 2017a), this is the first paper to study the effect of the DMZ on economic development empirically.

This paper is organized as follows. Section 2 describes the historical background of Colombia's civil war and *El Caguan*'s demilitarized zone. Section **??** presents the primary data sources and the construction process of the final analytical sample. Section 3 presents the empirical strategy and the identification assumptions. Section 4 outlines the main findings on developmental outcomes. I then test the robustness of my results to a set of complementary empirical exercises in section 6. In section 5 I discuss potential mechanisms behind my results. Finally, section 7 presents the concluding remarks.

2 Historical Background

Colombia's armed conflict dates back to the bipartisan violence between liberals and conservatives during the mid-20th century (Bushnell, 1993). While national parties resolved their political differences in the late '50s, peasants in remote areas remained armed due to agrarian disputes and political participation restrictions. By the early '60s, the central government's feeble State capacity allowed communist peasant guerrillas to control significantly large rural areas, called *independent republics* (Lopez-Uribe and Sanchez, 2018).

In 1964, the central government launched a military operation against the *Marquetalia Republic*, an enclave of communist peasant guerrillas in the southwest of Colombia. A group of survivors of around 350 men, led by Manuel Marulanda and inspired by the Cuban Revolution, founded the Revolutionary Armed Forces of Colombia (FARC) in May 1966. Their tactics went from communist civil disobedience to active mobile guerrilla attacks on the government. By 1978, FARC had expanded to nearly 1,000 men and deployed ten different battlefronts throughout Colombia. According to Rangel (1999), by providing social order, justice, and security, FARC earned the public's support in most regions where they were present. Moreover, in the absence of the State's rule of law, FARC was a suitable substitute (Medina, 1990).

Although the government had demobilized most guerrilla groups by the early '90s, FARC kept extending their military and political power. They took advantage of coca leaf growing production by taxing cocaine's supply chain. Additionally, the government had held peace talks on various occasions, allowing the guerrilla to get national and international political recognition (Velez, 2001).

2.1 *El Caguan*'s DMZ

By 1998, every attempt to demobilize FARC had failed and its rapid growth of power made military defeat unfeasible. Naturally, the public broadly favored negotiation over military defeat. Presidential candidate Andrés Pastrana exploited peace as his central political platform and was elected into office in June 1998. As president-elect, he met with FARC's leadership and agreed to start official negotiations after being signed in (see Figure B1).

In October 1998, Pastrana issued Resolution No. 85, which: i) officially started peace talks with FARC, ii) recognized FARC's political origins, and iii) demilitarized five municipalities[1]: La Uribe, Mesetas, Vistahermosa, La Macarena, and San Vicente del Caguan. Nearly 87,500 civilians lived under FARC's rule for more than three years, from November 1998 through February 2002.

The DMZ municipalities are part of a region in which FARC has always had a strong presence. In the mid-20th century, peasants from the Andean mountains had colonized the area to own land. The late colonization implied the State's absence and the active presence of FARC. By the time the DMZ started, the State and FARC were disputing territorial control over the region (CNMH, 2017a).

By demilitarizing *El Caguan*, the central government dictated the retreat of all military and police personnel. As a result, FARC settled in the DMZ on the 7th of January, 1999[2].

[1]Municipalities are Colombia's public administration smaller unit.

[2]Before arriving at the DMZ, FARC demanded in December 1998 the withdrawal of the last 92 admin-

According to a San Vicente resident: "Overnight we saw how the public forces left and how the guerrilla entered the town [...] We had no choice but to accept what was happening" (Reyes, 2012). While civilian staff from the mayor's office were the only public officials left in the area, the government did not restrict whether FARC could carry military equipment, allowing them to have full *de facto* control.

Qualitative and ethnographic evidence related to *El Caguan* suggest FARC completely governed inside the DMZ (e.g. Cadena, 2004; Espinosa and Ruiz, 2001; Reyes, 2012; Espinosa, 2010; El Tiempo, 1999; Gonzalez, 2012). So much so that Colombia's highest administrative court has ruled on various occasions that the State's absence violated the population's rights by failing to protect inhabitants of the area from FARC (El Universal, 2013). By permanently settling inside the DMZ and highly restricting mobility from the inside, FARC secured its position in *El Caguan*. FARC's top-ranking officials and nearly 4,000 militiamen lived and patrolled the DMZ between 1999 and 2002 (Semana, 2002).

During the first year of the DMZ, FARC moved to remove further State officials and social leaders from the area. To replace the judiciary system, FARC expelled the Attorney Generals Office's prosecutor in early 1999 (Caracol, 1999). FARC also campaigned against social leaders that questioned their actions. In mid-1999, they asked San Vicente's parish priest to leave the region (El Tiempo, 2002a). This strategy allowed FARC to replace State's presence inside the DMZ actively. They installed multiple complaints and claims offices, called "Oficinas", where the public could report civil disputes and requests to the FARC's high-ranking officials (Arjona, 2016; Espinosa and Ruiz, 2001).

FARC also provided public safety. As the replacement of police and military personnel, the government had designed a *civic* police force conformed by the civil population from each municipality in the DMZ and led by the mayor (see Figure B2). Nonetheless, FARC designated half of the *civic* police force, and ultimately they had the final say. Moreover, the *civic* police were mainly in charge of handling misdemeanors, for which the accused usually had to pay a fine at the complaints and claims offices (Reyes, 2012). As a result, the community often recognized FARC as a compelling source of justice and conflict resolution [3] (Espinosa, 2010).

Initially, the lack of infrastructure was an obstacle to govern *El Caguan* entirely. Hence, FARC promoted and financed roads between the five demilitarized municipalities to improve their access to the whole DMZ. In addition, they provided public goods by hiring and or-

istrative soldiers present in San Vicente's battalion. The government accepted the 23rd of December 1998, leaving the five municipalities of *El Caguan* without public force (Semana, 1999).

[3] For instance, according to Espinosa and Ruiz (2001), when the community couldn't resolve a dispute between two peasants in La Macarena, a woman affected stated: "How so? Is there no guerrilla? Where are the authorities (referring to FARC)?".

ganizing local communities (Gonzalez, 2012). As noted by Peñaranda Currie et al. (2021), road construction is typical where armed actors, especially FARC, provide social order.

2.2 The Aftermath

On February 20th, 2002, the government unilaterally decided to end the peace talks with FARC and retake the DMZ. While active conflict had declined dramatically during the DMZ, the government deployed nearly 4,000 troops, and combat intensified during 2002 (El Tiempo, 2002b). Even though official forces could control the municipalities' capitals, FARC's presence in the region remained constant until 2016 (see Figure B3).

Although the DMZ had legal validation, public officials often treated communities with stigma for their close relations with FARC during the peace talks. Military personnel in the area imposed additional restrictions on the civil population to limit FARC's ability to access market goods. For instance, police and military personnel restrained the trade of goods (i.e., food, fuel, personal care items) between municipalities' capitals and rural areas (CNMH, 2017b). The public force's inability to correctly differentiate between community members and under-covered FARC militia negatively affected communities' relation with the State (Semana, 2012).

3 Empirical Strategy

3.1 Data

Sample Selection and Outcomes

I use Colombia's 2014 National Agricultural Census (*Censo Nacional Agropecuario* - CNA). The CNA provides detailed information on living conditions and agricultural production of Colombia's entire rural population. I combine restricted data on households' precise GPS location from the CNA and historical maps of municipalities' boundaries in 1998 to assign whether the household lived in an area exposed to *El Caguan*.

The CNA has multiple levels of information. The central unit of observation is the Unit of Agricultural Production (*Unidad Productiva Agricola* - from now on, UPA or farm). The farm is defined by a plot of land with agricultural production and with a unique decision-making entity. If it has residents, they are organized into dwellings, and each dwelling can have one or more households. The available geographic coordinates vary at a farm level, as they correspond to the location where the interview took place, which is most likely the farm's decision-maker dwelling.

Table 1 summarizes the main variables in the CNA; columns 1-5 present basic summary statistics for the complete sample on the CNA. Column 1 shows that the CNA surveyed about 5 million people in rural Colombia during 2014, representing nearly 1.5 million households. Broadly, Colombia's rural population is relatively young, with an average age of 32 years. Almost three out of every four households self-identify as poor, and access to public conveniences is scarce. For example, only 6% and 40% of dwellings report having access to sewerage and aqueduct systems, respectively.

Columns 6-5 show the number of observations, the mean, and the standard deviation for the primary analysis sample, respectively. To construct this analysis sample, I first restrict the sample to households within 100 km buffer from *El Caguan*'s border. [4]Because *El Caguan* Next, because El Caguan is located at Colombia's southeast plains, I drop observations placed in the highlands of Colombia's Andean region. The red zone in Figure 2 illustrates the highlands area not considered in the analysis. Column 6 shows that these restrictions leave me with about 43,000 households and 133,000 people.

I then divide *El Caguan*'s border into five different boundary segments, corresponding to each municipality that made part of the DMZ (see Figure 2). Then, I assign the distance and information of the closest boundary segment to each household in my sample. In my analysis, I consider all boundary segments except for La Uribe, mainly located at the Eastern Ranges. If included, the Eastern Ranges would create a discontinuous jump on geographic, socioeconomic, and cultural characteristics, affecting the similarity of observations at each side of the boundary.

I study the DMZ's effect on three groups of outcomes. First, I examine the impact on human capital formation with a dummy variable indicating whether individuals can read and write, a variable of years of education, and a dummy variable indicating whether individuals have access to health insurance. Second, I assess whether the DMZ affected material welfare based on the dwelling's quality and access to public conveniences. Third, I consider variables related to agricultural yield, revenues, and crop specialization.[5]

Table 2 compares the primary outcome variables of individuals living in the DMZ's municipalities with neighboring municipalities. Overall, differences in column 5 suggest that individuals inside the DMZ have worse education, living conditions, and access to public services. For instance, Panel B indicates that households from DMZ municipalities are 7 p.p. more likely to self-identify as poor and have a higher probability of reporting exposure to land-related violence. Panel D suggests that farms inside the DMZ have higher revenues

[4]I do this mainly for computational reasons. Overall, my RDD specification optimal bandwidth ranges between 6 km and 26 km.

[5]To account for multiple hypothesis testing concerns, I present Romano and Wolf (2005)'s step-down adjusted p-values for the main results.

and are more specialized in cash crops, requiring higher investment levels.

Additional Data Sources

I use additional data sources to measure geographic characteristics, exposure to armed conflict, and migration patterns across exposed and non-exposed municipalities. I use data on elevation and rainfall between 1980 and 1989 from the World Clim Organization, and data on land suitability for agriculture and average cropland in 1992 from the UW - Madison's Atlas of the Biosphere. Data on the geographic location of rivers comes from official information of Colombia's National Mapping Institute - IGAC and information on conflict events by armed actors come from Violent Presence of Armed Actors in Colombia (ViPAA). To study migration patters between 2000 and 2005, I use Colombia's 2005 General Census. I also use data on fiscal outcomes from Colombia's National Planning Department (DNP) at the municipality-year level, to study differential investment by government post-DMZ as an alternative explanation for my results. Data on communities perceptions towards armed actors comes from the ELCA 2010, which I employ to study mechanism. Finally, I use data on educational inputs, such as teachers and schools, to study a differential educational policy after the DMZ. This data is provided by the Colombian Minister of Education and varies at the municipal-year level.

3.2 Research Design

Because exposed and non-exposed municipalities to the DMZ most likely differ on observable and unobservable characteristics, a simple regression of developmental outcomes on exposure would yield a biased estimation of the causal effect of interest. For instance, even before the DMZ started, exposed municipalities had had a more substantial FARC presence than their neighboring municipalities. Thus, I exploit the discontinuous change in exposure to *El Caguan* by comparing households located near the arbitrarily defined border of the DMZ.

The boundary forms a multi-dimensional discontinuity in a longitude-latitude space. Thus, my baseline specification approximates the location of each farm with a local linear polynomial of the geodesic distance to the DMZ border estimated separately on each side of the boundary.[6] Therefore, I can estimate the effect of exposure to the DMZ on

[6] Although a wide variety of empirical work has used the spacial RDD approach, there is no consensus on the best way to approximate the location in the regression analysis. While some researchers use latitude and longitude as running variables in a multi-score RDD setting (e.g. Dell, 2010; Grosfeld et al., 2013; Dell et al., 2018), others have favored a *normalizing-and-pooling approach* using the Euclidean distance to reduce the multi-dimensional space to a single running variable (e.g. Miguel and Roland, 2011; Michalopoulos and Papaioannou, 2013; Melnikov et al., 2020; Lowes and Montero, 2020).

the outcomes of interest with the following spatial regression discontinuity design (RDD) specification:

$$y_{i,u,v,s} = \alpha + \tau Caguan_u + \gamma Dist_u + \theta Dist_u \times Caguan_u + \phi_s + \psi X_{i,u,v,s} + \varepsilon_{i,u,v,s} \quad (1)$$

$y_{i,u,v,s}$ denotes the outcome for individual i from farm u in rural district (*vereda*) v assigned to boundary segment s. $Caguan_u$ is a variable indicating whether the farm u is located inside the DMZ. $Dist_u$ is the geodesic distance to the DMZ border, which controls for smooth functions of geographic location. ϕ_s are nearest boundary segment fixed effects. And $X_{i,u,v,s}$ is a covariates vector, which includes a dummy of sex, age, age-squared, and the farm's extension.

The coefficient of interest is τ, which captures the local causal effect of being just inside *El Caguan* on the outcome of interest. Because FARC rule was more vital at the municipalities' capitals -often located in the center of each municipality- this local intention-to-treat effect presumably underestimated the average treatment effect. Intuitively, I argue that when delimiting *El Caguan* the government arbitrarily allocated some peasants to live inside the DMZ -where FARC ruled- and others to live just outside the DMZ. Both groups presumably had similar geographical, cultural, historical, and institutional characteristics before the DMZ started. Therefore, an RDD design in this context allows me to identify the DMZ's effect on recent developmental outcomes.

I cluster standard errors at a rural district level. Accordingly, as benchmark, I use a linear distance polynomial function, triangular kernel weighting, and calculate the optimal bandwidth for each outcome variable following Calonico et al. (2019)'s one common MSE optimization. For completeness, I document robustness to using alternative approximations to standard error structures, the degree of the distance polynomial function, kernels weighting, and bandwidths (see Appendix D).

3.3 Threats to Identification

The RDD approach presented in equation (1) requires two key identifying assumptions. First, it needs the potential outcomes of all relevant factors to varying smoothly on the boundary before the DMZ[7]. Under this assumption, observations just outside of the boundary are a good approximation to those observations' counterfactuals just inside the boundary.

The main concern for identification is that the government chose the DMZ location strate-

[7]Let K_0 and K_1 denote the potential outcomes under treatment and control, respectively, d denote the linear distance to the boundary, and ϕ_m the boundary segment fixed effects. Identification requires that $E[K_0|d, \phi_m]$ and $E[K_1|d, \phi_m]$ are continuous at the discontinuity threshold.

gically for specific characteristics that could affect the outcomes of interest. For instance, the DMZ might have been located in strategic zones for cocaine production. Nonetheless, when defining the DMZ, the government didn't choose a specific area but instead used the already defined boundaries of existing municipalities. Thus, the government did not select the DMZ boundaries based on particular characteristics but Colombia's preexisting administrative division. Furthermore, I argue that what mattered when assigning the DMZ was each municipality's capital. Thus, the rest of the municipal area comes as an extension of the capital, making it presumably a *good as random* assignment for households near the edge.

I assess this assumption's plausibility by checking for the balance of geographic and essential demographic characteristics at both sides of the boundary. Table 3 presents mean differences and RDD estimates using specification (1) on these features. As most of these outcomes come from satellite information, results on geographic variables in Panel A and Panel B are at the 10 km by 10 km grid cell level. Demographic variables in Panel C come from CNA and vary at an individual level.

Consistent with the first identification assumption, I find balance in all the geographical characteristics. While unconditional mean differences in columns (3) and (5) yield statistically significant coefficients, the RDD approach in column (7) shows non of the variables in Panel A-B show significant differences statistically and in magnitude relative to the mean. Because these characteristics are pre-determined to the DMZ, results suggest areas are comparable along the boundary.

Demographic characteristics in Table 4 are also balanced around the border. Differences in these variables would suggest fundamental demographic changes along the border, threatening the sample comparability at each side of the boundary. For instance, significant differences in age or sex composition could be explained by the degree of exposure to civil conflict, as men and younger individuals are more likely to be recruited into armed organizations. Yet, results in column 7 imply that these demographic characteristics vary smoothly around the border, suggesting demographics are comparable along the boundary.

Second, the RDD approach requires that residents located inside the DMZ didn't selectively migrate from those areas to neighboring locations outside the boundary. This assumption guarantees that control observations were not directly exposed to the DMZ. During and after the DMZ, some migration likely took place, but ethnographic evidence suggests migration was mainly from the DMZ's municipalities and to major cities, such as Villavicencio and Bogotá (CNMH, 2017a).

Unfortunately, there is no data available to measure migration during *El Caguan* directly. Nonetheless, I provide three complementary exercises that suggest migration hardly explains my results. First, Figure 4 tests for a discontinuous change in population density at the boundary. If households inside the DMZ were more likely to migrate to neighboring municipalities, I would expect to see a decrease in household density at the boundary. Yet, results in Figure 4 do not support such a claim. Second, I use Colombia's 2005 General Census to suggest that migration patterns between the DMZ and neighboring municipalities are similar. Figure 3 shows the migration flow from DMZ municipalities to neighboring municipalities in my analysis. Descriptive results suggest migration to neighboring municipalities represented less than 10% of all migration between 2000 and 2005. Third, Section 6 presents a trimming to estimate what percentage of the sample I would have to drop for selective migration to explain my results fully.

4 Impacts on Economic Development

4.1 Human Capital

Table 5 presents the results of estimating equation (1) on human capital formation, using data from the 2014 rural census. For each outcome, I show both the RDD coefficient when using the optimal bandwidth suggested by (Cattaneo et al., 2020a) and the RDD coefficient for a fixed bandwidth of 10 km, which allows me to have a stable sub-sample to compare results across outcomes. To improve precision, I control for boundary segment fixed effects, age, age square, and sex. Columns (1) and (2) show a null impact on literacy; the coefficient is not only imprecise but very close to zero, with a point estimate of -0.01 for average literacy rate of 84%. Nonetheless, columns (3) and (4) show that the DMZ positively impacted the years of education. Twelve years after the *El Caguan*, I estimate the DMZ increased the years of schooling by 0.4; this accounts for an increase of 0.1 SD and 5% of the mean. Finally, columns (5) and (6) study the effect of health insurance, showing a small (1% of the mean) and non-significant effect. I present the respective RD plots for these outcomes in Figure 6.

4.2 Material Welfare

I now investigate whether rebel governance affected households' material welfare. To do so, I proxy welfare with available dwelling characteristics in the 2014 rural census. For each outcome, I show both the RDD coefficient when using the optimal bandwidth and a fixed bandwidth of 10 km and control for boundary segment fixed effects. Table 6 shows results on dwelling characteristics and access to public services. While columns (1) and (2) show results on the likelihood of having adequate walls materials (concrete or better), columns (3) and (4) show results on the probability of having adequate floor materials (concrete or better). I show the DMZ had no significant impact on neither of these outcomes. Although columns (5) and (6) document no effect on the probability of having electricity, columns (7) and (9) show a positive and statistically significant impact on the likelihood of having a sewerage and aqueduct system, respectively. I estimate that exposed areas are 2 percentage points more likely to have a sewerage system and 11 percentage points more likely to have an aqueduct system; both effects account for at least 100% of the sample mean. I present the respective RD plots for these outcomes in Figure 7.

4.3 Agricultural Production

Next, I look at agricultural production outcomes. Since I use Colombia's rural census, my population of interest are farmers for which agricultural production is their primary source of income. For each outcome, I show both the RDD coefficient when using the optimal bandwidth and a fixed bandwidth of 10 km and control for boundary segment fixed effects and the farms' total area.[8] I first study whether rebel governance had a persistent impact on agricultural yields and revenue. To do so, I proxy yield with a crop-specific measure of revenue per hectare, which I discuss in detail in Appendix A. I then examine whether the production is for the household's subsistence (self-consumption) or market sale.

Table 7 shows results on agricultural yields, revenue, and the production final use. The dependent variables in columns (1)-(2) and (3)-(4) are the inverse hyperbolic sine of total revenue and revenue per hectare, respectively, and the coefficient of interest can be interpreted as a semi-elasticity.[9] Column (1) shows that the DMZ led to an average increase of approximately 16% in the revenues per hectare, a proxy for yield,. This effect is precisely estimated (statistically significant at the 1% level) and hardly changes when using the fixed bandwidth of 10km in column (2). Similarly, results in columns (3) and (4) illustrate that the DMZ increased total revenue by at least 35% (statistically significant at the 1% level). Conversely, column (5) suggests that the DMZ increased the likelihood of using agricultural production only for self-consumption (i.e., subsistence) by nearly 12 percentage points. This result is significant at a 5% confidence level and is equivalent to 28% of the sample mean. While results in columns (7) and (8) on the likelihood of using agricultural production only for market sale are not statistically significant, they suggest a negative association with the DMZ. I present the respective RD plots for these outcomes in Figure 8.

Finally, I compare differences in crop choices by looking at how likely are farmers to specialize in cash crops, perennial crops, and transitory crops. Crop choices can reflect farmers'

[8]When studying the effect on agricultural yield and revenue, I run the regression at the farm-crop level and include crop fixed effects.

[9]I use arcsinh transformation on monetary variables because it approximates the natural logarithm while retaining zero-valued observations. For further details see Bellemare and Wichman (2020).

attitudes toward uncertainty or investment constraints. For example, cash crops require high levels of investment and vertical integration within the production chain. Similarly, Figure 5 shows that perennial crops have higher agricultural yields. Studying the effect on crop specialization is essential for two reasons. First, empirical evidence suggests that civil conflict distorts agricultural decisions such that producers prefer less profitable crops (Arias et al., 2019). Thus, resulting in a potential mechanism of poverty traps. Second, these distortions can persist over time and affect local long-run development (Brück and Schindler, 2009).

Table 8 presents the results on crop choices. I classify crops into three categories: cash crops, perennial crops, and transitory crops. Following Montero (2021), I define a crop as cash crop when it requires centralized processing to be valuable, and it cannot be directly consumed by an individual worker. For perennial and transitory crops, I use the official classification provided by DANE. Appendix A offers detailed information on crops classification. For example, column (1) to (4) shows that the DMZ led to a significant decrease on cash crops specialization. Column (1) illustrates that 12 years after rebel governance, exposed farms had 7% less area dedicated to cash crops, relative to an average share of cash crop area of 33%. Likewise, results in column (3) show that the DMZ led to an 11% decrease in the participation of cash crops in total revenue. I present the respective RD plots for these outcomes in Figure 9.

Columns (5) to (12) illustrate an accordant pattern across results: exposed farms are more likely to specialize in perennial crops and less likely to specialize in transitory crops. Column (5) and (7) show that the DMZ led to a 7% increase on the share of area dedicated to perennial crops and a 13% decrease on the participation of perennial crops on total revenue, respectively. Both results account for at least 10% of the sample mean. Moreover, columns (9) to (12) show that the DMZ had an opposite effect on transitory crops. Column (11) shows that the DMZ led to a 10% decrease on the participation of transitory crops on total revenue.

5 Mechanisms and Discussion

The results in Section 4 raise the intriguing question of how the DMZ, and more specifically FARC governance, positively affected economic development at the local level. This section draws from the historical background reviewed in Section 2 and presents additional empirical exercises to explore the theoretical mechanisms that could explain my main findings. Overall, I find two central mechanisms that explain my results. First, FARC invested in infrastructure (i.e., roads, bridges, water tanks), which explains the positive effects on public services and agricultural yields. Second, the DMZ significantly reduced violence, positively affecting education and promoting specialization in more profitable crops.

5.1 An Interventionist Rebel Rule

As noted in Section 2, FARC broadly intervened in the social order during the DMZ by providing security, infrastructure, dispute resolution, taxation, and regulating day-to-day life. Thus, the first question worth answering is why they decided to permeate social order. To answer this question, it is worth noting that the first implication of the DMZ was that FARC didn't have to fight for territorial control. The presidential decree that formalized *El Caguan* provided both *de facto* (i.e., public forces retreated) and *de jure* military control over the DMZ. Furthermore, the pooling together of FARC personnel and high-ranking commanders towards the DMZ, facilitated high levels of internal discipline. These two conditions likely made FARC "more likely to operate under long time horizons, establishing a social contract with the local population" (Arjona, 2016, p. 10).

Under this long time horizon, it follows from the theory of state formation that FARC had incentives to provide a minimal condition on government: monopoly of violence (Hobbes, 1948; Weber, 1922). Moreover, empirical evidence on state formation and armed conflict suggests that armed groups take over the essential functions of the state -monopoly of violence, taxation, and protection of property rights- to better appropriate local revenues (Sanchez de la Sierra, 2020). Ethnographic evidence suggests FARC co-opt the civil police, imposed restrictions on individuals' freedom of movement and patrolled roads and rivers to maintain territorial control (CNMH, 2014).

Yet, security provision for territorial control within the DMZ can't fully explain why FARC decided to adopt the functions of an interventionist state (e.g., infrastructure, a justice system, rules of conduct) rather than resembling a minimalist one. Therefore, I provide three complementary reasons that explain this decision in the context of *El Caguan*.

First, an interventionist strategy facilitated territorial control. For instance, FARC's military control over the DMZ resulted from a voluntary retreat from the public force; this increased the likelihood of a government military offensive, which was especially threatening as most high-ranking officials were permanently living in the DMZ (Espinosa and Ruiz, 2001). Thus, by intervening in most aspects of social order, FARC could directly monitor most activities inside the DMZ and fortify their control over the area.

Second, it allowed FARC to further modify social, political, and economic institutions in a territory where they had been present historically. As noted in Section 2, the DMZ was located where FARC had been present since the late 60s. Thus, to some extent, they recognized this territory as their own and expected to continue making a presence on it if the peace talks failed (CNMH, 2014). Consequently, the DMZ provided the opportunity to shape life in this territory without resistance from the central government. Moreover, by becoming a reliable source of social order in the region during the DMZ, they would secure the public's support in the future.

Third, the DMZ was an experimental setting to introduce various forms of rebel governance that FARC had used for decades in the territories under their control. As noted by Rangel (1999), FARC had been imposing local social order across Colombia since the 80s. Nonetheless, they did so in various ways across military fronts (CNMH, 2014). Thus, the DMZ became a suitable setting to share learned lessons on rebel governance strategies across commanders (Espinosa, 2016). In fact, in 2000 -a year after the DMZ started- FARC produced two formal documents -called FARC's Law 002 and 003- which included guidelines on taxation and administrative corruption (CNMH, 2014). Accordingly, after the DMZ, FARC fronts started using *Community Handbooks* to handle civilian affairs. Thus, these elements suggest that FARC leveraged the DMZ to formalize and unify rebel governance practice across fronts.

All together, FARC operating under a long time horizon and adopting an interventionist state could explain my findings to some extend. First, the construction of communal water tanks, documented in Section 2, could explain the positive effects on access to sewerage and aqueduct systems. Second, the construction of roads across the DMZ municipalities reasonably explains why the DMZ led to higher revenues per hectare in agricultural production. For instance, El Tiempo (2003) suggests that FARC build nearly 140 km of roads within the DMZ, significantly improving transport times across local rural markets. Finally, (CNMH, 2017b) also documents that FARC also promoted the construction of schools, which could explain the positive effects on years of education.

I use data on communities' perceptions towards armed conflict from the ELCA 2010 to provide suggestive evidence on the relationship between civilian populations exposed to the DMZ and FARC.[10] I compare attitudes towards the armed conflict in Vistahermosa with available communities in the same sub-region (i.e., Eastern Region). Tables B1-B4 present the results of this analysis. Due to the small number of observations and exposed communities, I use randomization inference to estimate statistical significance from my comparisons. While Table B1 suggests that communities in Vistahermosa have historically been more exposed to armed actors' presence, Table B2 shows that this presence is more likely to be monopolized by just one actor, especially at the end of the DMZ. These results are consistent with ethnographic on the monopolistic control of FARC within the DMZ.

Nonetheless, column (4) in Table B1 shows that the higher presence of armed actors

[10]ELCA surveyed four communities from Vistahermosa, a municipality exposed to the DMZ.

doesn't translate to higher perceptions of insecurity. This result might indicate that communities get used to the company of armed actors and don't perceive them as drivers of insecurity, especially in context with monopolistic presence. Furthermore, column (3) in Table B1 shows that communities in Vistahermosa have a higher probability of submitting to the will of armed groups, implying that armed actors do permeate social order. Finally, results in Table B4 do not show differential victimization patterns related to violent attacks, forced displacement, or taxation.

5.2 Violence

Another potential explanation mechanism that could explain my results is that the DMZ led to a substantial violence reduction. Figure 10 shows an appreciable reduction of armed conflict events between 1999 and 2000 within the DMZ municipalities. This conflict de-escalation is likely to result from the DMZ, as it entailed less competition over territorial control (Arjona et al., 2015). Moreover, anecdotal evidence discussed in Section 2 suggest FARC could control the DMZ without major adversaries; thus, the lack of armed competition made violence unnecessary.

To empirically assess whether the DMZ led to less violence related to conflict, in Table 9, I estimate equation (1) on the likelihood of being exposed to forced displacement, land dispossession, and land abandonment. Column (1) shows that the DMZ decreases the probability of land displacement by 10 p.p., representing a reduction of nearly 60% relative to the mean at the 99% confidence level. Accordingly, columns (2) and (3) show that the DMZ led to a decrease of 4 p.p. on the likelihood of being exposed to forced land dispossession and 3 p.p on reporting land abandonment. These estimates are highly precise and consistent with ethnographic evidence on violence reduction during *El Caguan*. Moreover, Figure 11 plots the year of victimization of forced displacement and land dispossession across the DMZ and neighboring municipalities. Note that victimization rates during the DMZ are consistently lower in DMZ municipalities for both variables. I find this evidence convincing that the DMZ led to a reduction in violence.

Less violence could explain the positive effect on the years of schooling, and it would be consistent with empirical evidence on the positive impact of violence decrease on education (Prem et al., 2021). If this is the case, one would expect that cohorts exposed to the DMZ during their schooling age would be the ones driving the effect. In Figure 12, I empirically test this hypothesis by separately estimating the RDD coefficient on three different subsamples: age cohorts that had their schooling age before the DMZ started, age cohorts that had their schooling age during the DMZ, and age cohorts that had their schooling age after the DMZ ended. Overall, results are consistent with this hypothesis and suggest that only exposed cohorts are driving the effect. Moreover, null effects on cohorts with schooling age after the DMZ suggest positive results were not persistent after the DMZ ended.

Furthermore, results on agricultural production could also be explained by lower levels of violence. Both theoretical and empirical evidence suggests that civil conflict has a strong negative effect on agricultural production, especially as it leads to lower levels of investment (Arias et al., 2014). As noted before, I argue that the DMZ reduced conflict victimization relative to neighboring territories. Then, I would expect farmers outside the DMZ to be more likely to have experienced active civil conflict. Thus, these farmers would prefer activities with short-term yields and lower profitability from activities that require high investments, as pointed out by Arias et al. (2019) and de Roux and Martínez (2021). Therefore, the positive effects on agricultural yield and revenue suggest that in the absence of the DMZ, farmers inside would have been exposed to high levels of violence and, thus, experiencing a lower-income trajectory. Moreover, this might also explain why farmers inside the DMZ are less likely to specialize in transitory crops and more likely to specialize in perennial crops.

5.3 Property Rights, Agricultural Practices and Collective Action

An alternative explanation to why the DMZ led to higher agricultural yields might be that FARC intervened in farmers' agricultural practices. For now, I have argued that better infrastructure (i.e., roads, bridges, river ports) and less violence could explain these positive effects. Nonetheless, FARC also had a solid agrarian agenda which they pushed forward during the DMZ. For instance, FARC's Law 001 (called "*The Revolutionary Agrarian Reform*") banned farmers from having estates of over 1,500 hectares, recognized property rights to small peasants under a lease, and promoted *de facto* occupations of private property and local agrarian collective action under the figure of local *Committees for Agrarian Reform* (CNMH, 2014).

Results on crop specialization are consistent with FARC's agrarian agenda. For example, FARC's agricultural practices favor local small farmers over corporate production. If so, this could explain the adverse effects on cash crops, which require high levels of investment, large estates, and vertical integration within the production chain. To further understand the role of FARC's agrarian policy on development outcomes, I estimate equation (1) on intermediate outcomes related to property rights, agricultural practices, capital investment, and partnership structure. In Table 10, while column (1) shows a small (relative to the sample mean) and imprecise effect on the probability of the farm being privately owned, column (2) suggests that the DMZ decreases the likelihood of having land of collective property.[11] In columns (3) to (5), I find no evidence that producers inside the DMZ were more likely to use irrigation systems, fertilizers, and professional pest control techniques. Next, in columns (6) and (7), I study whether exposed producers are more likely to have agricultural machinery and buildings. Results suggest a somewhat noisy increase of 7 percentage points in the likelihood of having machinery (e.g., harvester, power plant) and no effect on having buildings for agricultural use (e.g., greenhouse, barn). Finally, columns (8) to (10) show the estimates on producers' partnership structure. Whereas column (8) suggests that the DMZ decreased the likelihood of joining a formal producers association by 6 percentage points, results in column (10) show a statistically significant increase of 5 percentage points on using collective work.[12]

How should one think about my main results in light of these findings? First, notice that the negative effect on collective property suggests that institutional mechanisms might persist over time. At first, this result might seem counter-intuitively, as FARC follows a Marxism–Leninism doctrine. Yet, their rural agenda moved towards privately own property for small farmers. Second, the null results on agricultural practices reflect that FARC's agrarian policy didn't include elements that would improve agricultural operation. Third, positive effects on capital investment could respond to indirect incentives. Since FARC was primarily concerned with large unproductive estates, they promoted capital investments to enhance agricultural production. For instance, the *Revolutionary Agrarian Reform* stated that estates with more than 1,500 hectares would not be expropriated if they would "introduce modern agricultural technologies" (CNMH, 2014). Finally, results on partnership structure provide evidence of the persistent effect of informal institutions. These results suggest that the DMZ promoted informal collective action between peasants while counteracting formal vertical integration in agricultural production.

5.4 Alternative Explanation: Differential Public Investment

Differential public investment after the DMZ ended in early 2002 could be a potential alternative explanation to the positive results on public services, years of education, and agricultural revenues. However, this would imply that the central and municipal governments significantly increased expenditure in the DMZ municipalities. Therefore, I use fiscal outcomes at the municipal-year level to test for a disproportional increase of government in-

[11] Note that this variable has a very low frequency, with a sample average of 0.01. Thus, one has to interpret this result with caution, as the RD estimate might be sensitive to very rare events near the boundary.

[12] Collective work is defined as an informal labor relation where neighboring peasants collaborate during agricultural production.

vestment in the former DMZ's municipalities. To do so, I follow Abadie (2021) and estimate a synthetic control analysis that provides an alternative counterfactual for the evolution of fiscal investment across the DMZ municipalities. I study both transfers from the national government to the municipalities and municipal investment (i.e., capital expenditure).

Panel (a) in Figure C1 shows the total transfers from the national government to the municipalities. The solid line corresponds to the actual total transfer per 10,000 inhabitants, while the dashed line shows the prediction from the synthetic control. I observe that the synthetic control closely follows the realized time series up to 2000 and exceeds it afterward. While the synthetic control keeps growing, the actual series drops and stagnates. Placebo inference in panel (b) suggests this difference is not statistically significant (Abadie, 2021). Figure C5 show a similar pattern for municipal investment, meaning there is no significant difference between the synthetic control and the actual value observed. Moreover, figures C2-C4 show similar patterns on transfers from the central government for educational expenses, health expenses and free disposable revenues for the municipal government. I find this compelling evidence that a differential investment does not explain my results after the DMZ.

Yet, fiscal spreadsheets on revenues and expenses only show the funds flow toward specific categories but might be misleading as they do not show how the municipal government specifically spent money. Alternatively, I study educational inputs such as teachers and schools infrastructure. While Figure C6 shows synthetic control results for pupil-teacher ratio, Figure C7 shows similar results for the number of schools per 1,000 students. Altogether these results do not suggest that the DMZ municipalities provided more educational inputs than nearby non-DMZ municipalities.

6 Robustness

In this section, I provide an extensive battery of additional robustness tests available in Appendix D. First, I present robustness to alternative RDD bandwidths, polynomial functions, and kernel weightings. Second, I reproduce my results using a donut regression discontinuity design for accounting for measurement error on the distance to the DMZ. Third, I estimate my results using alternative assumptions when computing standard errors. Fourth, I use a sample trimming exercise to account for possible selective migration within the DMZ boundary. Overall, results discussed in Section 4 are robust to these robustness checks.

Alternative RDD Specifications

It could be the case that my main results only exist for a particular regression discontinuity specification. To examine the robustness of the results to alternative RDD specification choices, in Appendix D.1 I present the main results using additional bandwidths (i.e., 50%, 75%, 125%, and 150% of the optimal bandwidth), alternative RDD polynomials (i.e., quadratic and cubic), and varying the kernel choice for the estimation weighting (i.e., epanechnikov and uniform). Although some of the results (e.g., years of education, sewerage, and cash crops specialization) are sensitive to bandwidth selection as they lack the proper sample power, the point estimate is generally relatively constant across specifications. Overall, the results are highly robust to the degree of the RDD polynomial and the kernel choice.

Donut RDD

Given that I only observe the geographic coordinate within each estate where the interview took place and not the exact extend of the property, it might be the case that the farms very close to the border actually cross the DMZ's boundary. To alleviate further concerns over this possible measurement error in the independent variable, I use a Donut Regression Discontinuity Design. These modified RDD regressions involve dropping all data points within a specific distance of the threshold.[13] By construction, Donut RDD induces bias into my estimates because it compares less similar observations across the distance to the DMZ. Nonetheless, since it drops observations that are more likely to cross the DMZ's boundary, it helps me determine to what extent measurement error affects my results.

In Appendix D.2, I present the main results using Donut RDD. Overall, point estimates become more prominent as the Donut RDD hole increases. For instance, Column (6) in Table D5 suggests that the DMZ increased 1.4 years of education relative to an estimated increase of 0.4 years using the standard RDD approach. Altogether, Donut RDD results indicate that the main results on Section 4 are conservative due to classical measurement error on the independent variable, which biases my result towards zero.

Alternative Variance-covariance Matrix Assumptions

Considering that anecdotal evidence suggests that the rural district (*vereda*) groups similar individuals and that most of the day-to-day life in rural Colombia takes place at this level, my baseline specification uses Nearest Neighbor clustered (NN-cluster) standard errors at

[13] See Dahl et al. (2014); Canaan and Mouganie (2018); Zimmerman (2014); Melnikov et al. (2020) for similar applications of Donut RDDs.

the rural district level. Yet, I can't be sure if the error term correlates at this exact level. Therefore, in Appendix D.3, I present the main results using alternative structures for the error term.[14] In Tables D9 to D12, column (1) presents baseline results, column (2) uses NN-clustered standard errors at the municipality level, column (3) uses clustered standard errors at the rural district level, column (4) uses clustered standard errors at the municipality level, column (5) uses clustered standard errors within 250 meters bins of the distance to the boundary, and column (6) uses clustered standard errors within 500 meters bins of the distance to the boundary. Overall, both significant levels and point estimates are robust to alternative variance-covariance matrix assumptions.

Trimming for Selective Migration

As noted in Section 3.3, one possible alternative explanation for my results is that the DMZ led to selective migration across the boundary. Under this setting, the most capable individuals left neighboring places outside the DMZ by migrating to the former DMZ area. Thus, the area inside the DMZ appears more developed. In Appendix D.4, I follow Lowes and Montero (2020) in conducting a trimming exercise to empirically assess the magnitude of the selective migration to explain my results fully. I first estimate a simple factor model to produce a wellness score for human capital, dwelling characteristics and access to public services.[15] Based on this score, I then drop the $x\%$ of the most well-off individuals and reproduce my main results. By trimming the $x\%$ of the sample, I tried to determine what percentage of the most well-off individuals who reside inside the DMZ would I need to omit so that I don't observe statistically significant differences between former DMZ non-DMZ areas. I reproduce this process for yield, revenue, and crop specialization outcomes, but I use each continuous variable distribution instead of a factor model score.

Tables D13 to D16 show the main results if I trim the top 2.5, 5, 10, 15, and 20% of the most well-off individuals and farms. For the most part, my results suggest that after trimming for the top 5% of the sample inside the DMZ, the results remain of similar magnitude and statistical significance. However, for a higher trimming rate of 10%, coefficients become non-statistical significance while having similar magnitude; I hypothesize this reflects a loss of power induced by the smaller sample. Furthermore, coefficient's magnitude and statistical significance become increasingly noisy when trimming at least 15% of the sample inside the DMZ. Overall, these results suggest that at least 10% of the sample inside the DMZ had to reflect selective migration to fully explain the differences I observed. As a point of

[14] When using alternative error term structure, I allow the bandwidth to varying across regressions.

[15] I use a factor model to create a continuous measure of wellness because all of these variables are dummies.

reference, using Colombia's 2005 General Census, I estimate that between 2000 and 2005, the likelihood of migrating from a DMZ neighbor municipality to the DMZ was 14%. Thus, selective migration had to account for 2/3 of this migration pattern for my results to be only explained by selective sorting across the boundary. Therefore, these estimates demonstrate that my results remain consistent even under a strong assumption of high levels of selective migration.

Spatial Placebo Analysis

I now perform a placebo analysis using arbitrary alternative cutoffs for the RDD estimates. If the impacts reflect the effects of the DMZ on developmental outcomes, I wouldn't expect to find a discontinuous jump on the outcomes if I arbitrarily move the cutoff to a point inside or outside the DMZ. I present the results of this placebo exercise in section D.5. Overall, some results are sensitive to alternative placebo cutoff, while others are robust. For instance, Panel (b) in Figure D1 shows that I find null results for years of education in 3 out of 4 alternative cutoffs, while I find significant impacts in the actual cutoff. Nevertheless, Panel (a) in Figure D3 shows that 2 out of 4 placeboes yield statistically significant results when studying total revenues, which raises concerns on the validity of the original estimates. Further analysis is required to understand the sensitivity of some outcomes to this placebo test.

7 Conclusion

This paper examines the persistent effect of rebel governance in rural Colombia. By exploiting geographic discontinuities on the exposure to rebels rule, I empirically assess the impacts of rebels' social order on local economic development. My results provide evidence that rebel governance didn't negatively affect development in any of the dimensions I studied. Instead, I find modest positive effects on education, access to public services, and agricultural production.

I draw from historical and ethnographic evidence on the DMZ to explore potential mechanisms. First, I find that FARC took an interventionist approach to social order by providing public goods, security, and justice during the DMZ. Thus, this approach might explain the positive effect on access to public services and agricultural yield. Second, I show compelling evidence that lower levels of violence during rebel governance mediate my education results. Finally, I don't find persuasive evidence that rebels' agrarian policy had a persistent effect over time. Overall, most of the positive effects I see seem to be explained by cohorts directly exposed to the DMZ or initial investments in infrastructure. Thus, it doesn't appear to be a persistent effect of rebel rule over time, as I don't evidence any significant impact on historical institutions.

Although my results are relevant to understand better the persistent effects of civil conflict and, more specifically, rebel-based social order, my results reflect on a particular form of rebel governance and might not generalize to another context. FARC governance over El Caguan resulted from the specific circumstances that the peace negotiations allowed; thus, one must be careful because of the limited external validity.

Finally, my result on the positive gains from rebel governance not translating into better living standards suggests that these households face structural constraints that may restrict local development. Thus, this paper draws attention to a relevant aspect of development policy, especially in a post-conflict scenario. Nevertheless, further research is required to understand these restrictions better and address them from a public policy perspective.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *The American Economic Review*, 93(1):113–132.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401.

Arias, M. A., Camacho, A., Ibáñez, A. M., Mejía, D., and Rodríguez, C., editors (2014). *Costos económicos y sociales del conflicto en Colombia: ¿Cómo construir un posconflicto sostenible?* Universidad de los Andes, Colombia, 1 edition.

Arias, M. A., Ibáñez, A. M., and Zambrano, A. (2019). Agricultural Production Amid Conflict: Separating the Effects of Conflict into Shocks and Uncertainty. *World Development*, 119:165–184.

Arjona, A. (2016). *Rebelocracy: Social Order in the Colombian Civil War*. Cambridge Studies in Comparative Politics. Cambridge University Press.

Arjona, A., Kasfir, N., and Mampilly, Z., editors (2015). *Rebel Governance in Civil War*. Cambridge University Press.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 82(1):50–61.

Berman, E., Shapiro, J. N., and Felter, J. H. (2011). Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq. *Journal of Political Economy*, 119(4):766–819.

Blattman, C. and Miguel, E. (2010). Civil War. *Journal of Economic Literature*, 48(1):3–57.

Brück, T. and Schindler, K. (2009). The impact of violent conflicts on households: What do we know and what should we know about war widows? *Oxford Development Studies*, 37(3):289–309.

Bueno de Mesquita, E. (2013). Rebel Tactics. *Journal of Political Economy*, 121(2):323–357.

Bundervoet, T., Verwimp, P., and Akresh, R. (2009). Health and Civil War in Rural Burundi. *The Journal of Human Resources*, 44(2):536–563.

Bushnell, D. (1993). *The Making of Modern Colombia: A Nation in Spite of Itself*. University of California Press, 1 edition.

Cadena, J. L. (2004). La geografía y el poder: Territorialización del poder en Colombia - el caso de las FARC - de Marquetalia al Caguán. *Estudios Poíticos (Mexico)*.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2019). Optimal Bandwidth Choice for Robust Bias-corrected Inference in Regression Discontinuity Designs. *The Econometrics Journal*, 23(2):192–210.

Canaan, S. and Mouganie, P. (2018). Returns to education quality for low-skilled students: Evidence from a discontinuity. *Journal of Labor Economics*, 36(2):395–436.

Caracol (1999). Rechazan expulsión de fiscal regional de San Vicente del Caguán. News Report.

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2020a). *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020b). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

CNMH (2014). *Guerrilla y población civil: Trayectoria de las FARC 1949-2013.* Centro Nacional de Memoria Histórica.

CNMH (2017a). *La tierra no basta: Colonización, baldíos, conflicto y organizaciones sociales en el Caquetá*. Centro Nacional de Memoria Histórica.

CNMH (2017b). Recorriendos por los paisajes de la violencia: De San Vicente a La Macarena. Short Documentary; Centro Nacional de Memoria Histórica.

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–74.

de Roux, N. (2020). Weather Variability, Credit Scores and Access to Credit: Evidence from Colombian Coffee Farmers. Documentos CEDE 017800, Universidad de los Andes - CEDE.

de Roux, N., Gáfaro, M., Mahecha, N. M., Otero, G., and Parejo, A. (2019). Información Crediticia para la Formulación de Política Pública: El Caso del Banco Agrario y la Sustitución de Cultivos de Uso Ilícito. Documentos CEDE 017409, Universidad de los Andes - CEDE.

de Roux, N. and Martínez, L. R. (2021). Forgone Investment: Civil Conflict and Agricultural Credit in Colombia. *Mimeo*, (019236).

Dell, M. (2010). The Persistent Effect of Peru's Mining "Mita". *Econometrica*, 78(6):1863–1903.

Dell, M., Lane, N., and Querubin, P. (2018). The Historical State, Local Collective Action, and Economic Development in Vietnam. *Econometrica*, 86(6):2083–2121.

El Tiempo (1999). Despeje: El imperio de las FARC. News Report.

El Tiempo (2002a). Los incidentes del Caguán. News Report.

El Tiempo (2002b). Primer mes de la retoma del Caguán. News Report.

El Tiempo (2003). Las Carreteras de la Revolución. News Report.

El Tiempo (2020). La foto que cambió la historia de Colombia. News Report.

El Universal (2013). Primera condena a la Nación por el despeje del Caguán. News Report.

Espinosa, N. (2010). El justo comunitario, las leyes y la justicia en una regióncon fuerte presencia del conflicto armado. Etnografía del pluralismo jurídico en la Sierra de La Macarena. *Díalogos de Derecho y Politica*.

Espinosa, N. (2016). La justicia guerrillera en colombia. elementos de análisis para los retos de la transición política en una zona de control insurgente (el caso del piedemonte amazónico). *Estudios Latinoamericano*.

Espinosa, N. and Ruiz, D. (2001). Caminando el despeje. *Análisis Político*.

Fergusson, L., Ibañez, A. M., and Riaño, J. F. (2018). Conflict, Education Attainment and Structural Transformation: La Violencia in Colombia. *Economic Development and Cultural Change*.

Gonzalez, M. E. (2012). Diálogos y memorias en contextos subalternos. Algunos acercamientos a memorias disidentes en un caserío rural de San Vicente del Caguán, Caquetá. *Díalogos de Derecho y Política*.

Grosfeld, I., Rodnyansky, A., and Zhuravskaya, E. (2013). Persistent Antimarket Culture: A Legacy of the Pale of Settlement after the Holocaust. *American Economic Journal: Economic Policy*, 5(3):189–226.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4):451–464. European Association of Labour Economists 23rd annual conference, Paphos, Cyprus, 22-24th September 2011.

Hobbes, T. (1948). Leviathan or the matter, forme and power of a commonwealth, ecclesiasticall and civil. edited with an introduction by michael oakeshott. *Philosophy*, 23(85).

Kalyvas, S. N. and Balcells, L. (2010). International System and Technologies of Rebellion: How the End of the Cold War Shaped Internal Conflict. *The American Political Science Review*.

La Serna, M. (2012). *The Corner of the Living: Ayacucho on the Eve of the Shining Path Insurgency*. University of North Carolina Press.

León, G. (2012). Civil Conflict and Human Capital Accumulation: The Long-term Effects of Political Violence in Perú. *The Journal of Human Resources*, 47(4):991–1022.

Lopez-Uribe, M. d. P. and Sanchez, F. (2018). On the agrarian origins of civil conflict in Colombia. *Mimeo*.

Lowes, S. and Montero, E. (2020). Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State. *Mimeo*.

Medina, C. (1990). Autodefensas, paramilitares y narcotrafico en Colombia: origen, desarrollo y conslidación. *Mimeo*.

Melnikov, N., Schmidt-Padilla, C., and Sviatschi, M. M. (2020). Gangs, Labor Mobility, and Development. *Mimeo*.

Michalopoulos, S. and Papaioannou, E. (2013). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics*, 129(1):151–213.

Miguel, E. and Roland, G. (2011). The Long-run Impact of Bombing Vietnam. *Journal of Development Economics*, 96(1):1–15.

Montero, E. (2021). Cooperative property rights and development: Evidence from land reform in el salvador. *Journal of Political Economy*, 0(ja):null.

Orbegozo, G. (2021). Consequences of violence against social leaders in Colombia. *Mimeo*.

Peñaranda Currie, I., Otero-Bahamon, S., and Uribe, S. (2021). What is the State Made of? Coca, Roads, and the Materiality of State Formation in the Frontier. *World Development*, 141:105395.

Pool, D. (2001). *From Guerrillas to Government: The Eritrean People's Liberation Front*. Eastern African Studies. Ohio University Press.

Prem, M., Vargas, J. F., and Namen, O. (2021). The human capital peace dividend. *Journal of Human Resources*.

Rangel, A. (1999). *Roconocer la Guerra para construir la Paz*, chapter Las FARC-EP: una mirada actual. Norma.

Reyes, E. (2012). Las ruinas de un experimento llamado Caguán. Master's thesis, Universidad del Rosario.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Sanchez de la Sierra, R. (2020). On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo. *Journal of Political Economy*, 128(1).

Semana (1999). Ahora si a negociar. News Report.

Semana (2002). Las preguntas. News Report.

Semana (2012). A diez años del Caguán. News Report.

Velez, M. A. (2001). FARC-ELN: evolución y expansión territorial. *Desarrollo y Sociedad*, 1(47):151–225.

Weber, M. (1922). *Economy and Society: A New Translation.* Harvard University Press.

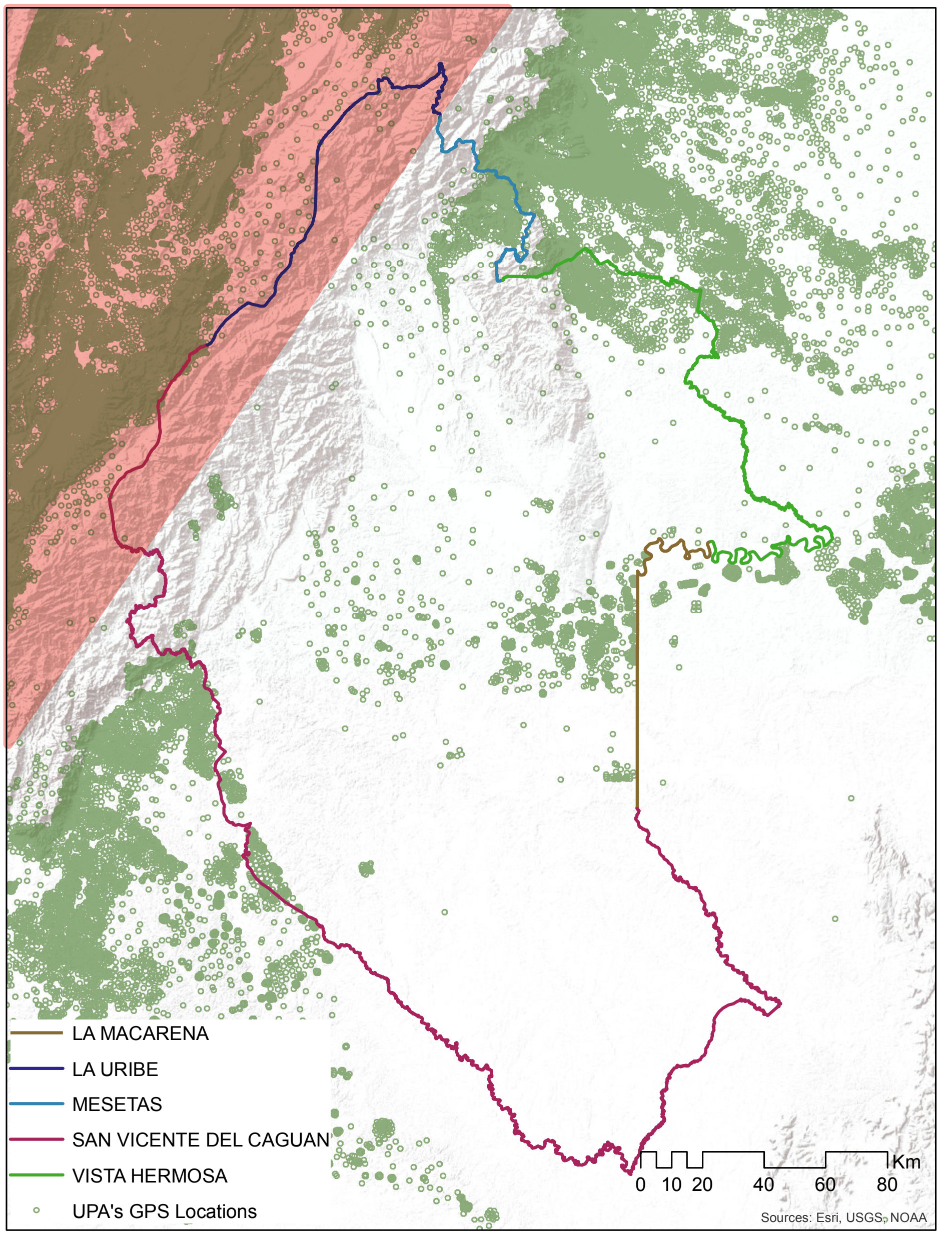
Zimmerman, S. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711 – 754.

Figures

Figure 1: *El Caguan*'s DMZ in Colombia

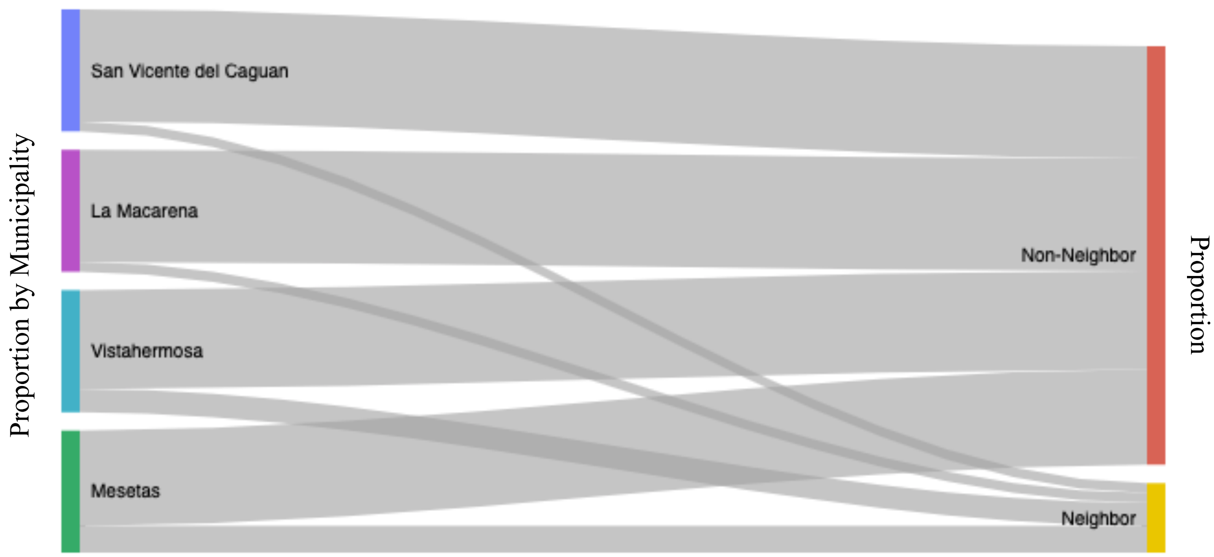
Notes: This figure plots the DMZ within Colombia's territory. The DMZ was located at Colombia's southeast region, between the Andes mountains and the Amazon.

Figure 2: Boundary Segments and farms Locations



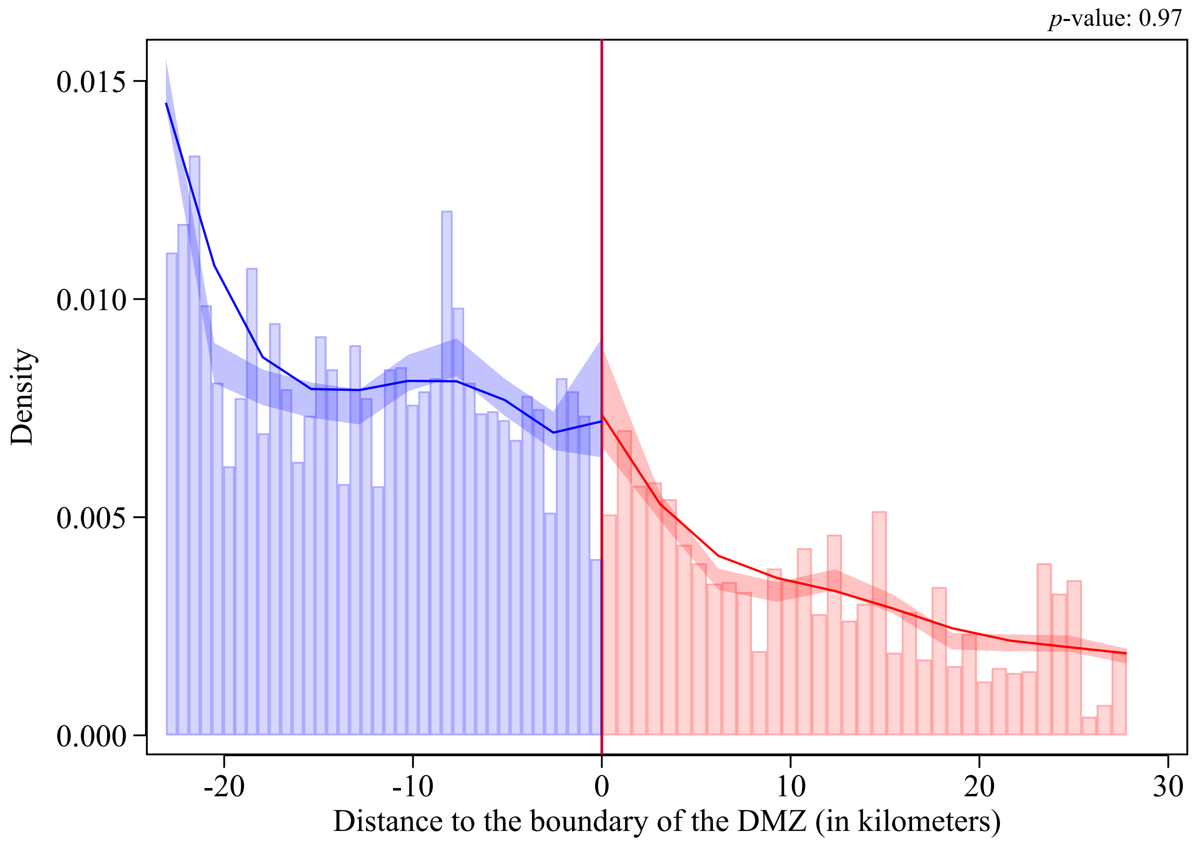
Notes: This figures plot farms located near *El Caguan*'s border. The red area corresponds to the areas located in the Andean highlands, I don't consider neighboring in the highlands, as they are most likely not comparable with those in the plains.

Figure 3: Migration from DMZ Municipalities (left) to Other Municipalities (right)



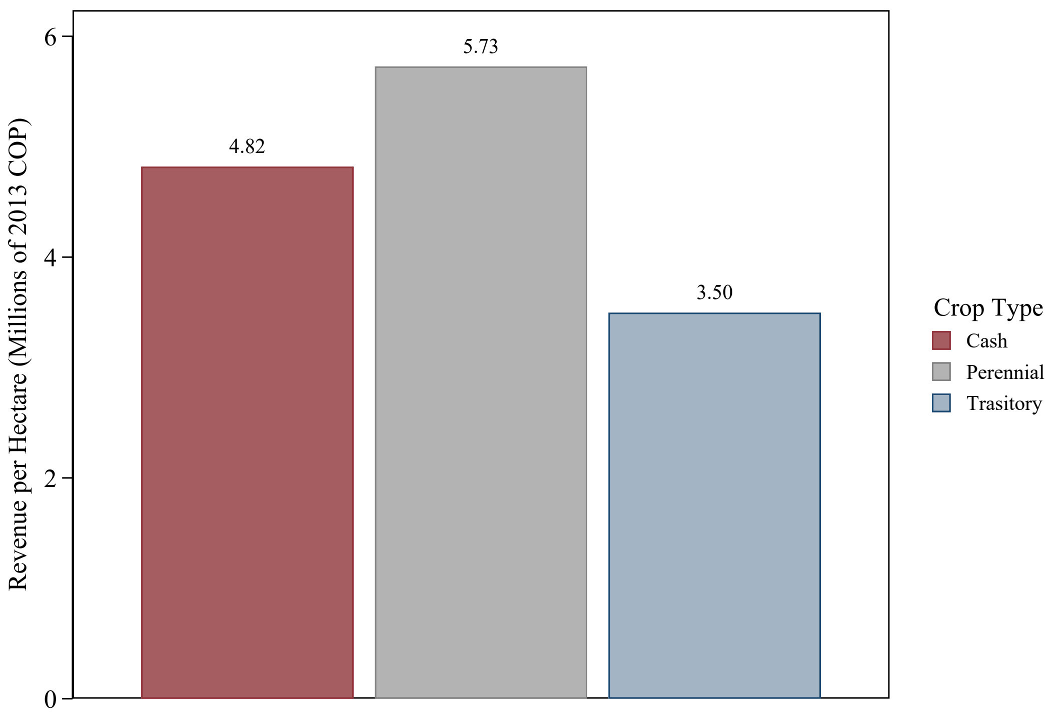
Notes: This figures plots the migration flow from DMZ municipalities (left) to neighboring municipalities (right) as a proportion of each DMZ municipality initial population. It is based on the 2005 General Census and studies migration flow between 2000 and 2005.

Figure 4: Households Density Test



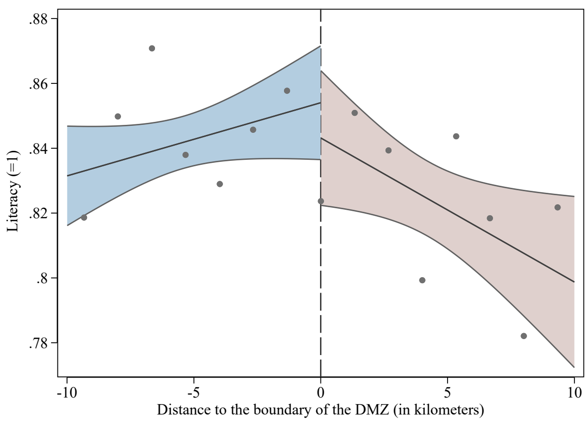
Notes: The figure shows the distribution of distance to the DMZ's boundary in kilometers. The red vertical line denotes the DMZ's border. I formally test for a discontinuity at the threshold using the Local Polynomial Density Estimators proposed by Cattaneo et al. (2020b).

Figure 5: Average Agricultural Yield by Crop Type

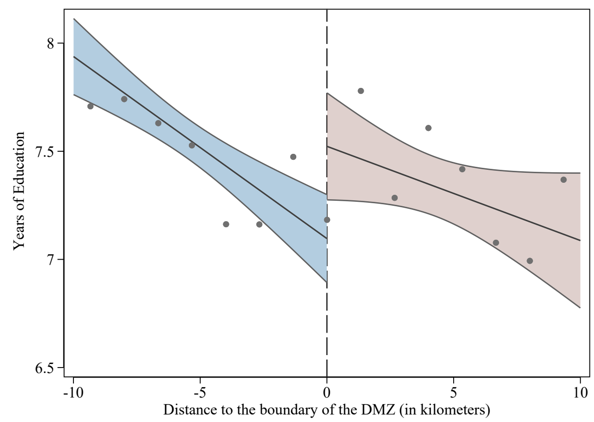


Notes: The figure shows the average revenue per hectare in the CNA. As noted in the text, perennial crops have higher yields that transitory crops.

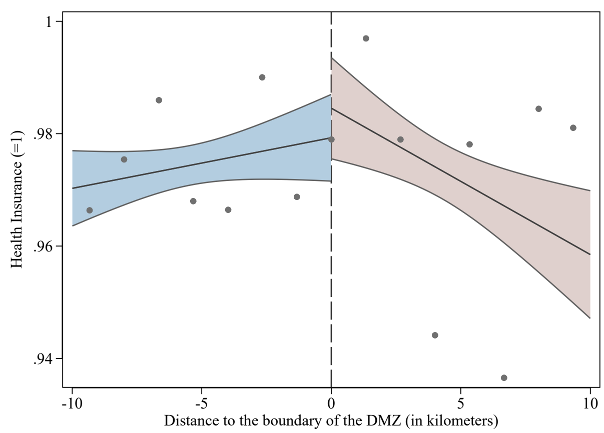
Figure 6: Human Capital: RD Plots



(a) Literacy (=1)



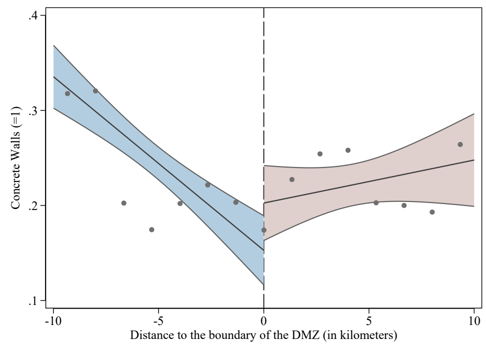
(b) Years of Education



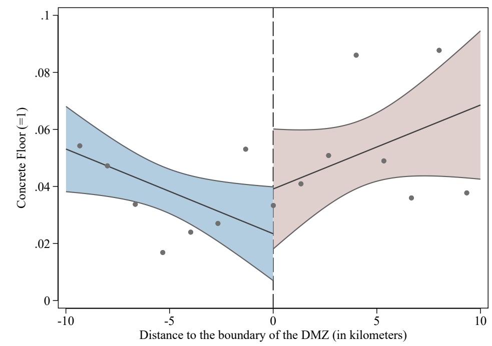
(c) Health Insurance (1)

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is an indicator variable that takes the value of one for individuals that know how to read and write, in panel (b) it is the total years of formal education, and panel (c) it is an indicator that takes the value of one for individuals with health insurance. The points represents the average value of the outcome variable in bins of width of 1.4 *km*. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. I present the corresponding estimate of τ in equation 1 in Table 5.

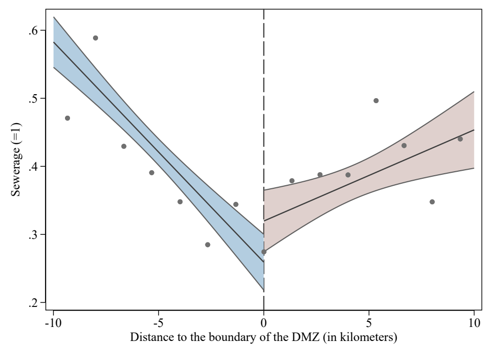
Figure 7: Dwelling Characteristics: RD Plots



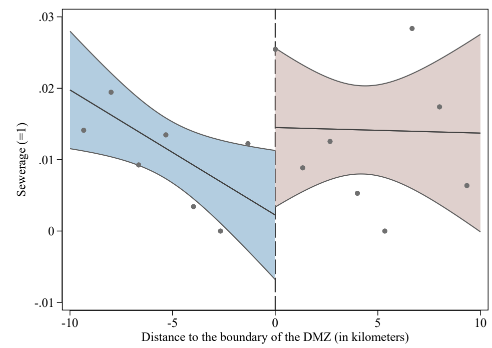
(a) Concrete Walls (=1)



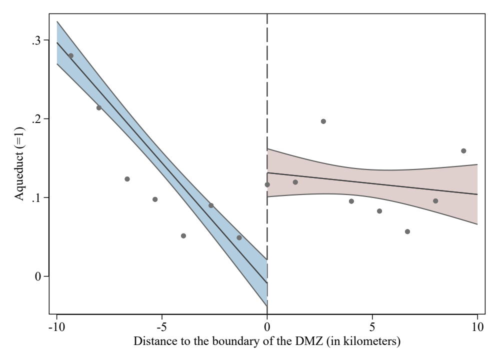
(b) Concrete Floor (=1)



(c) Electricity (=1)



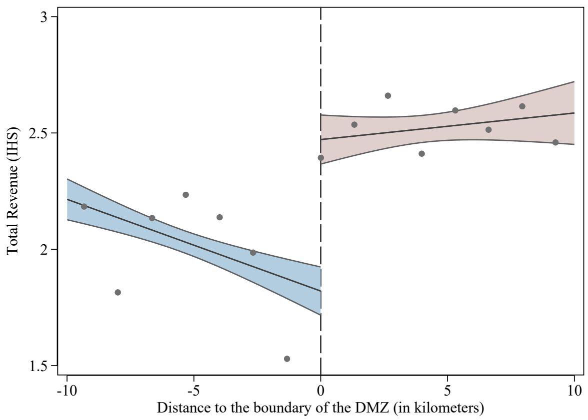
(d) Sewerage (=1)



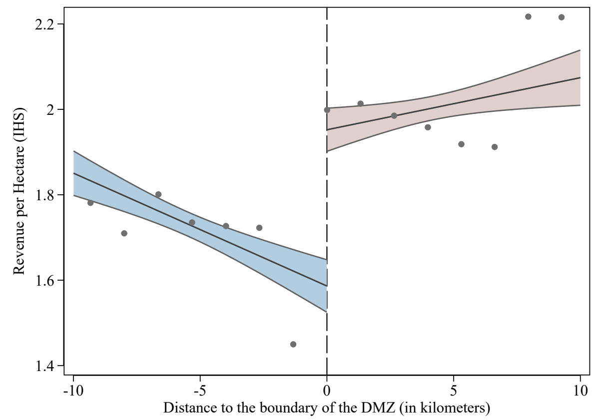
(e) Aqueduct (=1)

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in panel (b) it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in panel (c) is an indicator that equals one if the dwelling has access to electricity, in panel (d) it is an indicator that equals one if the dwelling has access to sewerage system, and in panel (e) it is an indicator that equals one if the dwelling has access to aqueduct system. The points represents the average value of the outcome variable in bins of width of 1.4 *km*. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. I present the corresponding estimate of τ in equation 1 in Table 6.

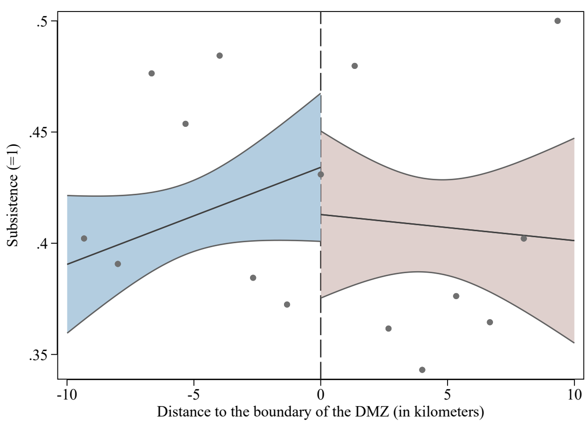
Figure 8: Yield & Agricultural Production: RD Plots



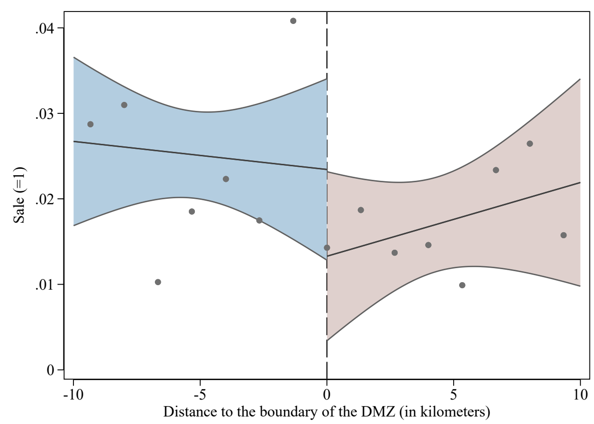
(a) Total Revenue (IHS)



(b) Revenue per Hectare (IHS)



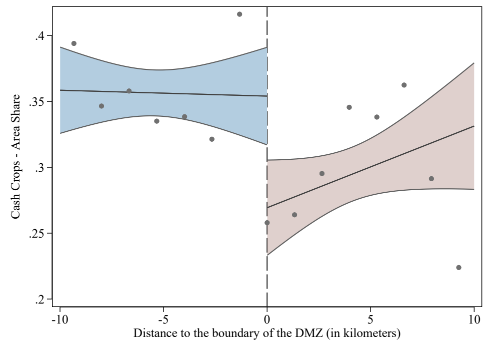
(c) Subsistence (=1)



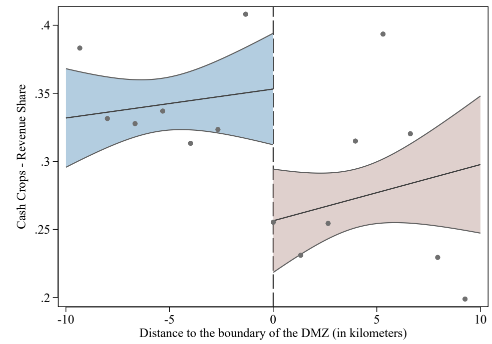
(d) Sale (=1)

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP, while in panel (b) it is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP. Dependent variable in (c) is an indicator that equals one if the farms agricultural production is used only for self-consumption, and in panel (d) it is an indicator that equals one if the farms agricultural production is used only for market sale. The points represents the average value of the outcome variable in bins of width of 1.4 km. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas.I present the corresponding estimate of τ in equation 1 in Table 7.

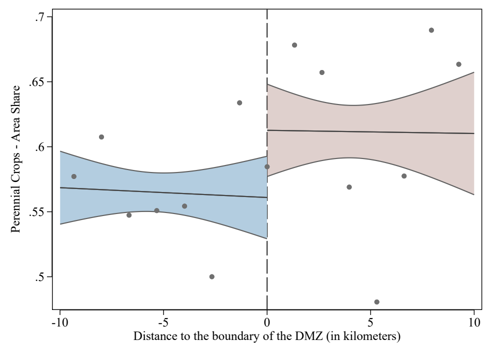
Figure 9: Crop Specialization: RD Plots



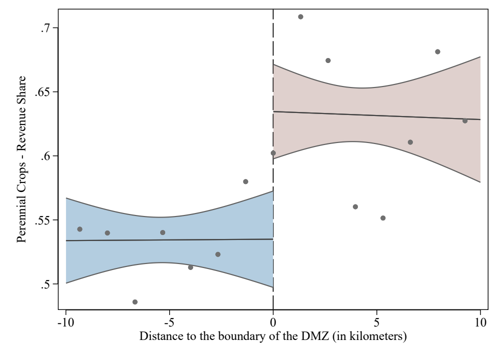
(a) Cash Crops - Area Share



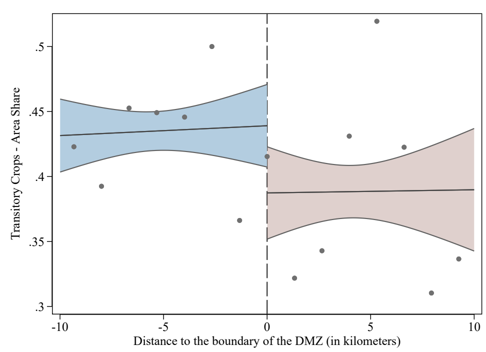
(b) Cash Crops - Revenue Share



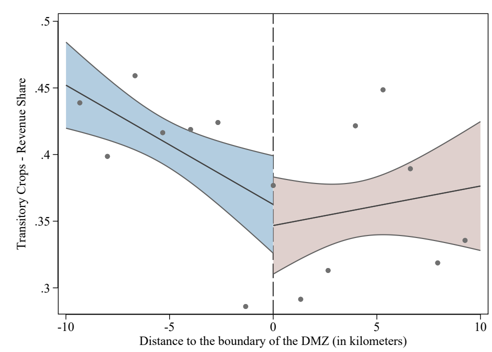
(c) Perennial Crops - Area Share



(d) Perennial Crops - Revenue Share



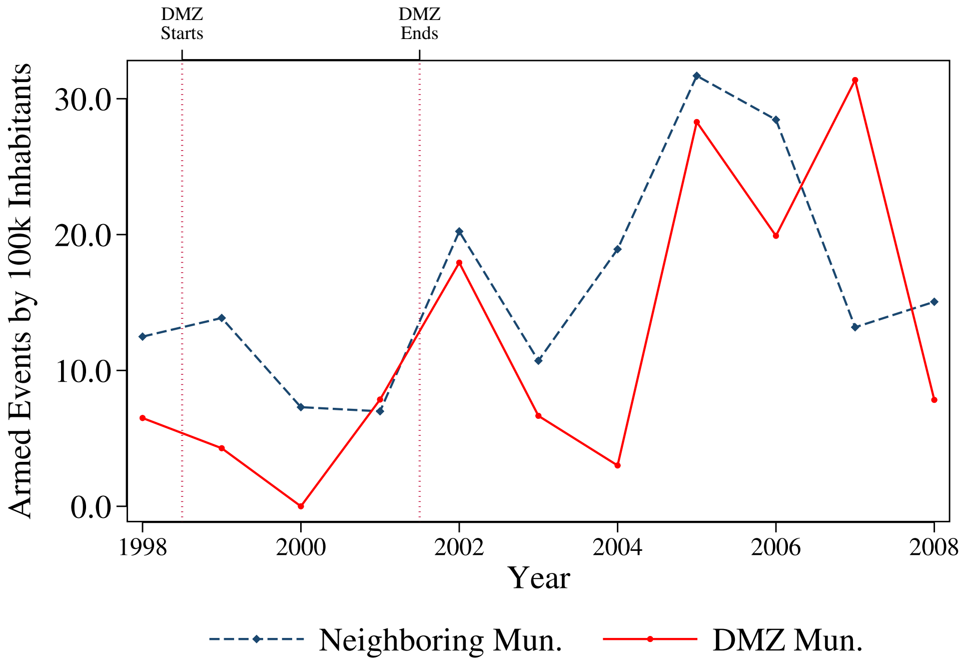
(e) Transitory Crops - Area Share



(f) Transitory Crops - Revenue Share

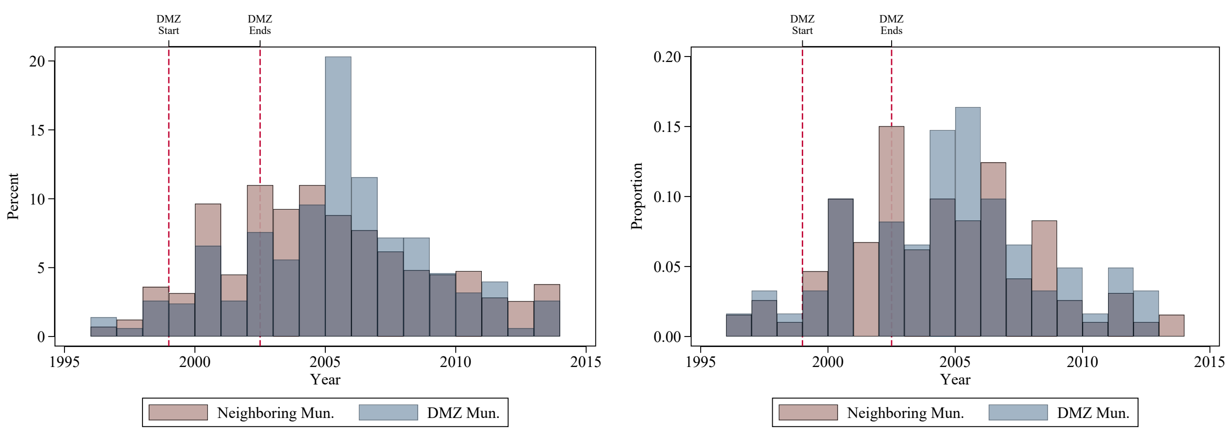
Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is the share of area with cash crops, while in panel (b) it is the share of revenue from cash crops. The dependent variable in (c) is the share of area with perennial crops, while in panel (d) it is the share of revenue from perennial crops.The dependent variable in panel (e) is the share of area with transitory crops, while in columns panel (f) it is the share of revenue from transitory crops. The points represents the average value of the outcome variable in bins of width of 1.4 km. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. I present the corresponding estimate of τ in equation 1 in Table 8.

Figure 10: Mechanism - Total Armed Events Over Time



Notes: This figure plots the average armed events by 100 thousand inhabitants over time. To compute the averages I only consider armed actors different from FARC.

Figure 11: Mechanism - Land-Related Violence Victimization Years

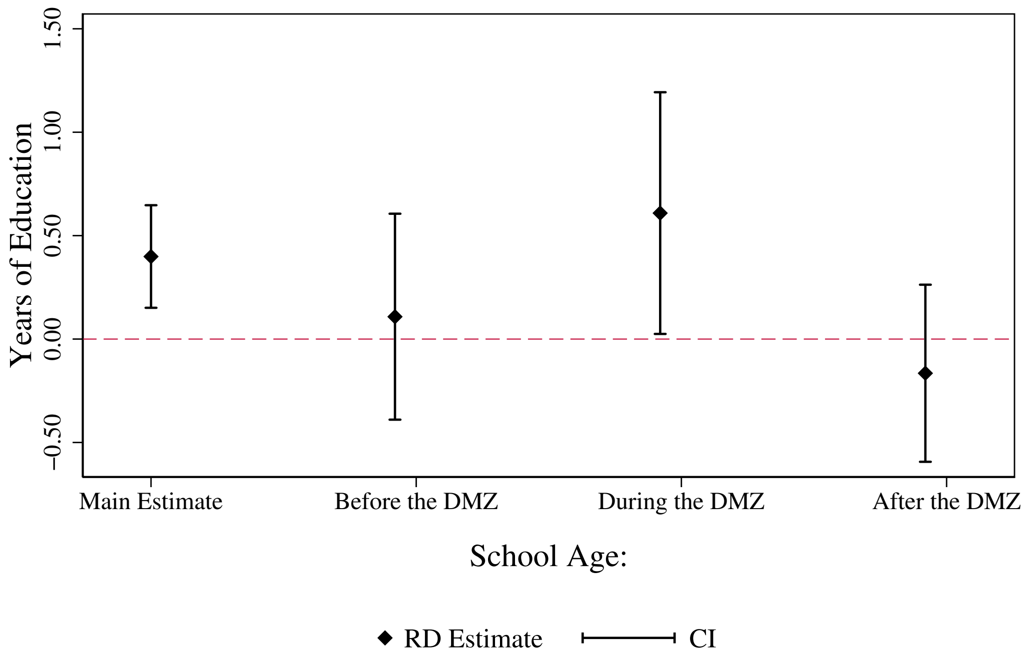


(a) Year of Forced Displacement

(b) Year of Land Dispossession

Notes: Each figure plots the year of victimization of the violent event by municipality's exposure to the DMZ.

Figure 12: Mechanism - Age Cohorts



Notes: This figure plots the RDD estimate for each age-specific cohorts. Estimates are obtain by running the RDD in equation 1 in a sub-sample for each cohort separately. Cohorts grouped in *Before the DMZ* correspond to individuals 21 years or older when the DMZ started in 1999. Cohorts grouped in *During the DMZ* correspond to individuals that were at most 20 years old in 1999 and at list 2 year old in 2002, when the DMZ ended. Cohorts grouped in *Before the DMZ* correspond to individuals 1 years or younger when the DMZ ended in 2002.

Tables

Table 1: Summary Statistics - Rural Census 2014

	Rural Census					Sample: 100 km Buffer			Inside *El Caguan*		
	Obs	Mean	SD	Min	Max	Obs	Mean	SD	Obs	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Individual-level Variables											
Men (=1)	5,126,734	0.517	0.500	0	1	133,716	0.542	0.498	13,668	0.559	0.496
Age in years	5,126,734	32.398	21.825	0	115	133,716	31.930	20.583	13,668	30.593	20.655
Natives (=1)	5,126,734	0.164	0.370	0	1	133,716	0.053	0.224	13,668	0.113	0.317
No Ethnicity (=1)	5,126,734	0.759	0.428	0	1	133,716	0.935	0.247	13,668	0.850	0.357
Knows how to read (=1)	5,126,734	0.798	0.401	0	1	133,716	0.841	0.365	13,668	0.797	0.402
Years of Education	4,649,993	7.624	4.637	0	23	121,261	8.225	4.447	12,162	7.044	4.088
Has health insurance (=1)	5,019,313	0.958	0.200	0	1	130,362	0.968	0.176	13,015	0.959	0.199
Panel B: Household-level Variables											
Self-identified as poor (=1)	1,457,519	0.679	0.467	0	1	42,415	0.598	0.490	4,409	0.657	0.475
Better now than 5 years ago (=1)	1,454,528	0.294	0.456	0	1	42,265	0.365	0.481	4,354	0.416	0.493
Internal displacement (=1)	1,543,134	0.132	0.338	0	1	43,345	0.171	0.376	4,550	0.128	0.334
Land dispossession (=1)	1,543,134	0.012	0.107	0	1	43,345	0.023	0.149	4,550	0.017	0.131
Land abandonment (=1)	1,543,134	0.014	0.117	0	1	43,345	0.026	0.159	4,550	0.021	0.143
Panel C: Dwelling-level Variables											
Concrete Walls (=1)	1,476,962	0.491	0.500	0	1	41,964	0.457	0.498	4,392	0.159	0.366
Concrete Floor (=1)	1,475,146	0.163	0.369	0	1	41,898	0.192	0.394	4,382	0.032	0.176
Electricity (=1)	1,488,807	0.833	0.373	0	1	42,578	0.590	0.492	4,519	0.336	0.472
Sewerage system (=1)	1,478,354	0.061	0.239	0	1	42,177	0.070	0.255	4,481	0.035	0.183
Aqueduct system (=1)	1,478,354	0.430	0.495	0	1	42,177	0.237	0.425	4,481	0.127	0.333
Panel D: UPA-level Variables											
Total Revenue (Millions of 2013 COP)	906,186	44.436	742.848	0	159,257	115,986	55.614	956.092	3,166	119.103	453.498
Perennial Crop (=1)	885,473	0.861	0.346	0	1	114,909	0.917	0.275	3,166	0.973	0.164
Transitory Crop (=1)	885,473	0.451	0.498	0	1	114,909	0.398	0.489	3,166	0.751	0.432
Cash Crops (=1)	906,186	0.743	0.437	0	1	115,986	0.818	0.386	3,166	0.642	0.480
Cattle per Hectare	1,340,788	0.018	0.421	0	306	163,412	0.014	0.156	5,308	0.001	0.016
Sells Crops (=1)	2,913,163	0.389	0.488	0	1	378,333	0.399	0.490	6,790	0.412	0.492
Sells Livestock (=1)	2,913,163	0.411	0.492	0	1	378,333	0.349	0.477	6,790	0.733	0.442

Notes: Columns (1) through (5) present basic summary statistics for the complete sample in Colombia's 2014 Rural Census, columns (6) through (8) present basic summary statistics for observations within a 100 km buffer around *El Caguan*'s border; and columns (9) through (11) present basic summary statistics for observations inside *El Caguan* . The number observations might differ within panels due to each question's response rate.

Table 2: Mean Differences with Neighboring Municipalities

	El Caguan		Neighbors		Difference	
	Mean	SD	Mean	SD	(1)-(3)	SE
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual-level Variables						
Knows how to read (=1)	0.79	0.41	0.83	0.38	-0.04	0.00***
Years of Education	6.91	4.08	7.55	3.98	-0.64	0.04***
Has health insurance (=1)	0.96	0.21	0.97	0.17	-0.02	0.00***
Obs.	11,744		38,369			
Panel B: Household-level Variables						
Self-identified as poor (=1)	0.68	0.47	0.62	0.49	0.07	0.01***
Better now than 5 years ago (=1)	0.40	0.49	0.36	0.48	0.04	0.01***
Internal displacement (=1)	0.14	0.34	0.14	0.34	0.00	0.01***
Land dispossession (=1)	0.02	0.13	0.02	0.13	0.00	0.00***
Land abandonment (=1)	0.02	0.15	0.02	0.13	0.00	0.00*
Obs.	4,157		12,621			
Panel C: Dwelling-level Variables						
Concrete Walls (=1)	0.17	0.37	0.21	0.41	-0.04	0.01***
Concrete Floor (=1)	0.03	0.18	0.05	0.22	-0.02	0.00***
Electricity (=1)	0.35	0.48	0.36	0.48	-0.01	0.01***
Sewerage system (=1)	0.04	0.19	0.04	0.20	-0.00	0.00***
Aqueduct system (=1)	0.14	0.35	0.14	0.35	-0.00	0.01***
Obs.	4,089		12,358			
Panel D: UPA-level Variables						
Total Revenue (Thousands of 2013 COP)	119.93	455.77	68.71	1,305.36	51.22	23.85***
Perennial Crop (=1)	0.97	0.16	0.94	0.24	0.04	0.00***
Transitory Crop (=1)	0.75	0.43	0.84	0.37	-0.08	0.01***
Cash Crops (=1)	0.64	0.48	0.61	0.49	0.04	0.01***
Sells Crops (=1)	0.41	0.49	0.47	0.50	-0.06	0.01
Sells Livestock (=1)	0.74	0.44	0.71	0.46	0.03	0.01
Obs.	6,708		16,452			
Observations						
Municipalities	5		10			
Rural Districts	337		553			

Notes: This tables show simple mean differences between exposed municipalities and neighboring municipalities. Sample is restricted to *El Caguan*'s DMZ and neighboring municipalities. I don't consider neighboring in the highlands, as they are most likely not comparable with those in the plains. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Balance on Geographic and Pre-Demilitarization Characteristics

	Within 50 km			*Within 25 km*			*RD Estimates*	
	Inside	Outside	SE	Inside	Outside	SE	RD Coefficient	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Geographic Characteristics								
Elevation	450.80	376.12	33.40	494.43	411.20	47.89	109.82	134.71
Rainfall (Avg.)	203.28	218.09	1.78***	204.45	213.69	2.43***	-4.57	5.87
Rainfall (St. Dev.)	3.15	3.09	0.28	3.37	3.12	0.34	0.74	0.66
Land Suitability	0.22	0.16	0.02***	0.24	0.19	0.02**	0.03	0.06
Cropland (1992)	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.01
River (=1)	0.51	0.41	0.04**	0.48	0.40	0.05*	0.09	0.10
Panel B: Location Characteristics								
Distance: Bogota	288.69	327.71	4.09***	291.22	314.46	5.15***	-1.78	12.95
Distance: Department's Capital	149.42	147.92	4.16	151.33	148.65	5.40	3.36	13.56
Observations - Grid cells	319	445		202	232		429	

Notes: Columns (1), (2), (4), and (5) present the mean of the corresponding variable. Columns (3) and (6) present clustered standard errors for the difference in means clustered at the municipality level. Inside and Outside indicate whether a grid cell's centroid is inside or outside the former DMZ area, respectively. Columns (7) and (8) show estimates of τ in equation 1 and it's standard error, respectively. The unit of observation is at the $10km$ by $10km$ grid level. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. I present robust standard errors. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 4: Balance on Basic Demopraphic Characteristics

	Within 50 km			*Within 25 km*			*RD Estimates*	
	Inside	Outside	SE	Inside	Outside	SE	RD Coefficient	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Demographic Characteristics								
Age	30.84	32.03	0.25***	31.75	31.94	0.29	0.44	0.77
Men (=1)	0.56	0.55	0.01*	0.56	0.55	0.01	-0.01	0.01
Natives (=1)	0.10	0.02	0.00***	0.08	0.03	0.00***	0.00	0.00
No Ethnicity (=1)	0.87	0.97	0.00***	0.90	0.95	0.00***	-0.00	0.00
Observations - People	11,565	49,471		8,265	21,601		9,152	

Notes: Columns (1), (2), (4), and (5) present the mean of the corresponding variable. Columns (3) and (6) present clustered standard errors for the difference in means clustered at the municipality level. Inside and Outside indicate whether a grid cell's centroid is inside or outside the former DMZ area, respectively. Columns (7) and (8) show estimates of τ in equation 1 and it's standard error, respectively. The unit of observation is an individual in the CNA. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. I present robust standard errors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Human Capital

	Literacy (=1)		Years of Education		Health Insurance (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)
RD Estimate	-0.01	0.00	0.40***	0.35***	0.01	0.01*
	(0.009)	(0.010)	(0.126)	(0.133)	(0.008)	(0.007)
	[0.520]		[0.107]		[0.520]	
BW Selection	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed
BW	15.3	10.0	11.3	10.0	7.58	10.0
Obs.	18250	12070	12338	10754	8750	11723
Dep. Var. Mean	0.84	0.84	7.53	7.48	0.98	0.97
Dep. Var. Std.	0.37	0.37	4.06	4.06	0.16	0.16

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns (1) and (2) is an indicator variable that takes the value of one for individuals that know how to read and write, in columns (3) and (4) it is the total years of formal education, and in (5) and (6) it is an indicator that takes the value of one for individuals with health insurance. The unit of observation is at the individual level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. I present the corresponding RD plot in Figure 6.

Table 6: Dwelling Characteristics and Public Conveniences

	Concrete Walls (=1)		Concrete Floor (=1)		Electricity (=1)		Sewerage (=1)		Aqueduct (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
RD Estimate	-0.02	-0.01	-0.02	-0.01	-0.01	-0.01	0.02**	0.01	0.11**	0.13***
	(0.040)	(0.038)	(0.025)	(0.021)	(0.066)	(0.068)	(0.010)	(0.008)	(0.046)	(0.044)
	[0.733]		[0.584]		[0.740]		[0.109]		[0.090]	
BW Selection	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed
BW	8.84	10.0	7.82	10.0	10.8	10.0	6.78	10.0	7.98	10.0
Obs.	3392	3822	2976	3815	4224	3901	2614	3836	3054	3836
Dep. Var. Mean	0.23	0.24	0.04	0.04	0.43	0.41	0.01	0.01	0.11	0.14
Dep. Var. Std.	0.42	0.43	0.19	0.20	0.49	0.49	0.10	0.11	0.31	0.35

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns (1) and (2) is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in columns (3) and (4) it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in columns (5) and (6) is an indicator that equals one if the dwelling has access to electricity, in columns (7) and (8) it is an indicator that equals one if the dwelling has access to sewerage system, and in columns (9) and (10) it is an indicator that equals one if the dwelling has access to aqueduct system. The unit of observation is at the dwelling level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, and include boundary segment fixed effects. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses and Romano and Wolf (2005)'s step-down adjusted p-values robust to multiple hypothesis testing in square brackets with 1,000 replications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. I present the corresponding RD plot in Figure 7.

Table 7: Agricultural Yield

	Revenue Per Hectare		Total Revenue		Subsistence (=1)		Sale (=1)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD Estimate	0.16***	0.17***	0.35***	0.38***	0.12**	0.08	-0.02	-0.02
	(0.020)	(0.022)	(0.084)	(0.080)	(0.060)	(0.053)	(0.013)	(0.013)
	[0.001]		[0.001]		[0.037]		[0.113]	
BW Selection	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed
BW	12.1	10.0	9.17	10.0	7.47	10.0	9.74	10.0
Obs.	8466	6994	6364	6994	4287	5773	5627	5773
Dep. Var. Mean	4.40	4.28	15.29	15.47	0.42	0.41	0.02	0.02
Dep. Var. Std.	4.20	4.13	37.81	37.56	0.49	0.49	0.15	0.15

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns (1) and (2) is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP, while in columns (3) and (4) it is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP. Dependent variable in columns (5) and (6) is an indicator that equals one if the farms agricultural production is used only for self-consumption, and in columns (7) and (8) it is an indicator that equals one if the farms agricultural production is used only for market sale. The unit of observation is at the farm-crop level in columns (1) to (4), while it is at the farm level in columns (5) to (8). All regressions use a triangular kernel, local linear polynomial at each side of the boundary, and include boundary segment fixed effects, and control for farm extension. In columns (1) to (4) I also include crop fixed effects. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses and Romano and Wolf (2005)'s step-down adjusted p-values robust to multiple hypothesis testing in square brackets with 1,000 replications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. I present the corresponding RD plot in Figure 8 .

Table 8: Agricultural Choices

	Cash Crops - Share				Perennial Crops - Share				Transitory Crops - Share			
	Area		Revenue		Area		Revenue		Area		Revenue	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
RD Estimate	-0.07*	-0.09**	-0.11**	-0.12**	0.07**	0.07	0.13***	0.07*	-0.07**	-0.07	-0.10***	-0.00
	(0.041)	(0.045)	(0.046)	(0.048)	(0.035)	(0.048)	(0.033)	(0.037)	(0.035)	(0.048)	(0.032)	(0.047)
	[0.026]		[0.012]		[0.012]		[0.001]		[0.012]		[0.001]	
BW Selection	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed	Optimal	Fixed
BW	14.3	10.0	11.6	10.0	21.8	10.0	18.2	10.0	21.8	10.0	25.6	10.0
Obs.	3476	2482	2907	2482	5391	2482	4437	2482	5391	2482	7271	2482
Dep. Var. Mean	0.33	0.33	0.31	0.32	0.63	0.58	0.63	0.57	0.37	0.42	0.34	0.39
Dep. Var. Std.	0.34	0.34	0.37	0.37	0.32	0.31	0.35	0.35	0.32	0.31	0.34	0.34

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns (1) and (2) is the share of area with cash crops, while in columns (3) and (4) it is the share of revenue from cash crops. The dependent variable in columns (1) and (2) is the share of area with perennial crops, while in columns (3) and (4) it is the share of revenue from perennial crops.The dependent variable in columns (1) and (2) is the share of area with transitory crops, while in columns (3) and (4) it is the share of revenue from transitory crops. The unit of observation is at the farm level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, and include boundary segment fixed effects, and control for farm extension. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses and Romano and Wolf (2005)'s step-down adjusted p-values robust to multiple hypothesis testing in square brackets with 1,000 replications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. I present the corresponding RD plot in Figure 9 .

Table 9: Mechanism - Violence

	Forced Displacement (=1)	Land Dispossession (=1)	Land Abandonment (=1)
	(1)	(2)	(3)
RD Estimate	-0.10**	-0.04***	-0.03***
	(0.039)	(0.011)	(0.007)
BW Selection	Optimal	Optimal	Optimal
BW	16.6	14.0	15.9
Obs.	6353	5386	6107
Dep. Var. Mean	0.17	0.02	0.02
Dep. Var. Std.	0.37	0.14	0.14

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns (1) is an indicator that equals one if the household head reports that at least one individual within the family was a victim of forced displacement, in columns (2) it is an indicator that equals one if the household head reports that at least one individual within the family was a victim of land dispossession, and in columns (3) it is an indicator that equals one if the household head reports that at least one individual within the family was a victim of land abandonment. The unit of observation is at the household level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, and include boundary segment fixed effects, and control for farm extension. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Mechanism - Agricultural Practices

	Panel A: Property Rights		**Panel B: Agricultural Practices**		
	Private Property (=1)	Collective Property (=1)	Irrigation System (=1)	Fertilizer Use (=1)	Pest Control (=1)
	(1)	(2)	(3)	(4)	(5)
RD Estimate	-0.01	-0.03***	0.00	-0.00	-0.00
	(0.028)	(0.010)	(0.036)	(0.046)	(0.045)
BW Selection	Optimal	Optimal	Optimal	Optimal	Optimal
BW	9.00	6.01	13.0	12.3	13.8
Obs.	4589	3108	7351	6966	7766
Dep. Var. Mean	0.86	0.01	0.82	0.45	0.67
Dep. Var. Std.	0.35	0.11	0.38	0.50	0.47

	Panel C: Capital Investment		**Panel D: Partnership Structure**		
	Machinery (=1)	Buildings (=1)	Producers Association (=1)	*Cooperativa* (=1)	Colective Work (=1)
	(6)	(7)	(8)	(9)	(10)
RD Estimate	0.07*	0.01	-0.06**	0.01	0.05***
	(0.043)	(0.039)	(0.028)	(0.008)	(0.015)
BW Selection	Optimal	Optimal	Optimal	Optimal	Optimal
BW	11.8	13.4	13.4	21.2	16.2
Obs.	6083	6707	5555	8303	8074
Dep. Var. Mean	0.53	0.43	0.11	0.02	0.05
Dep. Var. Std.	0.50	0.50	0.31	0.12	0.22

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns (1) is an indicator that equals that equals one if the farm is privately own, while in columns (2) it is an indicator that equals that equals one if the farm is collectively own. The dependent variable in column (3) is an indicator that equals that equals one if the farm has an irrigation system, in column (4) it is an indicator that equals that equals one if the farm use fertilizers, and in column (3) it is an indicator that equals that equals one if the farm uses modern pest control practices. The dependent variable in columns (6) is an indicator that equals that equals one if the farm has machinery equipment, while in columns (2) it is an indicator that equals that equals one if the farm has buildings for agricultural production. The dependent variable in column (3) is an indicator that equals that equals one if the farmers participate in a producers association, in column (4) it is an indicator that equals that equals one if the farmers participate in a *cooperativa*, and in column (3) it is an indicator that equals that equals one if the farm uses collective work for agricultural production. The unit of observation is at the household level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, and include boundary segment fixed effects, and control for farm extension. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Let the Rebels Rule?
EVIDENCE ON THE ECONOMIC EFFECTS OF REBEL GOVERNANCE IN COLOMBIA

Author: Santiago Pérez-Cardona

Table of Contents

A Data Appendix

This appendix provides detailed information on data sources, sample availability and other details for the different variables used in the paper.

Agricultural Product Prices

Information on agricultural product prices was provided by the Center for Production and Sectorial Trade Studies (*Centro de Estudios sobre Producción y Comercio Sectorial* - CEPCO) at the Central Bank of Colombia. CEPCO gathered data on crops prices from two primary sources: i) the Agricultural Price Information System (*Sistema de Información de Precios de Sector Agropecuario* - SIPSA) from the National Department of Statistics (*Departamento Adminsitrativo Nacional de Estadística* - DANE); ii) official record from producers associations (*agremiaciones*). To compute each price per kilogram, CEPCO takes the average price for each crop across 2013 and wholesale markets (*mercados mayoristas*).

Three aspects of the prices data are worth-noting. First, price information is only available for 86 crop types from 122 in the CNA. Nonetheless, these 86 crops represent 79.7% of the total sown area in the CNA. Second, the prices available include transportation costs and intermediation margins, which might exaggerate agricultural revenues for the local producer. Third, for sugar cane, crude palm oil and cottonseed production, CEPCO applies conversion factors to the final product price (i.e., sugar, oil, cotton) to retrieve a better estimate of the producers income. See de Roux (2020); de Roux et al. (2019) for previous work using this data.

Rural Census (2014)

Most of the information I use comes from Colombia's National Rural Census. The CNA covers the entire rural population in Colombia, and its information corresponds to the year 2013. Moreover, I access restrictive information of each estate's exact GPS location to compute the distance to the DMZ. I now explain in detail each variable relevant to my analysis.

Individuals and Dwelling Characteristics

- *Distance to the DMZ's Border*: Define as the geodesic distance from the farms GPS location to the nearest point of the DMZ's Border.
- *Literacy (=1)*: Indicator that equals one if the individual reports knowing how to read and write.
- *Years of Education*: I construct this variable using each individual's highest grade achieved. I then multiply each grade by the average number of years one has to study to complete it.
- *Health Insurance (=1)*: Indicator that equals one if the individual reports having health insurance.

- *Concrete Walls (=1)*: Indicator that equals one if the dwelling's walls materials are in concrete or better. Better materials include bricks, stones, or polished wood. Worse materials include, for example, cartons, raw lumber, *bahareque*, canned waste, or no walls.

- *Concrete Floor (=1)*: Indicator that equals one if the dwelling's floors materials are in concrete or better. Better materials include tile, bricks, marble, or polished wood. Worse materials include raw lumber, ground, or sand.

- *Electricity (=1)*: Indicator that equals one if the dwelling owner reports having electricity.

- *Sewerage (=1)*: Indicator that equals one if the dwelling owner reports having access to a sewerage system.

- *Aqueduct (=1)*: Indicator that equals one if the dwelling owner reports having access to a aqueduct system.

- *Forced Displacement (=1)*: Indicator that equals one if the household head reports that at least one individual within the family was a victim of forced displacement. Forced displacement refers to an involuntary or coerced migration of a person or people away from their home as a result of violence.

- *Land Dispossession (=1)*: Indicator that equals one if the household head reports that at least one individual within the family was a victim of land dispossession. Forced displacement refers to an involuntary transfer of right over a estate as a result of violence.

- *Land Abandonment (=1)*: Indicator that equals one if the household head reports that at least one individual within the family was a victim of land abandonment. Land abandonment refers to an involuntary or coerced migration of a person or people away from their estate.

Agricultural Production

- *Total Revenue*: Since CNA do not provide information on producers revenues, I have to compute them using information on the total production quantity and average prices per crop. To do so, I compute the total amount produced of each crop by farm and multiply it by price provided by CEPCO. This leaves me with the total revenue per crop at the crop-farm level. I further clean this variable winsorizing at the 1% and 99% levels.

- *Revenue Per Hectare (Yield)*: I then divide the revenue per crop at the crop-farm level by the total number of hectares each producer devotes to each crop. This leaves me with the revenue per hectare at the crop-farm level. I further clean this variable winsorizing at the 1% and 99% levels.

- *Subsistence (=1)*: Indicator that equals one if the farms agricultural production is used only for self-consumption.

- *Sale (=1)*: Indicator that equals one if the farms agricultural production is used only for market sale.

- *Cash Crops*: Indicator that equals one for crops that require centralized processing to be valuable, and it cannot be directly consumed by an individual worker. I categorize the following crops as cash crops: African oil palm, rice, yellow corn, white corn, cocoa, sugar cane, rubber tree, and coffee. I then compute the area share devoted to cash crops and the revenue share they yield to each producer.

- *Perennial Crops*: Indicator that equals one for crops that don't need to be replanted each year, as they automatically grow back after each harvest. I follow DANE's official classification of perennial crops. I then compute the area share devoted to perennial crops and the revenue share they yield to each producer.

- *Transitory Crops*: Indicator that equals one for crops that need to be replanted each year, as they do not grow back after each harvest. I follow DANE's official classification of transitory crops. I then compute the area share devoted to transitory crops and the revenue share they yield to each producer.

- *Private Property (=1)*: Indicator that equals one if the farm is privately own.

- *Collective Property (=1)*: Indicator that equals one if the farm is collectively own.

- *Irrigation System (=1)*: Indicator that equals one if the farm has an irrigation system for agricultural production.

- *Fertilizer Use (=1)*: Indicator that equals one if the producer uses fertilizer in the agricultural production process.

- *Pest Control (=1)*: Indicator that equals one if the producer uses pest control in the agricultural production process.

- *Machinery (=1)*: Indicator that equals one if the farm has machinery for agricultural production.

- *Buildings (=1)*: Indicator that equals one if the farm has buildings and infrastructure for agricultural production.

- *Producers Association (=1)*: Indicator that equals one if the producer belongs to a producers association.

- *Cooperativa (=1)*: Indicator that equals one if the producer belongs to a *cooperativa*. *Cooperativa* typically refers to a association of small producers.

- *Colective Work (=1)*: Indicator that equals one if the farms uses collective work within the production process.

Geographic Information

To study geographic characteristics at each side of the DMZ's boundary, I use grid of $10km$ by $10km$ cells with the average value of each characteristic within the cell. I use this size of cells, because the precision of most information varies at this level.

- *Elevation*: Average elevation comes from WorldClim, a online database of high spatial resolution global weather and climate data. I use elevation data with a resolution of 30 second resolution (about 1 km cells at the equator).

- *Rainfall*: Average rainfall comes from WorldClim, a online database of high spatial resolution global weather and climate data. I use rainfall data with a resolution of 2.5 minutes resolution (about 4.5 km cells at the equator). Specifically, I usec compute the rainfall's mean and standard deviation between 1980 and 1989.

- *Land Suitability*: This information comes from The Atlas of the Biosphere at the University of Wisconsin's Nelson Institute Center for Sustainability and the Global Environment (SAGE). The dataset represents the fraction of each grid cell that is suitable to be used for agriculture. It is based on the temperature and soil conditions of each grid cell, in a resolution of $10km$ by $10km$.

- *Cropland (1992)*: This information comes from The Atlas of the Biosphere at the University of Wisconsin's Nelson Institute Center for Sustainability and the Global Environment (SAGE). The dataset represents the fraction of each gridcell's surface that is covered by crops. It is based on the temperature and soil conditions of each grid cell, in a resolution of $10km$ by $10km$.

- *River (=1)*: This information comes from official information of Colombia's National Mapping Institute - IGAC. I construct a indicator variables that takes the value of one of there is a river within each cell.

Fiscal Outcomes

To study differential investment by the government after the DMZ as an alternative explanation, I use fiscal information from Colombia's National Planning Department (*Departamento Nacional de Planeación* - DNP):

- *Total SGP*: Measured in millions of 2010 COP. This variable measures each municipality-year total transfers from the central government to the municipal government (i.e., *Sistema General de Participaciones*- SGP). Information is available for all municipalities between 1994 and 2008, based on the data in the provided yearly by DNP.

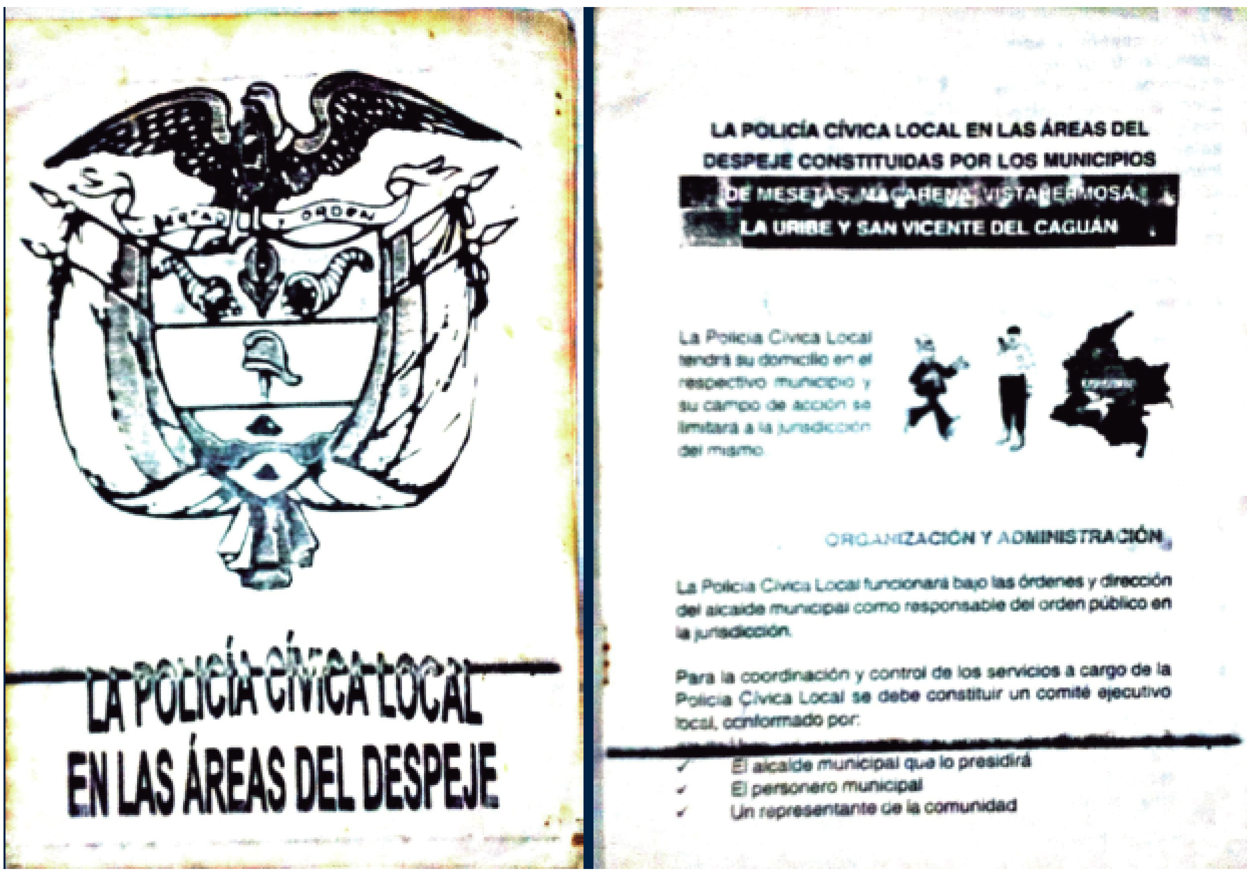
- *Municipal Investment*: Measured in millions of 2010 COP. This variable measures each municipality-year capital expenses (i.e., investment). Information is available for all municipalities between 1994 and 2008, based on the data in the municipal balance sheets provided yearly by DNP.

B Additional Figures and Tables

Figure B1: President-elect Pastrana meets Manuel Marulanda in 1998

Notes: Obtained from El Tiempo (2020). From left to right: Victor Julio Suárez (a high-ranking member of the FARC), Andrés Pastrana (president-elect), Manuel Marulanda (FARC leader), and Víctor Ricardo (high-ranking aide to the president-elect).

Figure B2: *El Caguan's civic* police pamphlet

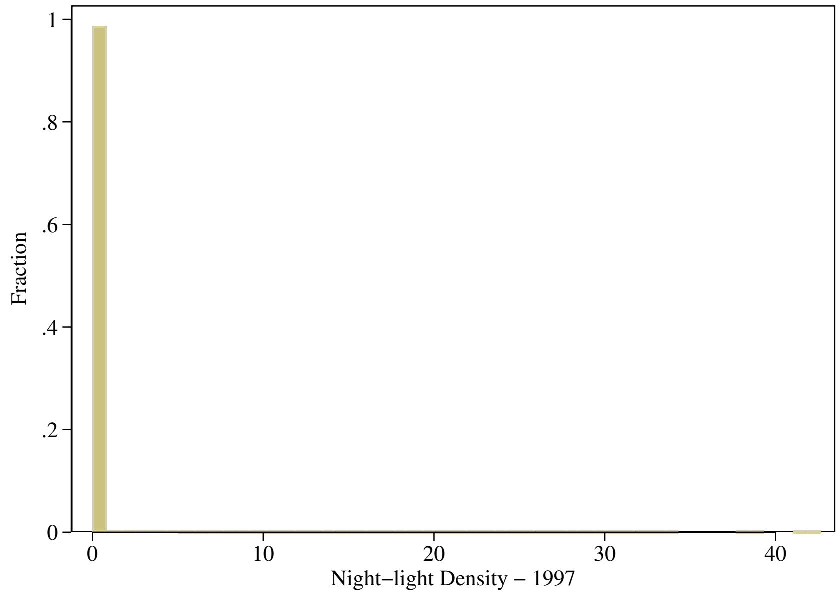


Notes: XXXXX Source Reyes (2012).

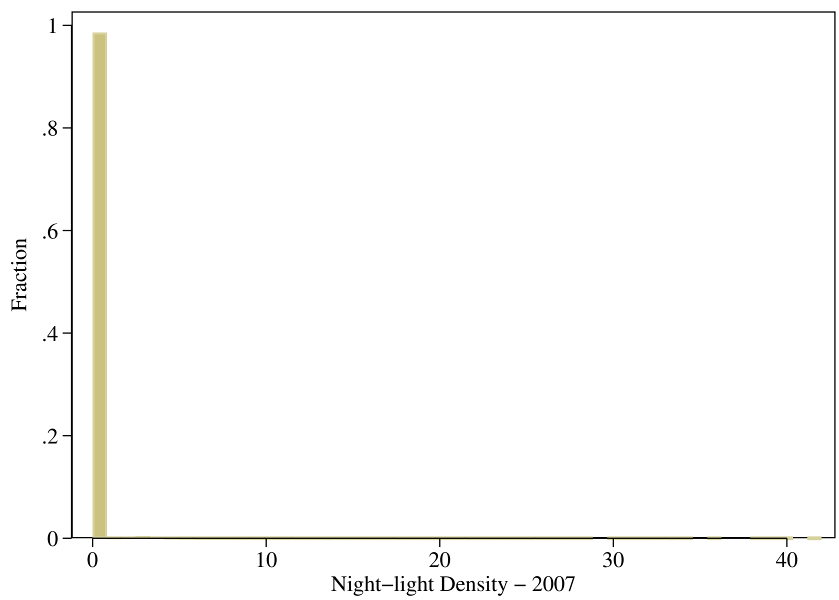
Figure B3: San Vicente's Police station in 2012

Notes: Source Reyes (2012). Sandbags surround San Vicente's police station in 2012 as a defense mechanism from FARC attacks.

Figure B4: Night-light Density



(a) 1997



(b) 2007

Notes: These figures show the nigh-light density distribution of 1km 1 1km cells within 100 km of the DMZ's border.

Table B1: Armed Actors Presence in Vistahermosa - ELCA

	Presence of Armed Actors (= 1)									
	2001	2002	2003	2004	2005	2006	2007	2009	2009	2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Vistahermosa (=1)	0.61**	0.44*	0.19	0.24	0.29	0.29	0.30	0.10	0.18	0.41**
$\Pr(\lvert\beta^{RI:Vistahermos}\rvert > \lvert\beta^{Non-Affected}\rvert)$	[0.014]	[0.066]	[0.562]	[0.594]	[0.142]	[0.156]	[0.142]	[0.990]	[0.218]	[0.044]
DV. Mean Non-Exposed	0.38	0.30	0.30	0.25	0.21	0.21	0.19	0.14	0.065	0.81

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Unit of observations are ELCA's communities (i.e., most likely neighborhoods). I compare theecha variable's mean of communities in Vistahermosa, which was part of the DMZ, with other communities in the oriental region of Colombia. In squared brackets, I present randomization inference p-values with 500 replications.

Table B2: Multiple Armed Actors Presence in Vistahermosa - ELCA

	More than One Armed Actors (= 1)									
	2001	2002	2003	2004	2005	2006	2007	2009	2009	2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Vistahermosa (=1)	-0.16	-0.24	-0.63**	-0.56*	0.03	0.03	0.00	-0.55	-0.50	-0.10
$\Pr(\lvert\beta^{RI:Vistahermos}\rvert > \lvert\beta^{Non-Affected}\rvert)$	[0.636]	[0.562]	[0.032]	[0.088]	[0.576]	[0.546]	[0.990]	[0.364]	[0.206]	[0.294]
DV. Mean Non-Exposed	0.66	0.57	0.63	0.56	0.46	0.46	0.60	0.55	0.050	0.40

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Unit of observations are ELCA's communities (i.e., most likely neighborhoods). I compare theecha variable's mean of communities in Vistahermosa, which was part of the DMZ, with other communities in the oriental region of Colombia. In squared brackets, I present randomization inference p-values with 500 replications.

Table B3: Armed Conflict Events in Vistahermosa - ELCA

	Migration (=1)	Forced Recruitment (=1)	Armed Actors Overpower (=1)	Safe (=1)
	(1)	(2)	(3)	(4)
Vistahermosa (=1)	0.48	0.52	0.66***	-0.08
$\Pr(\lvert\beta^{RI:Vistahermos}\rvert > \lvert\beta^{Non-Affected}\rvert)$	[0.228]	[0.152]	[0.000]	[0.990]
DV. Mean Non-Exposed	0.52	0.22	0.33	0.83

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Unit of observations are ELCA's communities (i.e., most likely neighborhoods). I compare eacha variable's mean of communities in Vistahermosa, which was part of the DMZ, with other communities in the oriental region of Colombia. In squared brackets, I present randomization inference p-values with 500 replications.

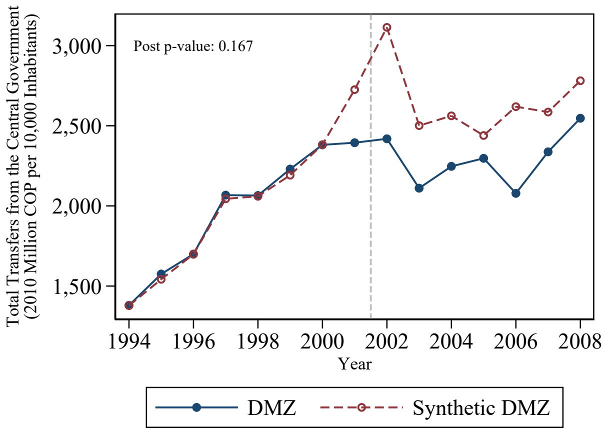
Table B4: Armed Actors Rule in Vistahermosa - ELCA

	Violent Attacks (=1)	Forced Displacement (=1)	Forced Rule of Law (=1)	Taxation (=1)
	(1)	(2)	(3)	(4)
Vistahermosa (=1)	-0.30	0.13	-0.20	-0.34
$\Pr(\lvert\beta^{RI:Vistahermos}\rvert > \lvert\beta^{Non-Affected}\rvert)$	[0.564]	[0.624]	[0.600]	[0.578]
DV. Mean Non-Exposed	0.55	0.37	0.70	0.40

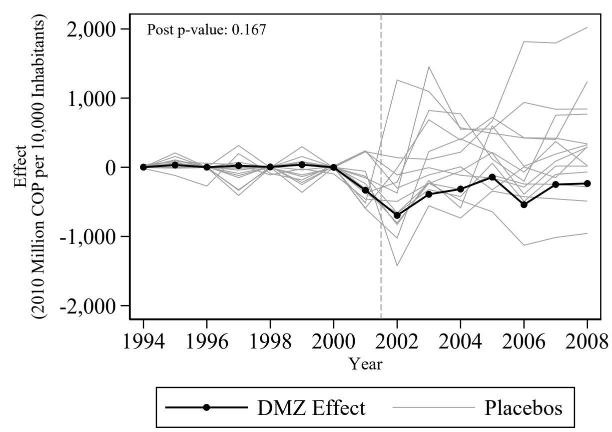
Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. Unit of observations are ELCA's communities (i.e., most likely neighborhoods). I compare theecha variable's mean of communities in Vistahermosa, which was part of the DMZ, with other communities in the oriental region of Colombia. In squared brackets, I present randomization inference p-values with 500 replications.

C Synthetic Control Analysis

Figure C1: Synthetic Control Results: Total SGP



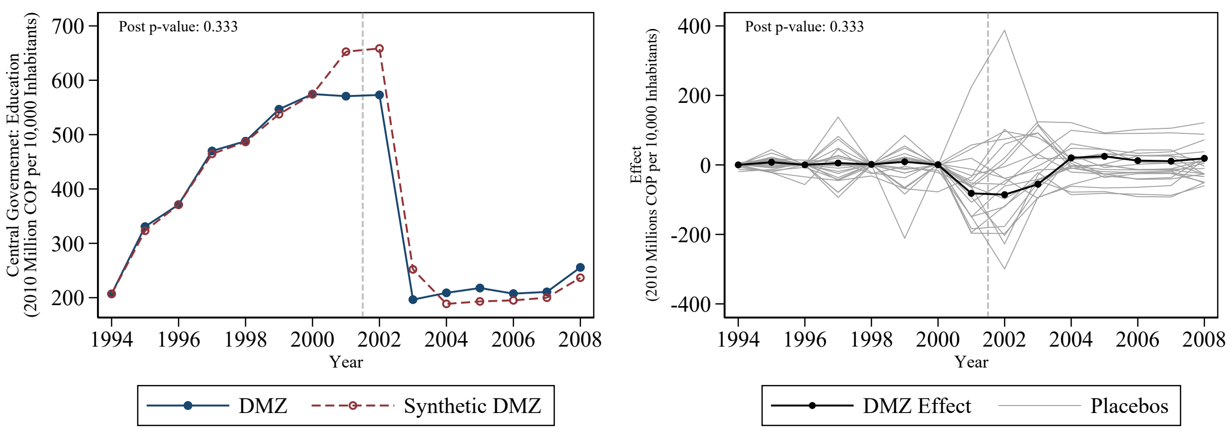
(a) Synthetic Control



(b) Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP transferred from the central government to the municipalities as part of the *Sistema General de Participaciones* (SGP). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

Figure C2: Synthetic Control Results: SGP Education

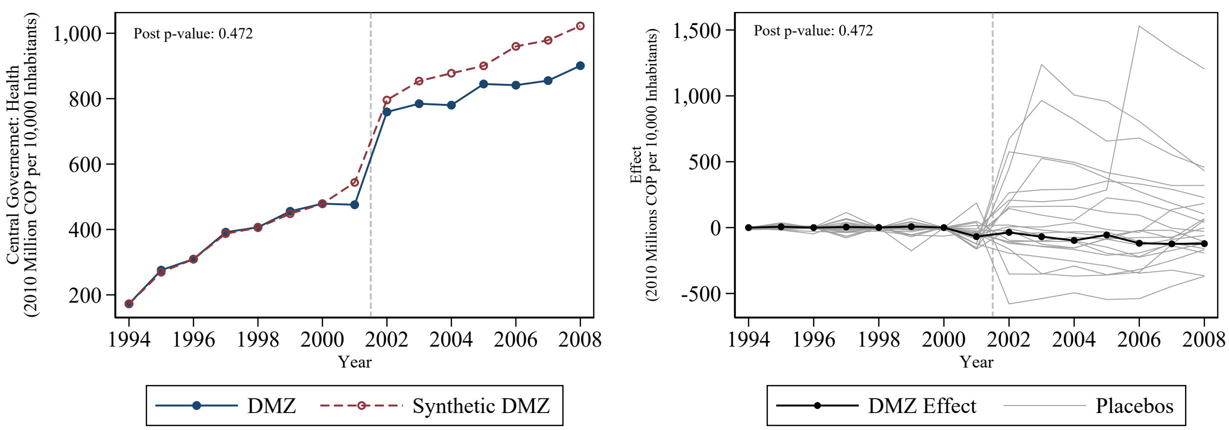


(a) Synthetic Control

(b) Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP transferred from the central government to the municipalities for the educational system as part of the *Sistema General de Participaciones* (SGP). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

Figure C3: Synthetic Control Results: SGP Health

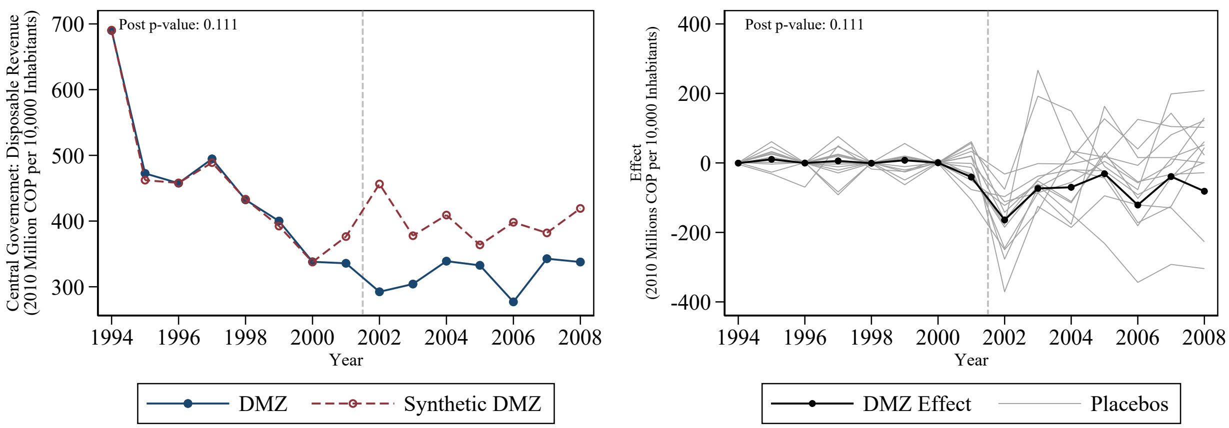


(a) Synthetic Control

(b) Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP transferred from the central government to the municipalities for the health system as part of the *Sistema General de Participaciones* (SGP). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

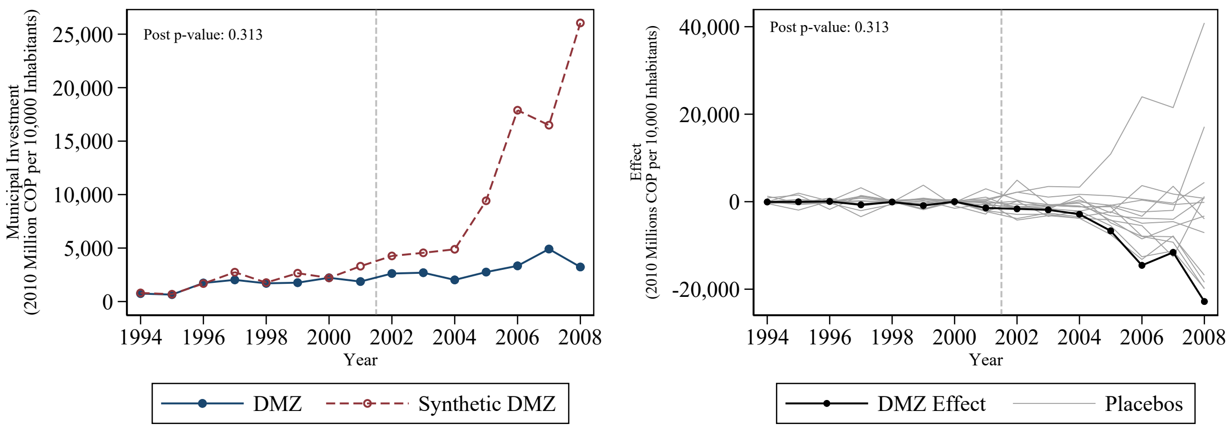
Figure C4: Synthetic Control Results: SGP Disposable Revenues



(a) Synthetic Control **(b)** Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP transferred from the central government to the municipalities as disposable income (i.e., *libre destinación)* as part of the *Sistema General de Participaciones* (SGP). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

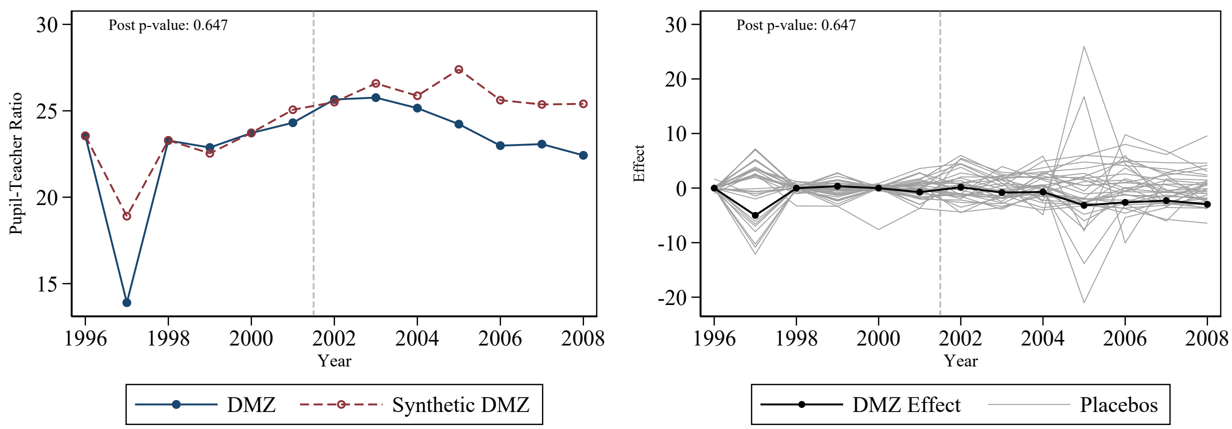
Figure C5: Synthetic Control Results: Municipal Investment



(a) Synthetic Control **(b)** Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP of municipal investment (i.e. capital expenditure). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

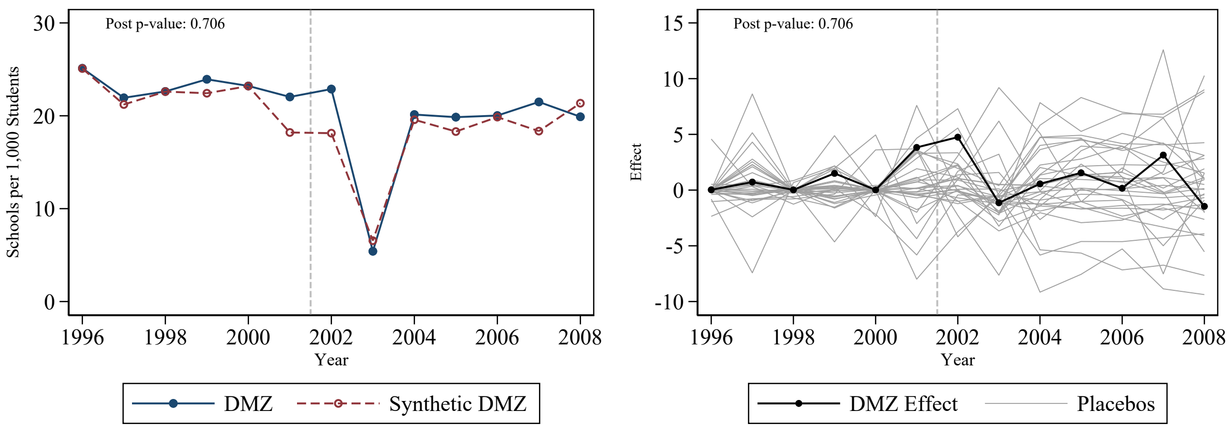
Figure C6: Synthetic Control Results: Pupil-Teacher Ratio



(a) Municipal Investment - Synthetic Control **(b)** Municipal Investment - Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP of municipal investment (i.e. capital expenditure). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

Figure C7: Synthetic Control Results: Schools per 1,000 Students



(a) Municipal Investment - Synthetic Control **(b)** Municipal Investment - Placebos

Notes: Panels (a) shows observed values of fiscal expenditure between 1995 and 2008 (solid line) and counterfactuals from the synthetic control (dashed line). The outcome variable is the total value of 2010 COP of municipal investment (i.e. capital expenditure). To feet the synthetic control I use the dependent variable in 1994, 1996, 1998 and 2000. I collapsed DMZ municipalities into one single entity, restricted sample to municpalities in the departments of Meta, Caqueta and Guaviare, and dropped each department's capital. The "Post p-value" follows Abadie (2021) and refers to the proportions of placebo standardized effects that are at least as large as the main standardized effect for the treated unit.

D Robustness Checks

D.1 RDD Specification

Table D1: RDD Specification Robustness - Human Capital

	Main	Bandwidth				Polynomial		Kernel	
	Estimate	0.50	0.75	1.25	1.50	Quadratic	Cubic	Epanechnikov	Uniform
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Literacy (=1)								
RD Estimate	-0.01	-0.00	-0.00	-0.02**	-0.02**	0.00	-0.00	-0.01	-0.02***
	(0.009)	(0.013)	(0.010)	(0.009)	(0.009)	(0.011)	(0.013)	(0.008)	(0.008)
BW	15.3	7.67	11.5	19.1	23.0	15.5	20.7	15.3	15.3
Obs.	18250	9171	13933	22119	26460	18388	23469	18250	18250
Dep. Var. Mean	0.84	0.84	0.84	0.83	0.83	0.84	0.83	0.84	0.84
Dep. Var. Std.	0.37	0.36	0.37	0.37	0.37	0.37	0.37	0.37	0.37
	Panel B: Years of Education								
RD Estimate	0.40***	-0.45**	0.18	0.34***	0.28**	0.27*	0.20	0.49***	0.46***
	(0.126)	(0.204)	(0.154)	(0.124)	(0.128)	(0.145)	(0.154)	(0.128)	(0.124)
BW	11.3	5.66	8.48	14.1	16.9	15.8	22.4	11.3	11.3
Obs.	12338	6014	9253	15026	17836	16807	23228	12338	12338
Dep. Var. Mean	7.53	7.37	7.46	7.52	7.51	7.52	7.48	7.53	7.53
Dep. Var. Std.	4.06	4.02	4.01	4.06	4.03	4.03	4.04	4.06	4.06
	Panel C: Health Insurance (=1)								
RD Estimate	0.01	0.00	0.01	0.01*	0.01	0.01	0.02	0.02*	0.02**
	(0.008)	(0.011)	(0.009)	(0.007)	(0.006)	(0.007)	(0.010)	(0.008)	(0.007)
BW	7.58	3.79	5.68	9.47	11.3	18.8	17.0	7.58	7.58
Obs.	8750	4499	6604	11199	13478	21166	19420	8750	8750
Dep. Var. Mean	0.98	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98
Dep. Var. Std.	0.16	0.14	0.16	0.16	0.16	0.16	0.16	0.16	0.16

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator variable that takes the value of one for individuals that know how to read and write, in Panel B it is the total years of formal education, and in Panel C it is an indicator that takes the value of one for individuals with health insurance. The unit of observation is at the individual level in all columns. Estimates in column (1) include regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. Columns (2)-(5) vary the optimal bandwidthl columns (6) and (7) vary the local linear polynomial function at each side of the boundary, and columns (8) and (9) vary the kernel weighting. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D2: RDD Specification Robustness - Dwelling Characteristics and Public Conveniences

	Main	Bandwidth				Polynomial		Kernel	
	Estimate	0.50	0.75	1.25	1.50	Quadratic	Cubic	Epanechnikov	Uniform
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Concrete Walls (=1)								
RD Estimate	-0.02	-0.02	-0.04	0.00	0.02	-0.03	-0.05	-0.02	0.00
	(0.040)	(0.056)	(0.050)	(0.050)	(0.047)	(0.044)	(0.046)	(0.041)	(0.042)
BW	8.84	4.42	6.63	11.0	13.2	14.6	21.0	8.84	8.84
Obs.	3392	1751	2542	4268	4993	5535	7579	3392	3392
Dep. Var. Mean	0.23	0.21	0.21	0.25	0.28	0.28	0.31	0.23	0.23
Dep. Var. Std.	0.42	0.41	0.41	0.44	0.45	0.45	0.46	0.42	0.42
	Panel B: Concrete Floor (=1)								
RD Estimate	-0.02	-0.02	-0.03	-0.01	0.00	-0.03	-0.04	-0.02	-0.01
	(0.025)	(0.036)	(0.028)	(0.021)	(0.020)	(0.027)	(0.030)	(0.024)	(0.020)
BW	7.82	3.91	5.87	9.78	11.7	13.6	21.1	7.82	7.82
Obs.	2976	1571	2279	3728	4469	5131	7597	2976	2976
Dep. Var. Mean	0.04	0.04	0.04	0.04	0.05	0.05	0.07	0.04	0.04
Dep. Var. Std.	0.19	0.19	0.19	0.20	0.22	0.22	0.25	0.19	0.19
	Panel C: Electricity (=1)								
RD Estimate	-0.01	-0.02	-0.01	-0.02	-0.02	-0.04	-0.02	-0.02	-0.04
	(0.066)	(0.044)	(0.062)	(0.075)	(0.073)	(0.069)	(0.084)	(0.065)	(0.062)
BW	10.8	5.41	8.12	13.5	16.2	19.6	21.0	10.8	10.8
Obs.	4224	2158	3201	5188	6155	7314	7737	4224	4224
Dep. Var. Mean	0.43	0.35	0.40	0.47	0.49	0.50	0.50	0.43	0.43
Dep. Var. Std.	0.49	0.48	0.49	0.50	0.50	0.50	0.50	0.49	0.49
	Panel D: Sewerage (=1)								
RD Estimate	0.02**	0.01	0.02	0.02*	0.01	0.03**	0.00	0.02**	0.02**
	(0.010)	(0.012)	(0.011)	(0.009)	(0.008)	(0.012)	(0.015)	(0.009)	(0.009)
BW	6.78	3.39	5.08	8.47	10.1	9.40	9.43	6.78	6.78
Obs.	2614	1380	2008	3287	3902	3624	3635	2614	2614
Dep. Var. Mean	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Dep. Var. Std.	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.10	0.10
	Panel E: Aqueduct (=1)								
RD Estimate	0.11**	0.04	0.05*	0.13**	0.11**	0.08*	0.07	0.12***	0.13***
	(0.046)	(0.026)	(0.028)	(0.049)	(0.054)	(0.048)	(0.064)	(0.046)	(0.048)
BW	7.98	3.99	5.98	9.97	11.9	19.8	16.0	7.98	7.98
Obs.	3054	1609	2330	3826	4548	7268	5988	3054	3054
Dep. Var. Mean	0.11	0.10	0.10	0.14	0.16	0.19	0.20	0.11	0.11
Dep. Var. Std.	0.31	0.30	0.30	0.35	0.37	0.39	0.40	0.31	0.31

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in columns Panel B it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in Panel C is an indicator that equals one if the dwelling has access to electricity, in Panel D it is an indicator that equals one if the dwelling has access to sewerage system, and in Panel E it is an indicator that equals one if the dwelling has access to aqueduct system. The unit of observation is at the dwelling level in all columns. Estimates in column (1) include regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. Columns (2)-(5) vary the optimal bandwidth columns (6) and (7) vary the local linear polynomial function at each side of the boundary, and columns (8) and (9) vary the kernel weighting. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D3: RDD Specification Robustness - Agricultural Yield

	Main	Bandwidth				Polynomial		Kernel	
	Estimate	0.50	0.75	1.25	1.50	Quadratic	Cubic	Epanechnikov	Uniform
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Total Per Hectare								
RD Estimate	0.16***	0.17***	0.17***	0.15***	0.16***	0.20***	0.20***	0.16***	0.12***
	(0.020)	(0.035)	(0.032)	(0.034)	(0.032)	(0.030)	(0.034)	(0.020)	(0.019)
BW	12.1	6.06	9.09	15.1	18.1	12.4	15.7	12.1	12.1
Obs.	8466	4157	6307	10500	12295	8665	10831	8466	8466
Dep. Var. Mean	4.40	4.09	4.19	4.52	4.59	4.42	4.57	4.40	4.40
Dep. Var. Std.	4.20	3.79	4.04	4.25	4.36	4.21	4.31	4.20	4.20
	Panel B: Total Revenue								
RD Estimate	0.35***	0.32***	0.41***	0.42***	0.40***	0.49***	0.53***	0.35***	0.42***
	(0.084)	(0.123)	(0.091)	(0.096)	(0.103)	(0.093)	(0.098)	(0.082)	(0.079)
BW	9.17	4.59	6.88	11.4	13.7	16.0	24.9	9.17	9.17
Obs.	6364	3274	4705	8083	9405	11066	17541	6364	6364
Dep. Var. Mean	15.29	15.19	15.73	15.69	15.81	15.97	18.85	15.29	15.29
Dep. Var. Std.	37.81	38.48	39.39	37.01	37.21	37.17	46.02	37.81	37.81
	Panel C: Subsistence (=1)								
RD Estimate	0.12**	0.15*	0.16**	0.09	0.05	0.13**	0.13*	0.12**	0.08
	(0.060)	(0.080)	(0.079)	(0.079)	(0.082)	(0.065)	(0.072)	(0.059)	(0.056)
BW	7.47	3.74	5.60	9.34	11.2	14.4	21.2	7.47	7.47
Obs.	4287	2320	3312	5400	6499	8087	11699	4287	4287
Dep. Var. Mean	0.42	0.42	0.41	0.41	0.40	0.38	0.35	0.42	0.42
Dep. Var. Std.	0.49	0.49	0.49	0.49	0.49	0.49	0.48	0.49	0.49
	Panel D: Sale (=1)								
RD Estimate	-0.02	-0.03	-0.02	-0.02	-0.01	-0.03*	-0.03*	-0.02	-0.02**
	(0.013)	(0.028)	(0.020)	(0.014)	(0.012)	(0.014)	(0.016)	(0.012)	(0.011)
BW	9.74	4.87	7.31	12.1	14.6	17.2	26.8	9.74	9.74
Obs.	5627	2937	4216	6908	8216	9484	16777	5627	5627
Dep. Var. Mean	0.02	0.02	0.02	0.02	0.03	0.03	0.05	0.02	0.02
Dep. Var. Std.	0.15	0.14	0.14	0.16	0.16	0.17	0.22	0.15	0.15

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP, while in Panel B it is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP. Dependent variable in Panel C is an indicator that equals one if the farms agricultural production is used only for self-consumption, and in Panel D it is an indicator that equals one if the farms agricultural production is used only for market sale. The unit of observation is at the farm-crop level in columns Panel A and B, while it is at the farm level in panel C and D. Estimates in column (1) include regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and include boundary segment fixed effects, and control for farm extension. Columns (2)-(5) vary the optimal bandwidth columns (6) and (7) vary the local linear polynomial function at each side of the boundary, and columns (8) and (9) vary the kernel weighting. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D4: RDD Specification Robustness - Agricultural Choices

	Main	Bandwidth				Polynomial		Kernel	
	Estimate	0.50	0.75	1.25	1.50	Quadratic	Cubic	Epanechnikov	Uniform
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Panel A: Cash Crops - Area Share								
RD Estimate	-0.07*	-0.10*	-0.09	-0.06	-0.05	-0.08*	-0.09	-0.06	-0.05
	(0.041)	(0.057)	(0.063)	(0.064)	(0.063)	(0.046)	(0.054)	(0.041)	(0.040)
BW	14.3	7.16	10.7	17.8	21.4	22.8	24.2	14.3	14.3
Obs.	3476	1743	2667	4343	5278	5731	6205	3476	3476
Dep. Var. Mean	0.33	0.32	0.33	0.31	0.31	0.31	0.30	0.33	0.33
Dep. Var. Std.	0.34	0.35	0.34	0.33	0.34	0.34	0.34	0.34	0.34
	Panel B: Cash Crops - Revenue Share								
RD Estimate	-0.11**	-0.13*	-0.13*	-0.09	-0.08	-0.11**	-0.10*	-0.10**	-0.09**
	(0.046)	(0.070)	(0.073)	(0.075)	(0.075)	(0.049)	(0.056)	(0.045)	(0.044)
BW	11.6	5.83	8.74	14.5	17.4	21.2	27.2	11.6	11.6
Obs.	2907	1469	2166	3577	4246	5225	7920	2907	2907
Dep. Var. Mean	0.31	0.31	0.32	0.31	0.29	0.29	0.27	0.31	0.31
Dep. Var. Std.	0.37	0.37	0.38	0.37	0.36	0.36	0.35	0.37	0.37
	Panel C: Perennial Crops - Area Share								
RD Estimate	0.07**	0.07	0.07	0.07	0.08	0.07	0.05	0.08**	0.09***
	(0.035)	(0.063)	(0.069)	(0.060)	(0.055)	(0.042)	(0.057)	(0.034)	(0.031)
BW	21.8	10.9	16.3	27.2	32.7	32.9	29.0	21.8	21.8
Obs.	5391	2707	3989	7935	9653	9710	8603	5391	5391
Dep. Var. Mean	0.63	0.59	0.63	0.63	0.64	0.64	0.64	0.63	0.63
Dep. Var. Std.	0.32	0.30	0.31	0.33	0.33	0.33	0.33	0.32	0.32
	Panel D: Perennial Crops - Revenue Share								
RD Estimate	0.13***	0.06	0.11**	0.14***	0.17***	0.10***	0.08*	0.14***	0.14***
	(0.033)	(0.048)	(0.045)	(0.042)	(0.039)	(0.035)	(0.042)	(0.033)	(0.032)
BW	18.2	9.10	13.6	22.7	27.3	26.5	25.0	18.2	18.2
Obs.	4437	2249	3304	5680	7947	7616	6515	4437	4437
Dep. Var. Mean	0.63	0.57	0.60	0.64	0.64	0.64	0.64	0.63	0.63
Dep. Var. Std.	0.35	0.35	0.35	0.36	0.35	0.35	0.35	0.35	0.35
	Panel E: Transitory Crops - Area Share								
RD Estimate	-0.07**	-0.07	-0.07	-0.07	-0.08	-0.07	-0.05	-0.08**	-0.09***
	(0.035)	(0.063)	(0.069)	(0.060)	(0.055)	(0.042)	(0.057)	(0.034)	(0.031)
BW	21.8	10.9	16.3	27.2	32.7	32.9	29.0	21.8	21.8
Obs.	5391	2707	3989	7935	9653	9710	8603	5391	5391
Dep. Var. Mean	0.37	0.41	0.37	0.37	0.36	0.36	0.36	0.37	0.37
Dep. Var. Std.	0.32	0.30	0.31	0.33	0.33	0.33	0.33	0.32	0.32
	Panel F: Transitory Crops - Revenue Share								
RD Estimate	-0.10***	-0.02	-0.07	-0.11**	-0.10**	-0.05	0.03	-0.12***	-0.13***
	(0.032)	(0.073)	(0.070)	(0.053)	(0.048)	(0.041)	(0.058)	(0.031)	(0.028)
BW	25.6	12.8	19.2	32.0	38.4	32.9	24.1	25.6	25.6
Obs.	7271	3146	4743	9486	10994	9710	6177	7271	7271
Dep. Var. Mean	0.34	0.37	0.34	0.33	0.34	0.33	0.33	0.34	0.34
Dep. Var. Std.	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns Panel A is the share of area with cash crops, while in columns Panel B it is the share of revenue from cash crops. The dependent variable in columns Panel C is the share of area with perennial crops, while in columns Panel D it is the share of revenue from perennial crops.The dependent variable in columns Panel E is the share of area with transitory crops, while in columns Panel F it is the share of revenue from transitory crops. The unit of observation is at the farm level in all columns. Estimates in column (1) use a triangular kernel, local linear polynomial at each side of the boundary, and include boundary segment fixed effects, and control for farm extension. Columns (2)-(5) vary the optimal bandwidth columns (6) and (7) vary the local linear polynomial function at each side of the boundary, and columns (8) and (9) vary the kernel weighting. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Donut Regression Discontinuity Designs

Table D5: Donut RDD - Human Capital

	Main Estimate	Donut Size 0.5 *km*	1.0 *km*	1.5 *km*	2.0 *km*	2.5 *km*
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Literacy (=1)					
RD Estimate	-0.01	-0.01	-0.01	0.01	0.00	-0.01
	(0.009)	(0.009)	(0.011)	(0.013)	(0.013)	(0.014)
BW	15.3	15.3	15.3	15.3	15.3	15.3
Obs.	18250	17924	17186	16454	15740	15176
Dep. Var. Mean	0.84	0.84	0.84	0.84	0.84	0.84
Dep. Var. Std.	0.37	0.37	0.37	0.37	0.37	0.37
	Panel B: Years of Education					
RD Estimate	0.40***	0.71***	1.10***	1.25***	1.69***	1.40***
	(0.126)	(0.149)	(0.186)	(0.251)	(0.301)	(0.279)
BW	11.3	11.3	11.3	11.3	11.3	11.3
Obs.	12338	12050	11432	10824	10180	9674
Dep. Var. Mean	7.53	7.54	7.54	7.55	7.57	7.58
Dep. Var. Std.	4.06	4.06	4.07	4.07	4.08	4.07
	Panel C: Health Insurance (=1)					
RD Estimate	0.01	0.01	0.02	0.02**	0.05***	0.05**
	(0.008)	(0.010)	(0.012)	(0.009)	(0.014)	(0.023)
BW	7.58	7.58	7.58	7.58	7.58	7.58
Obs.	8750	8444	7764	7046	6355	5802
Dep. Var. Mean	0.98	0.98	0.98	0.98	0.97	0.97
Dep. Var. Std.	0.16	0.16	0.15	0.16	0.16	0.17

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator variable that takes the value of one for individuals that know how to read and write, in Panel B it is the total years of formal education, and in Panel C it is an indicator that takes the value of one for individuals with health insurance. The unit of observation is at the individual level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. In columns (2) to (6) I implement a RD Donut design by dropping observations close to the boundary. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table D6: Donut RDD - Dwelling Characteristics and Public Conveniences

	Main	Donut Size				
	Estimate	0.5 *km*	1.0 *km*	1.5 *km*	2.0 *km*	2.5 *km*
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Concrete Walls (=1)					
RD Estimate	-0.02	-0.06	-0.09	-0.02	0.12	0.10
	(0.040)	(0.045)	(0.054)	(0.066)	(0.089)	(0.110)
BW	8.84	8.84	8.84	8.84	8.84	8.84
Obs.	3392	3280	3045	2794	2576	2384
Dep. Var. Mean	0.23	0.23	0.23	0.23	0.24	0.24
Dep. Var. Std.	0.42	0.42	0.42	0.42	0.43	0.43
	Panel B: Concrete Floor (=1)					
RD Estimate	-0.02	-0.04	-0.05**	-0.01	0.03	0.06
	(0.025)	(0.028)	(0.024)	(0.017)	(0.029)	(0.040)
BW	7.82	7.82	7.82	7.82	7.82	7.82
Obs.	2976	2865	2630	2379	2162	1970
Dep. Var. Mean	0.04	0.04	0.04	0.04	0.04	0.04
Dep. Var. Std.	0.19	0.19	0.19	0.19	0.19	0.20
	Panel C: Electricity (=1)					
RD Estimate	-0.01	0.02	-0.03	-0.08	0.05	-0.04
	(0.066)	(0.068)	(0.071)	(0.076)	(0.091)	(0.103)
BW	10.8	10.8	10.8	10.8	10.8	10.8
Obs.	4224	4111	3870	3616	3392	3198
Dep. Var. Mean	0.43	0.43	0.44	0.45	0.45	0.46
Dep. Var. Std.	0.49	0.50	0.50	0.50	0.50	0.50
	Panel D: Sewerage (=1)					
RD Estimate	0.02**	0.02**	0.05***	0.03***	0.04*	0.04*
	(0.010)	(0.012)	(0.012)	(0.012)	(0.021)	(0.020)
BW	6.78	6.78	6.78	6.78	6.78	6.78
Obs.	2614	2502	2264	2010	1788	1595
Dep. Var. Mean	0.01	0.01	0.01	0.01	0.01	0.01
Dep. Var. Std.	0.10	0.10	0.10	0.09	0.09	0.09
	Panel E: Aqueduct (=1)					
RD Estimate	0.11**	0.14***	0.16***	0.15**	0.35***	0.45***
	(0.046)	(0.047)	(0.058)	(0.061)	(0.078)	(0.086)
BW	7.98	7.98	7.98	7.98	7.98	7.98
Obs.	3054	2942	2704	2450	2228	2035
Dep. Var. Mean	0.11	0.11	0.12	0.11	0.12	0.12
Dep. Var. Std.	0.31	0.32	0.32	0.32	0.33	0.32

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in columns Panel B it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in Panel C is an indicator that equals one if the dwelling has access to electricity, in Panel D it is an indicator that equals one if the dwelling has access to sewerage system, and in Panel E it is an indicator that equals one if the dwelling has access to aqueduct system. The unit of observation is at the dwelling level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects. In columns (2) to (6) I implement a RD Donut design by dropping observations close to the boundary. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D7: Donut RDD - Agricultural Yield

	Main	Donut Size				
	Estimate	0.5 *km*	1.0 *km*	1.5 *km*	2.0 *km*	2.5 *km*
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Revenue Per Hectare					
RD Estimate	0.16***	0.16***	0.20***	0.21***	0.19***	0.13***
	(0.020)	(0.021)	(0.024)	(0.026)	(0.032)	(0.039)
BW	12.1	12.1	12.1	12.1	12.1	12.1
Obs.	8466	8288	7805	7376	7013	6690
Dep. Var. Mean	4.40	4.41	4.46	4.48	4.51	4.53
Dep. Var. Std.	4.20	4.19	4.27	4.27	4.31	4.34
	Panel B: Total Revenue					
RD Estimate	0.35***	0.40***	0.64***	0.91***	0.56***	0.16
	(0.084)	(0.093)	(0.091)	(0.111)	(0.156)	(0.178)
BW	9.17	9.17	9.17	9.17	9.17	9.17
Obs.	6364	6186	5703	5274	4911	4588
Dep. Var. Mean	15.29	15.46	15.71	15.79	15.99	15.55
Dep. Var. Std.	37.81	38.04	38.54	38.52	39.01	36.93
	Panel C: Subsistence (=1)					
RD Estimate	0.12**	0.12*	0.16**	0.14*	-0.02	-0.15
	(0.060)	(0.065)	(0.070)	(0.072)	(0.093)	(0.107)
BW	7.47	7.47	7.47	7.47	7.47	7.47
Obs.	4287	4071	3725	3345	3030	2734
Dep. Var. Mean	0.42	0.42	0.42	0.43	0.43	0.43
Dep. Var. Std.	0.49	0.49	0.49	0.49	0.50	0.49
	Panel D: Sale (=1)					
RD Estimate	-0.02	-0.01	-0.01	-0.01	-0.02*	-0.02
	(0.013)	(0.007)	(0.007)	(0.009)	(0.013)	(0.018)
BW	9.74	9.74	9.74	9.74	9.74	9.74
Obs.	5627	5411	5065	4685	4370	4074
Dep. Var. Mean	0.02	0.02	0.02	0.02	0.02	0.02
Dep. Var. Std.	0.15	0.14	0.14	0.14	0.14	0.14

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP, while in Panel B it is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP. Dependent variable in Panel C is an indicator that equals one if the farms agricultural production is used only for self-consumption, and in Panel D it is an indicator that equals one if the farms agricultural production is used only for market sale. The unit of observation is at the farm-crop level in columns Panel A and B, while it is at the farm level in panel C and D. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. In columns (2) to (6) I implement a RD Donut design by dropping observations close to the boundary. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table D8: Donut RDD - Agricultural Choices

	Main Estimate	Donut Size 0.5 km	1.0 km	1.5 km	2.0 km	2.5 km
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Cash Crops - Area Share					
RD Estimate	-0.07*	-0.05	-0.06	-0.05	-0.06	-0.04
	(0.041)	(0.044)	(0.052)	(0.059)	(0.059)	(0.065)
BW	14.3	14.3	14.3	14.3	14.3	14.3
Obs.	3476	3398	3227	3065	2939	2811
Dep. Var. Mean	0.33	0.33	0.33	0.33	0.33	0.33
Dep. Var. Std.	0.34	0.34	0.34	0.33	0.34	0.33
	Panel B: Cash Crops - Revenue Share					
RD Estimate	-0.11**	-0.08*	-0.12*	-0.12*	-0.14**	-0.11
	(0.046)	(0.049)	(0.060)	(0.068)	(0.069)	(0.079)
BW	11.6	11.6	11.6	11.6	11.6	11.6
Obs.	2907	2829	2658	2496	2370	2242
Dep. Var. Mean	0.31	0.31	0.31	0.31	0.32	0.32
Dep. Var. Std.	0.37	0.37	0.37	0.37	0.37	0.37
	Panel C: Perennial Crops - Area Share					
RD Estimate	0.07**	0.09**	0.12***	0.11***	0.09**	0.07*
	(0.035)	(0.036)	(0.036)	(0.036)	(0.038)	(0.041)
BW	21.8	21.8	21.8	21.8	21.8	21.8
Obs.	5391	5313	5142	4980	4854	4726
Dep. Var. Mean	0.63	0.63	0.63	0.63	0.64	0.64
Dep. Var. Std.	0.32	0.32	0.32	0.32	0.32	0.32
	Panel D: Perennial Crops - Revenue Share					
RD Estimate	0.13***	0.15***	0.21***	0.18***	0.15***	0.15***
	(0.033)	(0.035)	(0.038)	(0.040)	(0.047)	(0.055)
BW	18.2	18.2	18.2	18.2	18.2	18.2
Obs.	4437	4359	4188	4026	3900	3772
Dep. Var. Mean	0.63	0.63	0.63	0.63	0.63	0.63
Dep. Var. Std.	0.35	0.35	0.35	0.35	0.35	0.35
	Panel E: Transitory Crops - Area Share					
RD Estimate	-0.07**	-0.09**	-0.12***	-0.11***	-0.09**	-0.07*
	(0.035)	(0.036)	(0.036)	(0.036)	(0.038)	(0.041)
BW	21.8	21.8	21.8	21.8	21.8	21.8
Obs.	5391	5313	5142	4980	4854	4726
Dep. Var. Mean	0.37	0.37	0.37	0.37	0.36	0.36
Dep. Var. Std.	0.32	0.32	0.32	0.32	0.32	0.32
	Panel F: Transitory Crops - Revenue Share					
RD Estimate	-0.10***	-0.11***	-0.15***	-0.15***	-0.14***	-0.15***
	(0.032)	(0.033)	(0.032)	(0.032)	(0.033)	(0.036)
BW	25.6	25.6	25.6	25.6	25.6	25.6
Obs.	7271	7193	7022	6860	6734	6606
Dep. Var. Mean	0.34	0.34	0.34	0.34	0.34	0.34
Dep. Var. Std.	0.34	0.34	0.34	0.34	0.34	0.34

Notes: This table shows estimates of τ in equation 1. The dependent variable in columns Panel A is the share of area with cash crops, while in columns Panel B it is the share of revenue from cash crops. The dependent variable in columns Panel C is the share of area with perennial crops, while in columns Panel D it is the share of revenue from perennial crops.The dependent variable in columns Panel E is the share of area with transitory crops, while in columns Panel F it is the share of revenue from transitory crops. The unit of observation is at the farm level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. In columns (2) to (6) I implement a RD Donut design by dropping observations close to the boundary. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses. * p < 0.1, ** p < 0.05, *** p < 0.01 .

D.3 Robustness to Variance-covariance Matrix Assumptions

Table D9: Variance-covariance Matrix Assumptions - Human Capital

	Standard Errors Structure					
	NN-Cluster		Cluster			
	Vereda	Municipality	*Vereda*	Municipality	250m Bins	500m Bins
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Literacy (=1)					
RD Estimate	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.009)	(0.009)	(0.018)	(0.028)	(0.015)	(0.015)
BW	15.3	15.3	16.9	16.8	16.3	15.8
Obs.	18250	18250	19818	19713	19244	18771
Dep. Var. Mean	0.84	0.84	0.84	0.84	0.84	0.84
Dep. Var. Std.	0.37	0.37	0.37	0.37	0.37	0.37
	Panel B: Years of Education					
RD Estimate	0.40***	0.40***	0.36	0.35	0.40*	0.39*
	(0.126)	(0.126)	(0.363)	(0.358)	(0.220)	(0.233)
BW	11.3	11.9	14.5	13.7	12.7	12.9
Obs.	12338	12847	15640	14712	13660	13936
Dep. Var. Mean	7.53	7.53	7.50	7.54	7.54	7.55
Dep. Var. Std.	4.06	4.06	4.04	4.06	4.06	4.06
	Panel C: Health Insurance (=1)					
RD Estimate	0.01	0.01*	0.01	0.01	0.01	0.01
	(0.008)	(0.007)	(0.018)	(0.013)	(0.013)	(0.012)
BW	7.58	9.83	9.95	9.37	10.7	10.2
Obs.	8750	11500	11687	11045	12659	11971
Dep. Var. Mean	0.98	0.97	0.97	0.97	0.97	0.97
Dep. Var. Std.	0.16	0.16	0.16	0.16	0.16	0.16

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator variable that takes the value of one for individuals that know how to read and write, in Panel B it is the total years of formal education, and in Panel C it is an indicator that takes the value of one for individuals with health insurance. The unit of observation is at the individual level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. While column (1) presents baseline specification for the standard errors, columns (2)-(6) show robustness of the results to alternative specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D10: Variance-covariance Matrix Assumptions - Dwelling Characteristics and Public Conveniences

	Standard Errors Structure					
	NN-Cluster		Cluster			
	Vereda	Municipality	*Vereda*	Municipality	250m Bins	500m Bins
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Concrete Walls (=1)					
RD Estimate	-0.02	-0.02	-0.01	-0.02	-0.02	-0.02
	(0.040)	(0.051)	(0.052)	(0.054)	(0.045)	(0.047)
BW	8.84	9.99	9.78	10.1	8.76	9.01
Obs.	3392	3816	3738	3871	3377	3472
Dep. Var. Mean	0.23	0.24	0.24	0.24	0.23	0.23
Dep. Var. Std.	0.42	0.43	0.43	0.43	0.42	0.42
	Panel B: Concrete Floor (=1)					
RD Estimate	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.025)	(0.024)	(0.026)	(0.024)	(0.022)	(0.022)
BW	7.82	7.30	8.24	7.92	8.28	8.21
Obs.	2976	2776	3149	3018	3165	3148
Dep. Var. Mean	0.04	0.04	0.04	0.04	0.04	0.04
Dep. Var. Std.	0.19	0.19	0.20	0.19	0.20	0.20
	Panel C: Electricity (=1)					
RD Estimate	-0.01	-0.02	-0.02	-0.02	-0.01	-0.01
	(0.066)	(0.075)	(0.108)	(0.123)	(0.057)	(0.062)
BW	10.8	12.2	11.6	12.0	9.71	10.0
Obs.	4224	4678	4551	4642	3797	3912
Dep. Var. Mean	0.43	0.45	0.44	0.44	0.41	0.41
Dep. Var. Std.	0.49	0.50	0.50	0.50	0.49	0.49
	Panel D: Sewerage (=1)					
RD Estimate	0.02**	0.02*	0.02	0.02	0.02	0.02
	(0.010)	(0.011)	(0.016)	(0.014)	(0.017)	(0.017)
BW	6.78	4.96	5.32	7.49	5.25	5.19
Obs.	2614	1973	2093	2839	2069	2038
Dep. Var. Mean	0.01	0.01	0.01	0.01	0.01	0.01
Dep. Var. Std.	0.10	0.10	0.10	0.10	0.10	0.10
	Panel E: Aqueduct (=1)					
RD Estimate	0.11**	0.13***	0.13*	0.13	0.12**	0.12**
	(0.046)	(0.049)	(0.077)	(0.079)	(0.052)	(0.056)
BW	7.98	9.90	10.0	10.1	9.11	9.50
Obs.	3054	3791	3846	3872	3518	3666
Dep. Var. Mean	0.11	0.14	0.14	0.14	0.13	0.14
Dep. Var. Std.	0.31	0.35	0.35	0.35	0.34	0.34

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in columns Panel B it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in Panel C is an indicator that equals one if the dwelling has access to electricity, in Panel D it is an indicator that equals one if the dwelling has access to sewerage system, and in Panel E it is an indicator that equals one if the dwelling has access to aqueduct system. The unit of observation is at the dwelling level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. While column (1) presents baseline specification for the standard errors, columns (2)-(6) show robustness of the results to alternative specifications. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table D11: Variance-covariance Matrix Assumptions - Agricultural Yield

	Standard Errors Structure					
	NN-Cluster		Cluster			
	Vereda	Municipality	*Vereda*	Municipality	250m Bins	500m Bins
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Revenue Per Hectare					
RD Estimate	0.16***	0.17***	0.15***	0.17**	0.15***	0.15***
	(0.020)	(0.034)	(0.037)	(0.076)	(0.038)	(0.043)
BW	12.1	12.0	12.8	10.9	12.8	12.5
Obs.	8466	8434	8886	7704	8877	8706
Dep. Var. Mean	4.40	4.40	4.45	4.33	4.44	4.43
Dep. Var. Std.	4.20	4.20	4.23	4.13	4.23	4.22
	Panel B: Total Revenue					
RD Estimate	0.35***	0.36***	0.39**	0.39	0.42***	0.42***
	(0.084)	(0.087)	(0.180)	(0.243)	(0.150)	(0.157)
BW	9.17	8.23	12.2	11.0	12.2	11.5
Obs.	6364	5728	8557	7847	8506	8139
Dep. Var. Mean	15.29	15.13	15.70	15.72	15.68	15.65
Dep. Var. Std.	37.81	37.98	37.18	37.11	36.99	36.90
	Panel C: Subsistence (=1)					
RD Estimate	0.12**	0.15**	0.09	0.12	0.12**	0.12**
	(0.060)	(0.077)	(0.084)	(0.109)	(0.052)	(0.056)
BW	7.47	6.80	9.82	9.02	7.37	7.42
Obs.	4287	3946	5654	5219	4227	4239
Dep. Var. Mean	0.42	0.42	0.41	0.41	0.42	0.42
Dep. Var. Std.	0.49	0.49	0.49	0.49	0.49	0.49
	Panel D: Sale (=1)					
RD Estimate	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.013)	(0.016)	(0.016)	(0.021)	(0.013)	(0.016)
BW	9.74	9.87	10.1	11.0	9.74	10.6
Obs.	5627	5669	5826	6333	5619	6140
Dep. Var. Mean	0.02	0.02	0.02	0.02	0.02	0.02
Dep. Var. Std.	0.15	0.15	0.15	0.15	0.15	0.15

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP, while in Panel B it is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP. Dependent variable in Panel C is an indicator that equals one if the farms agricultural production is used only for self-consumption, and in Panel D it is an indicator that equals one if the farms agricultural production is used only for market sale. The unit of observation is at the farm-crop level in columns Panel A and B, while it is at the farm level in panel C and D. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and form's size. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. While column (1) presents baseline specification for the standard errors, columns (2)-(6) show robustness of the results to alternative specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D12: Variance-covariance Matrix Assumptions - Agricultural Choices

	Standard Errors Structure					
	NN-Cluster		Cluster			
	Vereda	Municipality	*Vereda*	Municipality	250m Bins	500m Bins
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Cash Crops - Area Share					
RD Estimate	-0.07*	-0.07	-0.07	-0.07	-0.08**	-0.08**
	(0.041)	(0.065)	(0.044)	(0.079)	(0.031)	(0.034)
BW	14.3	17.2	14.7	17.0	12.0	12.6
Obs.	3476	4183	3634	4140	2988	3095
Dep. Var. Mean	0.33	0.32	0.33	0.32	0.33	0.33
Dep. Var. Std.	0.34	0.33	0.34	0.33	0.34	0.34
	Panel B: Cash Crops - Revenue Share					
RD Estimate	-0.11**	-0.10	-0.10**	-0.10	-0.12***	-0.12***
	(0.046)	(0.077)	(0.052)	(0.098)	(0.038)	(0.042)
BW	11.6	14.3	12.3	14.8	10.0	9.80
Obs.	2907	3481	3036	3660	2490	2437
Dep. Var. Mean	0.31	0.31	0.32	0.31	0.32	0.32
Dep. Var. Std.	0.37	0.37	0.37	0.36	0.37	0.37
	Panel C: Perennial Crops - Area Share					
RD Estimate	0.07**	0.07	0.08*	0.08	0.07***	0.07**
	(0.035)	(0.070)	(0.047)	(0.102)	(0.028)	(0.031)
BW	21.8	24.5	23.9	24.7	18.7	18.7
Obs.	5391	6314	6104	6418	4638	4640
Dep. Var. Mean	0.63	0.64	0.64	0.64	0.63	0.63
Dep. Var. Std.	0.32	0.33	0.32	0.33	0.32	0.32
	Panel D: Perennial Crops - Revenue Share					
RD Estimate	0.13***	0.13***	0.13***	0.13*	0.12***	0.12***
	(0.033)	(0.043)	(0.042)	(0.067)	(0.031)	(0.037)
BW	18.2	19.6	17.0	17.0	15.9	15.9
Obs.	4437	4817	4147	4147	3889	3889
Dep. Var. Mean	0.63	0.63	0.62	0.62	0.62	0.62
Dep. Var. Std.	0.35	0.35	0.35	0.35	0.35	0.35
	Panel E: Transitory Crops - Area Share					
RD Estimate	-0.07**	-0.07	-0.08*	-0.08	-0.07***	-0.07**
	(0.035)	(0.070)	(0.047)	(0.102)	(0.028)	(0.031)
BW	21.8	24.5	23.9	24.7	18.7	18.7
Obs.	5391	6314	6104	6418	4638	4640
Dep. Var. Mean	0.37	0.36	0.36	0.36	0.37	0.37
Dep. Var. Std.	0.32	0.33	0.32	0.33	0.32	0.32
	Panel F: Transitory Crops - Revenue Share					
RD Estimate	-0.10***	-0.11*	-0.10**	-0.11	-0.08***	-0.08**
	(0.032)	(0.063)	(0.045)	(0.104)	(0.028)	(0.034)
BW	25.6	27.5	25.0	26.8	20.0	20.9
Obs.	7271	8078	6522	7707	4902	5128
Dep. Var. Mean	0.34	0.33	0.33	0.34	0.34	0.34
Dep. Var. Std.	0.34	0.34	0.34	0.34	0.34	0.34

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is the share of area with cash crops, while in Panel B it is the share of revenue from cash crops. The dependent variable in Panel C is the share of area with perennial crops, while in Panel D it is the share of revenue from perennial crops.The dependent variable in Panel E is the share of area with transitory crops, while in Panel F it is the share of revenue from transitory crops. The unit of observation is at the farm level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and farm's size. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. While column (1) presents baseline specification for the standard errors, columns (2)-(6) show robustness of the results to alternative specifications. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.4 Selective Migration: Trimming

Table D13: Trimming for Selective Migration - Human Capital

	% Trimmed					
	0%	2.5%	5%	10%	15%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Literacy (=1)					
RD Estimate	-0.01	-0.01	-0.01	-0.02*	-0.02*	-0.02*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)
BW	15.3	15.3	15.3	15.3	15.3	15.3
Obs.	18250	18064	17789	17470	17155	16767
Dep. Var. Mean	0.84	0.84	0.83	0.83	0.83	0.83
Dep. Var. Std.	0.37	0.37	0.37	0.37	0.37	0.38
	Years of Education					
RD Estimate	0.40***	0.35***	0.31**	0.22	0.23*	0.30**
	(0.126)	(0.126)	(0.130)	(0.137)	(0.136)	(0.145)
BW	11.3	11.3	11.3	11.3	11.3	11.3
Obs.	12338	12277	12109	11890	11630	11378
Dep. Var. Mean	7.53	7.51	7.49	7.44	7.47	7.49
Dep. Var. Std.	4.06	4.05	4.06	4.03	4.04	4.02
	Panel C: Health Insurance (=1)					
RD Estimate	0.01	0.01	0.01	0.01	0.01	0.02**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
BW	7.58	7.58	7.58	7.58	7.58	7.58
Obs.	8750	8703	8602	8452	8278	8138
Dep. Var. Mean	0.98	0.97	0.98	0.98	0.97	0.97
Dep. Var. Std.	0.16	0.16	0.15	0.16	0.16	0.16

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator variable that takes the value of one for individuals that know how to read and write, in Panel B it is the total years of formal education, and in Panel C it is an indicator that takes the value of one for individuals with health insurance. The unit of observation is at the individual level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and control for age, age squared and sex. In columns (2) to (6), I trim the top 2.5, 5, 10, 15, and 20% of the most well-off individuals and farms. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D14: Trimming for Selective Migration - Dwelling Characteristics and Public Conveniences

	% Trimmed					
	0%	2.5%	5%	10%	15%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Concrete Walls (=1)					
RD Estimate	-0.02	-0.02	-0.05	-0.08**	-0.11***	-0.11***
	(0.040)	(0.040)	(0.039)	(0.038)	(0.039)	(0.039)
BW	8.84	8.84	8.84	8.84	8.84	8.84
Obs.	3392	3373	3328	3270	3203	3148
Dep. Var. Mean	0.23	0.22	0.21	0.20	0.19	0.19
Dep. Var. Std.	0.42	0.42	0.41	0.40	0.39	0.39
	Panel B: Concrete Floor (=1)					
RD Estimate	-0.02	-0.02	-0.03	-0.04*	-0.04**	-0.05**
	(0.025)	(0.024)	(0.024)	(0.022)	(0.023)	(0.022)
BW	7.82	7.82	7.82	7.82	7.82	7.82
Obs.	2976	2958	2918	2864	2798	2747
Dep. Var. Mean	0.04	0.03	0.03	0.03	0.03	0.02
Dep. Var. Std.	0.19	0.18	0.17	0.16	0.16	0.15
	Panel C: Electricity (=1)					
RD Estimate	-0.01	-0.02	-0.04	-0.06	-0.09	-0.11*
	(0.066)	(0.066)	(0.066)	(0.065)	(0.065)	(0.066)
BW	10.8	10.8	10.8	10.8	10.8	10.8
Obs.	4224	4202	4151	4083	4005	3928
Dep. Var. Mean	0.43	0.42	0.42	0.41	0.40	0.39
Dep. Var. Std.	0.49	0.49	0.49	0.49	0.49	0.49
	Panel D: Sewerage (=1)					
RD Estimate	0.02**	0.01	0.01	0.00	0.00	0.00
	(0.010)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
BW	6.78	6.78	6.78	6.78	6.78	6.78
Obs.	2614	2596	2558	2506	2447	2401
Dep. Var. Mean	0.01	0.01	0.01	0.01	0.01	0.01
Dep. Var. Std.	0.10	0.09	0.08	0.08	0.08	0.08
	Panel E: Aqueduct (=1)					
RD Estimate	0.11**	0.10**	0.08*	0.05	0.04	0.00
	(0.046)	(0.045)	(0.043)	(0.042)	(0.040)	(0.038)
BW	7.98	7.98	7.98	7.98	7.98	7.98
Obs.	3054	3035	2994	2940	2873	2821
Dep. Var. Mean	0.11	0.11	0.10	0.09	0.09	0.08
Dep. Var. Std.	0.31	0.31	0.30	0.29	0.29	0.27

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in columns Panel B it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in Panel C is an indicator that equals one if the dwelling has access to electricity, in Panel D it is an indicator that equals one if the dwelling has access to sewerage system, and in Panel E it is an indicator that equals one if the dwelling has access to aqueduct system. The unit of observation is at the dwelling level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects. In columns (2) to (6), I trim the top 2.5, 5, 10, 15, and 20% of the most well-off individuals and farms. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D15: Trimming for Selective Migration - Agricultural Yield

	% Trimmed					
	0%	2.5%	5%	10%	15%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Revenue Per Hectare					
RD Estimate	0.16***	0.15***	0.15***	0.15***	0.14***	0.13***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)
BW	12.1	12.1	12.1	12.1	12.1	12.1
Obs.	8466	8409	8368	8284	8189	8043
Dep. Var. Mean	4.40	4.28	4.23	4.17	4.11	4.04
Dep. Var. Std.	4.20	3.90	3.85	3.83	3.81	3.80
	Panel B: Total Revenue					
RD Estimate	0.35***	0.33***	0.29***	0.25***	0.22***	0.18**
	(0.084)	(0.084)	(0.085)	(0.086)	(0.084)	(0.085)
BW	9.17	9.17	9.17	9.17	9.17	9.17
Obs.	6364	6342	6287	6209	6092	5966
Dep. Var. Mean	15.29	14.00	12.75	12.05	11.44	10.94
Dep. Var. Std.	37.81	30.70	27.27	26.68	26.56	26.62

Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP, while in Panel B it is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP. The unit of observation is at the farm-crop level in columns Panel A and B, while it is at the farm level in panel C and D. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and form's size. In columns (2) to (6), I trim the top 2.5, 5, 10, 15, and 20% of the most well-off individuals and farms. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. The unit of observation is at the individual level. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

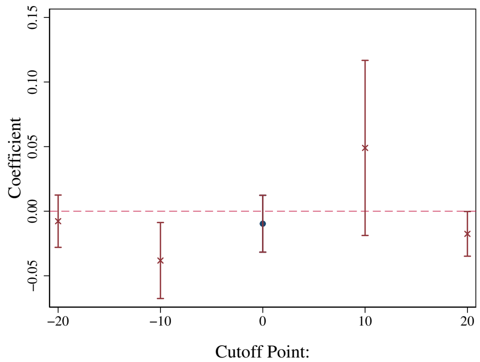
Table D16: Trimming for Selective Migration - Agricultural Choices

	% Trimmed					
	0%	2.5%	5%	10%	15%	20%
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Cash Crops - Area Share					
RD Estimate	-0.07*	-0.07*	-0.07*	0.07*	0.10***	0.15***
	(0.041)	(0.041)	(0.041)	(0.037)	(0.035)	(0.034)
BW	14.3	14.3	14.3	14.3	14.3	14.3
Obs.	3476	3476	3476	3184	3076	2972
Dep. Var. Mean	0.33	0.33	0.33	0.27	0.25	0.24
Dep. Var. Std.	0.34	0.34	0.34	0.28	0.27	0.26
	Panel B: Cash Crops - Revenue Share					
RD Estimate	-0.11**	-0.11**	-0.11**	0.13***	0.17***	0.19***
	(0.046)	(0.046)	(0.046)	(0.041)	(0.034)	(0.032)
BW	11.6	11.6	11.6	11.6	11.6	11.6
Obs.	2907	2907	2907	2537	2487	2392
Dep. Var. Mean	0.31	0.31	0.31	0.21	0.20	0.18
Dep. Var. Std.	0.37	0.37	0.37	0.28	0.27	0.25
	Panel C: Perennial Crops - Area Share					
RD Estimate	0.07**	0.07**	0.07**	0.07**	0.07**	0.07**
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
BW	21.8	21.8	21.8	21.8	21.8	21.8
Obs.	5391	5391	5391	5391	5391	5391
Dep. Var. Mean	0.63	0.63	0.63	0.63	0.63	0.63
Dep. Var. Std.	0.32	0.32	0.32	0.32	0.32	0.32
	Panel D: Perennial Crops - Revenue Share					
RD Estimate	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
BW	18.2	18.2	18.2	18.2	18.2	18.2
Obs.	4437	4437	4437	4437	4437	4437
Dep. Var. Mean	0.63	0.63	0.63	0.63	0.63	0.63
Dep. Var. Std.	0.35	0.35	0.35	0.35	0.35	0.35
	Panel E: Transitory Crops - Area Share					
RD Estimate	-0.07**	-0.07**	-0.07**	-0.02	0.01	0.05*
	(0.035)	(0.035)	(0.035)	(0.035)	(0.033)	(0.030)
BW	21.8	21.8	21.8	21.8	21.8	21.8
Obs.	5391	5391	5391	4950	4817	4595
Dep. Var. Mean	0.37	0.37	0.37	0.31	0.30	0.28
Dep. Var. Std.	0.32	0.32	0.32	0.28	0.27	0.26
	Panel F: Transitory Crops - Revenue Share					
RD Estimate	-0.10***	-0.10***	-0.10***	-0.02	0.01	0.05*
	(0.032)	(0.032)	(0.032)	(0.032)	(0.030)	(0.030)
BW	25.6	25.6	25.6	25.6	25.6	25.6
Obs.	7271	7271	7271	6552	6367	6151
Dep. Var. Mean	0.34	0.34	0.34	0.27	0.25	0.24
Dep. Var. Std.	0.34	0.34	0.34	0.28	0.27	0.26

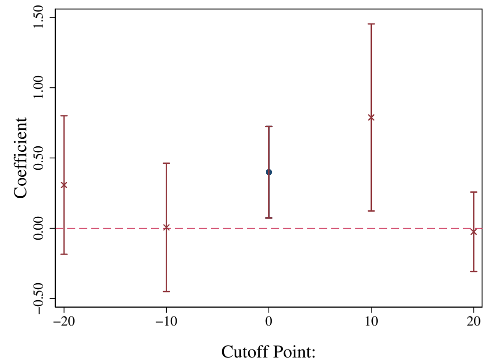
Notes: This table shows estimates of τ in equation 1. The dependent variable in Panel A is the share of area with cash crops, while in Panel B it is the share of revenue from cash crops. The dependent variable in Panel C is the share of area with perennial crops, while in Panel D it is the share of revenue from perennial crops.The dependent variable in Panel E is the share of area with transitory crops, while in Panel F it is the share of revenue from transitory crops. The unit of observation is at the farm level in all columns. All regressions use a triangular kernel, local linear polynomial at each side of the boundary, include boundary segment fixed effects, and farm's size. In columns (2) to (6), I trim the top 2.5, 5, 10, 15, and 20% of the most well-off individuals and farms. The RDD MSE optimal bandwidths are determined using the procedure suggested by Cattaneo et al. (2020a). Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. Standard errors clustered at the rural districts level in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.5 Spatial Placebo Analysis

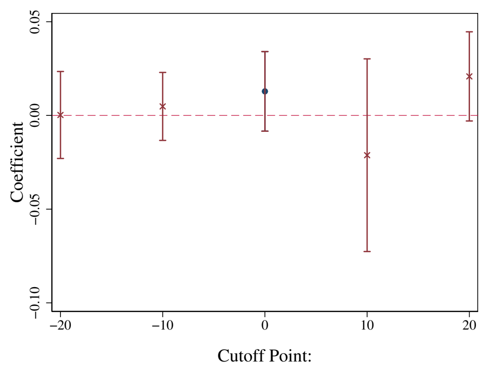
Figure D1: Spatial Placebo: Human Capital



(a) Literacy (=1)



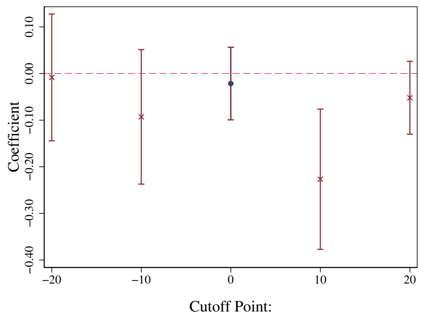
(b) Years of Education



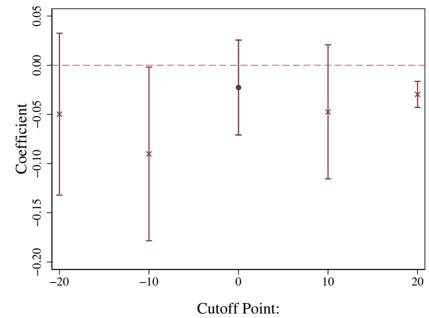
(c) Health Insurance (1)

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is an indicator variable that takes the value of one for individuals that know how to read and write, in panel (b) it is the total years of formal education, and panel (c) it is an indicator that takes the value of one for individuals with health insurance. The points represents the average value of the outcome variable in bins of width of 1.4 *km*. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. I present the corresponding estimate of τ in equation 1 in Table 5.

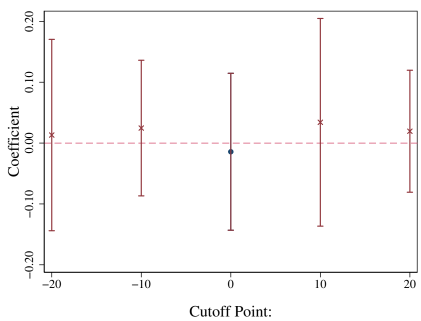
Figure D2: Spatial Placebo: Dwelling Characteristics



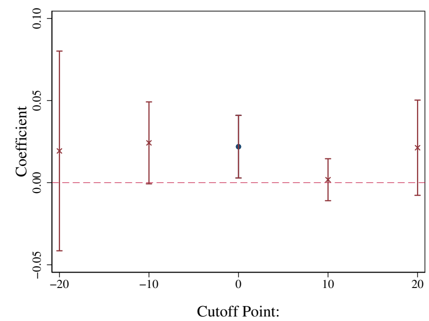
(a) Concrete Walls (=1)



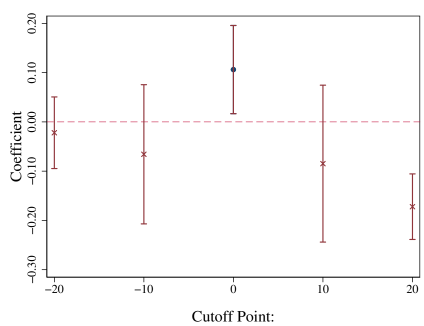
(b) Concrete Floor (=1)



(c) Electricity (=1)



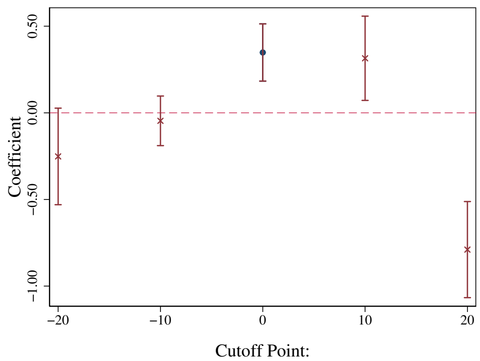
(d) Sewerage (=1)



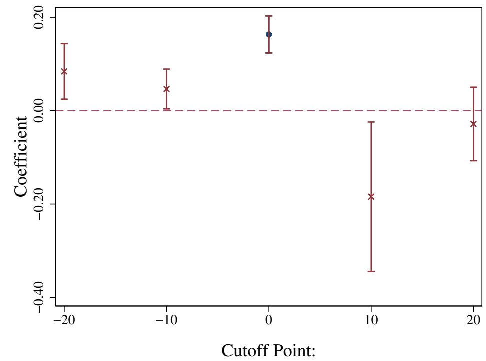
(e) Aqueduct (=1)

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is an indicator that equals one if the dwelling's walls materials are in concrete or better, while in panel (b) it is an indicator that equals one if the dwelling's floor materials are in concrete or better. Dependent variable in panel (c) is an indicator that equals one if the dwelling has access to electricity, in panel (d) it is an indicator that equals one if the dwelling has access to sewerage system, and in panel (e) it is an indicator that equals one if the dwelling has access to aqueduct system. The points represents the average value of the outcome variable in bins of width of 1.4 km. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. I present the corresponding estimate of τ in equation 1 in Table 6.

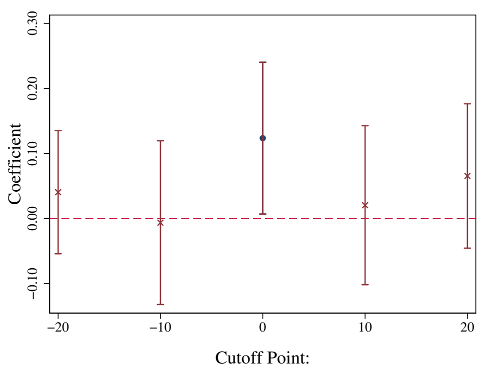
Figure D3: Spatial Placebo: Yield & Agricultural Production



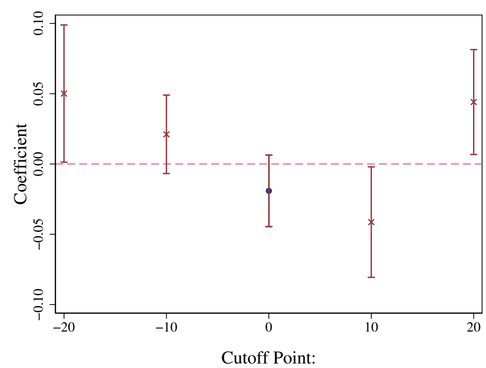
(a) Total Revenue (IHS)



(b) Revenue per Hectare (IHS)



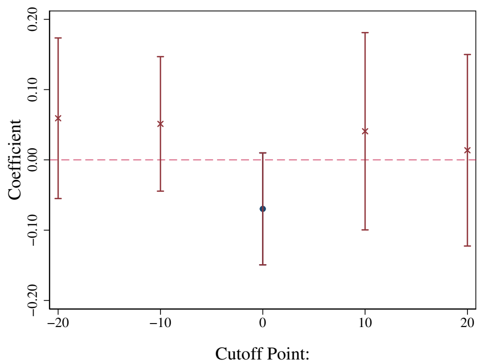
(c) Subsistence (=1)



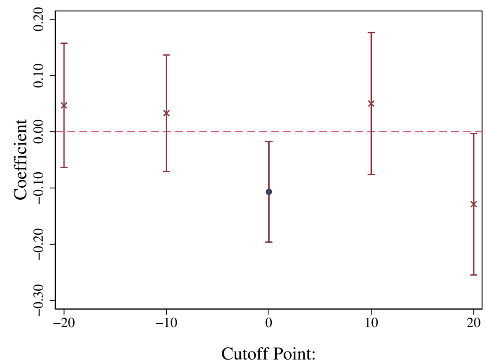
(d) Sale (=1)

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is the inverse hyperbolic sine of the total revenue in Colombian 2013 millions of COP, while in panel (b) it is the inverse hyperbolic sine of the revenue per hectare in Colombian 2013 millions of COP. Dependent variable in (c) is an indicator that equals one if the farms agricultural production is used only for self-consumption, and in panel (d) it is an indicator that equals one if the farms agricultural production is used only for market sale. The points represents the average value of the outcome variable in bins of width of 1.4 km. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas.I present the corresponding estimate of τ in equation 1 in Table 7.

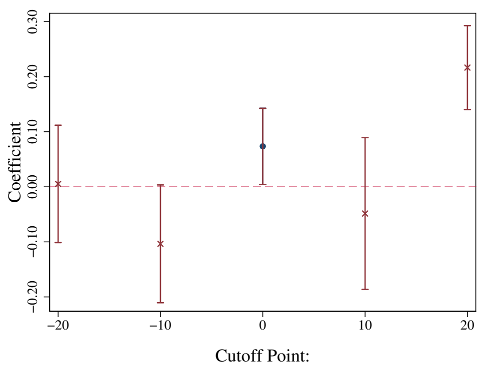
Figure D4: Spatial Placebo: Crop Specialization



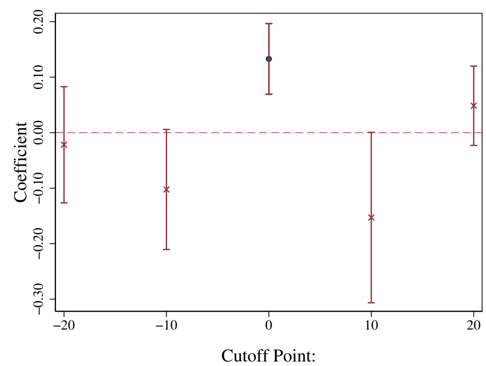
(a) Cash Crops - Area Share



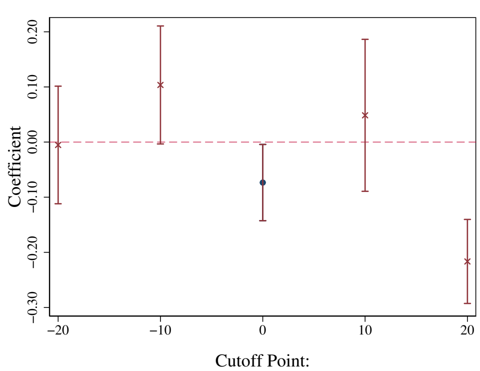
(b) Cash Crops - Revenue Share



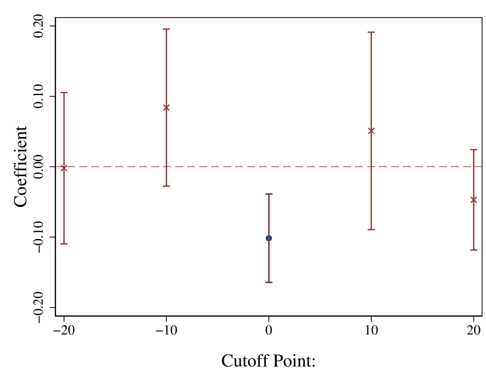
(c) Perennial Crops - Area Share



(d) Perennial Crops - Revenue Share



(e) Transitory Crops - Area Share



(f) Transitory Crops - Revenue Share

Notes: This figures plots the discontinuity at the boundary. The dependent variable in panel (a) is the share of area with cash crops, while in panel (b) it is the share of revenue from cash crops. The dependent variable in (c) is the share of area with perennial crops, while in panel (d) it is the share of revenue from perennial crops.The dependent variable in panel (e) is the share of area with transitory crops, while in columns panel (f) it is the share of revenue from transitory crops. The points represents the average value of the outcome variable in bins of width of 1.4 *km*. The regressions are estimated using local linear polynomials in the outcome of interest estimated separately on each side of the border within a fixed bandwidth of 10 km. Sample is restricted to border segments of San Vicente, La Macarena, Vistahermosa, and Mesetas. I present the corresponding estimate of τ in equation 1 in Table 8.