# Essays on the Economics of Indigenous Peoples and Violence in Latin America

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2023

vorgelegt
von
Eduardo Hidalgo, M.Sc. Volkswirtschaftslehre
aus
Ecuador

Referent: Univ.-Prof. Dr. Erik Hornung
Korreferent: Univ.-Prof. Dr. Susanne Prantl
Tag der Promotion: 31.12.2022

*To my family*

# Acknowledgment

This dissertation was an extraordinarily enriching experience. I want to thank all the people who helped me make this possible.

First and foremost, I would like to express my sincere gratitude to my supervisor Erik Hornung. Erik provided me with great advice and guidance throughout my PhD studies. Further, Erik supported me in all my endeavours, and his doors were always open when I needed help.

I also want to thank my co-authors. I had the privilege to work on two projects with Aldo Elizalde, and I would like to thank him for long hours of discussions, great laughs, and a special visit to Glasgow. A special thanks to Pablo Selaya for his trust and guidance while studying in Copenhagen and during my PhD studies in Cologne. In the same vein, I thank Nayeli Salgado whom I also co-authored a chapter of this thesis.

I want to thank the Department of Economics and the Center for Macroeconomic Research at the University of Cologne for providing the necessary tools for doing a fruitful doctoral study. First, I want to thank Susanne Prantl for being my second supervisor. For countless discussions and comments in seminars and coffee break I thank Martin Barbier, Peter Funk, Paul Schempp, Emanuel Hansen, Marius Vogel, Jonas Löbbing, Philipp Giesa, Florian Wicking, Lucas Radke, Vangjel Bita and Francesco Giovanardi. A special thanks to the Econ History group for the great comments and discussions, Carola Stapper, Melissa Rubio, Ann-Kristin Becker and the research assistants. I must thank Ina Dinstühler, Sylvia Hoffmeyer, Kristin Winnefeld, Yvonne Havertz, and Diana Frangenberg, whose support in administrative matters was vital to surviving German bureaucracy. Last, I would like to thank Stefan Hasenclever, and Matthias Kaldorf for making the past 5 years much more enjoyable.

I would like to thank my friends outside of academia for keeping me sane throughout the whole process. Thanks to all of you.

Last, I want to thank my family for their unlimited support and trust throughout all my life. Gracias Carlos, Sofía, Hubertus, Tita and Sophie. Last but most importantly, I want to thank Mapi. Without you, it would not have been possible. For this, I am forever grateful.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

The global Indigenous population is estimated to be 476 million people, which represents 6.2 % of the total population, yet they represent around 19% of the extreme poor (ILO, 2019). Latin America has an estimated 55 million Indigenous people (12% of the total population), who belong to 500 different ethnic groups. The largest Indigenous populations in Latin America are in Mexico, Guatemala, Peru, and Bolivia (Lustig and Tommasi, 2020). This is the main reason I focus my studies in Mexico and Peru in this thesis.

There are no doubts that the Indigenous people are among the most vulnerable social groups in the region. For example, in Mexico, regions with a higher Indigenous population have a higher illiteracy rate, poorer health, higher poverty and lower access to public goods (Servan-Mori et al., 2014). In Chapter 2, I will study the underprovision of public goods from the demand side perspective.

Furthermore, according to the United Nations, the Indigenous communities face a great deal of discrimination, which has culminated in genocides over the 20th century. For example, In 1932 between 10,000 and 40,000 people were killed in El Salvador, most of whom where Indigenous people, in the genocide called La Matanza ("The Massacre"). Another example is the *"Maya Genocide"* in Guatemala, where between 40,000 and 60,000 individuals where killed. In chapter 4, I study the consequences of the Peruvian civil conflict, where around 70,000 people were killed, and 85% of the victims were Indigenous people.

Violence is an important issue that not only affects the Indigenous population, but most of society faces its consequences. In Latin America, the escalation of drug-related violence in the last two decades is one of the main concerns. In Mexico, between 2007 and 2010, over 50,000 drug-related homicides were registered by the National Security Council. In Chapter 3, I study the supply chain structure of Mexican drug trafficking organizations and how a positive economic shock could terminate in an escalation of drug-related violence.

This thesis comprises three self-contained articles that provide new insights on the economics of Indigenous people, and violence in Latin America.

Chapter 2, which is joint work with Aldo Elizalde and Nayeli Salgado, expands the literature on the intersection between public goods provision and ethnic diversity (Alesina, Baqir, and Easterly, 1999, 2003; Alesina and La Ferrara, 2005; Montalvo and Reynal-Querol, 2005; Montalvo and Reynal-Querol, 2017). One recurrent finding in this literature is the negative correlation between ethnic diversity and public goods provisions because, in more heterogenous societies, it is more difficult to meet individual preferences in policy decisions, hindering public goods' supply. Yet, there is a lack of understanding about the rejection of public goods by some social groups, especially if these goods are perceived as undermining identities and traditions.

It is well known among human rights activists that Indigenous groups oppose large-scale infrastructure projects despite their potential economic benefits. Such rejection is evident in large-scale infrastructure projects by the Indigenous population across the globe. This raises several interesting questions. First, why do they reject large-scale infrastructure projects? Second, which Indigenous groups are better able to reject the implementation of a large-scale infrastructure project? And what are the mechanisms that explain their success?

We use post-revolutionary Mexico (1920 to 1950) as a natural experiment to answer these questions. After the Mexican revolution (1910 to 1920), the new state used the construction of road infrastructure as a tool for building the Mexican nation. In the words of Waters (2006), roads in post-revolutionary Mexico represented both "*a change in individual identities*...[by extending] *the process of Mexicanidad*" and a way for individuals to "*come into more direct contact with the market forces and institutions of the state*". Therefore, during this period, roads represented a threat to Indigenous identity and traditions.

We study road infrastructure development in two sets of municipalities based on the proportion of Indigenous people who descended from politically centralized societies in pre-colonial times. Using a difference-in-differences strategy, we find that municipalities with Indigenous people who descended from centralized societies in pre-colonial times received a lower share of road infrastructure than municipalities with indigenous population who descended from fragmented societies in pre-colonial times.

We propose two mechanisms that could explain why Indigenous people coming from politically centralized societies in precolonial times were better able to reject the implementation of road infrastructure in their municipalities than those with fragmented political institutions.

First, we argue that because of the legacy of pre-colonial centralized institutions, better capacities for collective action were developed in municipalities with "centralized" Indigenous institutions. Therefore, Indigenous peoples in politically centralized municipalities coordinate more effectively to get the state to meet their collective demands than in municipalities with "fragmented" Indigenous institutions. To provide evidence for this mechanism, We study an important land restitution policy from the early 20th century, in which 16 million hectares of ancestral land were redistributed to Indigenous people. The redistribution of this land was not a straightforward process, and the Indigenous people had to collectively organized to acquire their ancestral land. We show that the Indigenous population originated from politically centralized societies of the pre-colonial period, received a larger share of ancestral land. We argue that this is a result of a better collective action capability of the Indigenous people from politcally centralized societies in pre-colonial times. Furthermore, using survey data, we show that individuals who descend from politically centralized societies are more likely to engage in collective actions such as participation in demonstrations, seizure of private property or land, and road blockades than individuals from politically fragmented ethnic groups in pre-colonial times.

The second mechanism argues that Indigenous people from politically centralized societies prefer Indigenous identity more than those from politically fragmented societies. Using survey data, we find that individuals who come from politically centralized societies in pre-colonial times are more proud of their Indigenous identity, more likely to preserve their Indigenous culture and traditional practises, and more likely to identify with their Indigenous identity than with national identity than individuals who come from politically fragmented societies. Since road infrastructure represented a threat to Indigenous identity and traditions, a higher preference for Indigenous identity could explain the rejection of road infrastructure by Indigenous people from politically centralized societies in pre-colonial times.

Last, we study the long-run economic consequences of rejecting road infrastructure. We find that those more successful municipalities rejecting road infrastructure have lower access to different public goods today.

The results from chapter 2 have important policy implications. Qualitative evidence suggests that Indigenous people are not against large-scale infrastructure projects per se but they are against the imposition of "development" models that undermine their Indigenous identity and traditions within their communities. Understanding which types of Indigenous groups can better reject large-scale infrastructure projects may ensure better forms of project implementation, thus facilitating successful projects that are essential for economic growth.

Chapter 3, which is joint work with Erik Hornung, and Pablo Selaya, contributes to the understanding of the consequences of economic shocks on violence. In this chapter, we propose a novel mechanism that could explain how a positive trade shock triggered drug-related violence in Mexico. We argue that trade liberalization increases the exchange of legal goods between two countries and facilitates the traffic of illegal goods, leading to increased profits for actors in illicit markets. This increase in profits could ultimately lead to violence since firms in the illicit goods sector compete over profits by using violence due to the absence of legally enforced property rights.

We use the introduction of the North America Free Trade Agreement (NAFTA) in 1994 to study the consequences of a positive trade income shock on drug-related violence. When entering into force, NAFTA eliminated most trade barriers between Mexico and the United States, which massively increased trade between the two countries and unintentionally increased the possibility for drug-trafficking organizations to smuggle drugs into the United States (Andreas, 1996, 2012). We hypothesize that the increase in illegal goods trafficking induced by NAFTA could have increased drug-related profits, leading to an increase in drug-related violence in Mexican municipalities traversed by drug-trafficking routes.

To test the aforementioned hypothesis, we identify optimal routes for drug trafficking in Mexico by connecting locations of major drug eradication and seizures of illegal drugs in Mexico with all U.S. land ports of entry using Dijkstra's algorithm (Dijkstra, 1959). This allows us to predict which municipalities are traversed by drug-trafficking routes. Further, we combine this information with drug-related homicides at the municipality level from 1990 to 1999 and compare drug-related homicide differences between municipalities with and without a predicted drug-trafficking route before and after the introduction of NAFTA in 1994.

We find that the introduction of NAFTA increased drug-related homicides by 2.3 per 100,000 inhabitants (27% with respect to the mean) in municipalities with a predicted drug-trafficking route. We argue this increase in drug-related homicides is a consequence of the increasingly violent competition between drug-trafficking organizations to control municipalities traversed by drug-trafficking routes due to the fact that NAFTA increased trafficking possibilities.

The findings in chapter 3 could be generalized to all types of illegal goods that are smuggled across borders and human trafficking. Furthermore, the results show that policymakers should consider the unintended negative consequences of trade liberalization for future trade agreements.

In Chapter 4, I study the consequences of exposure to violence on ethnic identity. In recent years, the economic literature has shown that our (ethnic) identity influences who we are, our behaviour and how we interact with others (Akerlof and Kranton, 2010, p. 13, 28). Therefore, our ethnic identity affects our economic decisions, such as how much charity we give (Chen and Li, 2009), the amount of taxes we are willing to pay (Xin Li, 2010), the supply of public goods (Alesina, Baqir, and Easterly, 1999) the demand of public goods (Chapter 2), or the efficiency of governments and corruption (La Porta et al., 1999).

A well-established finding in this literature is the positive correlation between ethnic heterogeneity and conflict. However, our understanding is still limited regarding the consequences of violence on ethnic identity. Rohner, Thoenig, and Zilibotti (2013b) propose a theory of the relationship between violent conflict, trust, and ethnic identity. The authors argue that violent conflict between two ethnic groups leads to a collapse in generalized trust, increases in-group trust, and leads to a higher own ethnic group identification. Rohner, Thoenig, and Zilibotti (2013a) provide empirical evidence for the theoretical predictions of Rohner, Thoenig, and Zilibotti (2013b).

In chapter 4, I intend to expand the understanding between violence and ethnic identity by using the Shining Path conflict in Peru as a natural experiment. This conflict expanded between 1982 and 1992, and it was one of the most brutal civil conflicts in Latin America during the 20th century. In this conflict, more than 70.000 people were killed, and many suffered torture, rape, and displacement.

The Shining Path was a Marxist-Leninist- Maoist organization that upraised against the Peruvian state with the idea of implementing a communist society. They started their revolutionary activity by settling in rural areas and pressuring the population to embrace class identity. In this context, all people, including the Indigenous people, had the decision to abandon their identity to embrace the class identity or suffer potentially violent consequences. The Indigenous population was the most affected ethnic group in the conflict, representing 75% of the victims of this conflict while only representing 30% of the total population.

To study the consequences of exposure to violence on ethnic identity, I combine individual-level data on ethnic identity and violent event-level data from 1958 until 1992, and implement a difference-in-differences strategy to exploit quasi-random variation levels of exposure to violence across individuals. The results show that individuals exposed to violence are less likely to speak an Indigenous language and less likely to identify as Indigenous, which suggests a negative relationship between violence and ethnic identity.

These results are interesting compared to the results from Rohner, Thoenig, and Zilibotti (2013a,b), which show a positive relationship between exposure to violent conflict and ethnic identity. The main difference between the Shining Path conflict and the studies by Rohner, Thoenig, and Zilibotti (2013a,b) is the perpetrator of violence. In the case of Rohner, Thoenig, and Zilibotti (2013a,b) the perpetrator of violence is a different ethnic group than your own. Hence, this leads to lower trust in other ethnic groups and a higher ethnic identification.

In the case of the Shining Path conflict, violence was perpetuated by individuals within the same village. According to the CVR (2003), each community or family had a quota of children that had to be given to the guerrilla, and families that opposed giving their children were killed to enrol their kids later in the guerrilla. Furthermore, the Shining

Path rhetoric was mostly accepted by the young population, mainly individuals between 12 and 30 years old, because of the promise of a "new world order" would bring justice and social equality. These new guerrilla fighters would be in charge of their community and inflict violence if necessary.

I hypothesize that if violence was perpetuated by your own Indigenous group, the probability of an individual identifying Indigenous is lower than if a different Indigenous group perpetuated violence, which could explain the difference results between Rohner, Thoenig, and Zilibotti (2013a,b) and the main results of chapter 4.

To test the hypothesis mentioned above, I create an Indigenous diversity index at the district level, which measures the probability that two randomly drawn people from the same district belong to a different Indigenous group. Using this index, I find that in more Indigenous homogenous districts, when the probability of suffering violence from your own Indigenous group is higher, individuals exposed to violence are less likely to identify as Indigenous than in Indigenous heterogeneous districts.

The results in chapter 4 expand our understanding of how violence shapes ethnic identity. I provide empirical evidence that violence has heterogeneous effects on ethnic identity depending on whether violence was perpetrated by a different ethnic group or your own ethnic group. If a different ethnic group committed violence, we could expect an increase in ethnic identity. However, if your ethnic group committed violence, we expect a decrease in ethnic identity.

### Contribution

Chapter 2 is based on joint work with Aldo Elizalde and Nayeli Salgado. The research idea and the framework were developed in collaboration. Nayeli Salgado provided the road infrastructure data. Aldo Elizalde provided the rest of the data, and worked on the background, and edited the draft. I worked on the theoretical framework, the empirical analysis and the presentation of the results. The latter was largely discussed with Aldo Elizalde as well as the contribution of this chapter.

Chapter 3 is based on joint work with Erik Hornung and Pablo Selaya. The research idea and the framework were developed in collaboration. Erik Hornung worked on placing the project in the literature and the contribution to the literature, editing the draft, and on designing different robustness checks and result extensions. Pablos Selaya provided background knowledge. I contributed to the project by working on the data-cleaning process, the empirical strategy, and the presentation of the main results.

Chapter 4 is a solo project.

# Chapter 2

# Public good or public bad? Indigenous institutions and the demand for public goods

This chapter is based on Elizalde, Hidalgo, and Salgado (2021).[1]

## 2.1 Introduction

A prominent literature in economics argues that the underprovision of public goods is associated with ethnic diversity.[2] In particular, it is well documented that in countries with more ethnically diverse societies, it is more difficult to meet individual preferences in policy decisions. Less well documented is that a low provision of public goods may be associated with their rejection by some social groups, especially if these goods are perceived as undermining identities and traditions. Such rejection is evident in large-scale infrastructure projects. While these infrastructure projects can have positive economic effects, they often have negative impacts on social cohesion, livelihoods, the environment and the health of Indigenous people.[3] It is well known among human rights activists that Indigenous groups oppose large-scale infrastructure projects despite their potential economic benefits.[4] However, such explanations for the failure to provide public goods are

[1]We are indebted to Erik Hornung for detailed comments and discussions. We also thank Siwan Anderson, Gustav Agneman, Patrick Allmis, Luis Angeles, Luz Marina Arias, Timothy Birabi, Martin Benedikt Busch, Kasper Brandt, Carola Engelke, Alfonso Herranz, James Fenske, Susanne Prantl and Marius Vogel for helpful comments, and seminar and conference audiences at Almeria, The University of British Columbia, Cardiff, Cologne, Glasgow, IESG, EHS Annual Conference, NEUDC2020, LANE-HOPE Seminar, EEA Congress, EHES Conference, Virtual Economic History Seminar, Ibero-American Economic History Webinar, BCDE conference, VfS Conference, SAEe2020 and SIE-RSA. Aldo Elizalde gratefully acknowledges financial support from the British Academy Post-Doctoral Fellowship. Nayeli Salgado gratefully acknowledges financial support from the Institute of Economic and Social History of the Department of Socioeconomics at the Vienna University of Economics and Business (WU) and from the Austrian National Bank.

[2]See Alesina, Baqir, and Easterly (1999), Alesina et al. (2003); Alesina and La Ferrara (2005); Montalvo and Reynal-Querol (2005); Montalvo and Reynal-Querol (2017).

[3]See Duflo and Pande (2007); Lipscomb, Mobarak, and Barham (2013); Hornung (2015) Baum-Snow et al. (2017); Donaldson (2018)

[4]See UN Human Rights Council reports A/HRC/39/17/Add.2 for Mexico; A/HRC/39/17/Add.3 for Guatemala; A/HRC/33/42/Add.2 for Honduras; A/HRC/36/46/Add.1 for the United States

largely neglected in the economics literature. Moreover, we do not know exactly which Indigenous groups are better able to achieve rejection and which mechanisms explain their success.

This paper investigates whether Indigenous people who descend from politically centralised societies in pre-colonial times are better able to reject the implementation of large-scale infrastructure projects than Indigenous people who descend from politically fragmented societies. To this end, we use heterogeneity in the proportion of Indigenous people who descend from different levels of political organisation in pre-colonial times. In particular, we use an index of Indigenous institutions developed by Elizalde (2020a).[5] This measure is the Indigenous population-weighted average of Murdock (1967)'s *Jurisdictional Hierarchy Beyond the Local Community* index for all Indigenous groups within each municipality.[6] Following Gennaioli and Rainer (2007), the degree of political complexity of ethnic institutions captures the ability of ethnic groups to coordinate and implement policies at the local level. In this study, we use this variation to examine the demand for large-scale infrastructure projects in municipalities with varying degrees of ethnic institutions. Our aim is therefore to understand what kind of Indigenous groups reject large-scale infrastructure projects.

To explore the above, we use a quasi-experimental framework from post-revolutionary Mexico. After the Mexican Revolution (1920 to 1950), the new state used the creation of a large-scale road-building programme as a tool for nation-building. In the words of Waters (2006), roads in post-revolutionary Mexico represented both "*a change in individual identities*...[by extending] *the process of Mexicanidad*" and a way for individuals to "*come into more direct contact with the market forces and institutions of the state*".[7] Therefore, during this period, roads represented both a threat to Indigenous identity and traditions and a public good that could bring economic development.

Our study setting offers two key advantages over contexts where policies target specific populations or contexts where there is insufficient ethnic heterogeneity. First, because the new state aimed to create a national identity by reaching out to all strata of society, no distinction was made between Indigenous and non-Indigenous populations in the provision of road infrastructure (Waters, 2006). However, roads posed a threat to Indigenous identity and traditions as they were replaced by a new national identity, which is why Indigenous people rejected road infrastructure (Bess, 2017). Second, Indigenous groups in Mexico before colonisation had very different political structures and institutional complexity:

[5] See also Gennaioli and Rainer (2007); Michalopoulos and Papaioannou (2013); Michalopoulos and Papaioannou (2014); Dippel (2014); Angeles and Elizalde (2017)

[6] This index of ethnic institutions ranges from 0-2, with higher values representing communities with indigenous populations descended from ethnic groups with more complex political organisation in pre-colonial times.

[7] *Mexicanidad* refers to a process of creating a unified nation under a single national identity and language: Mexican or Spanish. It should therefore be understood as a process of cultural unification in which pre-existing cultures and ethnic identities are abolished

from politically centralised systems administered by the Aztec Empire to numerous scattered nomadic groups whose political organisation did not go beyond the village level (e.g. the Huichol group in northern Mexico) (Adams and Macleod, 2000; Angeles and Elizalde, 2017). As Europeans were unable to rule over thousands of villages and towns, existing systems of governance remained in the form of collective forms of organisation at the local level (Angeles and Elizalde, 2017; Elizalde, 2020a). We use these two unique sources of variation to improve our understanding of which Indigenous groups - politically centralised and politically fragmented - are better able to reject the implementation of a large-scale road-building project.

To test our main hypothesis, we construct a dataset containing a new digital map of the Mexican road network, in which the expansion of roads since 1920 has been coded.[8] To our knowledge, we are the first to evaluate the evolution over time of the road network in Mexico since the beginning of the 20th century.

In our main econometric strategy, we exploit the fact that the state used the provision of road infrastructure as a nation-building tool after the Mexican Revolution. We use a difference-in-differences approach to compare differences in road infrastructure between municipalities with different levels of Indigenous institutions before and after the start of the road construction programme (1925). We find that municipalities with a higher share of Indigenous people who descend from politically centralised societies in pre-colonial times experienced lower road expansion than municipalities with a larger share of Indigenous people whose ancestors belong to politically fragmented groups.[9] To capture the timing of differential road infrastructure development, we use a flexible difference-in-differences approach. We show that differences in road infrastructure provision became significant after 1940, when road infrastructure in Mexico was significantly expanded. Our results are robust to different specifications of our main outcome and explanatory variables.

The validity of our identification strategy rests on the assumption that road infrastructure in municipalities with different levels of Indigenous institutions would have followed parallel trends had it not been for the process of nation-building and the use of roads as a tool to create a national sense of *Mexicanidad*. We provide evidence for this assumption in two ways. First, using newly digitised data on colonial roads and the railway network

[8] Roads are defined as all federal and state highways that are part of the Mexican road network. According to the Law on Roads, Bridges and Federal Motor Transport, federal and state highways are those that connect with roads from abroad, link two or more subnational states and are generally built by the government with federal funds.

[9] We classify politically centralised and politically fragmented municipalities according to Dippel (2014). Specifically, he classifies politically centralised Indigenous groups as those with a value of 1 (small chiefdoms) and 2 (larger chiefdoms) from Murdock's Jurisdictional Hierarchy Index, and as politically fragmented those with a value of zero (stateless societies). Therefore, we classify a municipality as politically *centralised* if the value of our index of Indigenous institutions is greater than zero; if the value of this index is zero, we classify municipalities as politically *fragmented*.

up to 1920, we find that our measure of Indigenous institutions does not correlate with the provision of infrastructure projects in Mexico before the revolution. Second, we take advantage of the shift in road infrastructure policy in the mid-1950s, when road construction increasingly focused on modernisation policy rather than nation-building. We find that post-1950s road infrastructure follows parallel trends between municipalities with different levels of Indigenous institutions. This suggests that the different trends in road infrastructure were a consequence of the way the state used roads for nation-building in post-revolutionary Mexico (1920s to 1950s).

Furthermore, we perform an additional robustness check to confirm our main hypothesis. We hypothesise that Indigenous people reject the provision of large-scale infrastructure projects when the state attempts to use these projects as a tool for nation building. Therefore, Indigenous people in post-revolutionary Mexico did not see roads as a public good that could bring prosperity to their communities, but rather as a public "bad" that could affect their identity and traditions by integrating Indigenous communities into the notion of *Mexicanidad*. Therefore, if our hypothesis is correct, we expect not only a reduced provision of road infrastructure, but also a reduced state presence in politically centralised municipalities. We explore this question using data on the number of bureaucrats as a proxy for state presence for each Mexican municipality between 1900 and 1940, drawn from Garfias (2018). We show that during the nation-building process, the new state was significantly less intrusive in municipalities with more Indigenous populations that had politically centralised characteristics in the pre-colonial period. This result suggests that Indigenous people in politically centralised municipalities were better able to reject the state presence than in politically fragmented municipalities.

It is indeed a valid concern that the link between Indigenous institutions and road infrastructure may not show that Indigenous people rejected roads. To validate the main argument of this paper, we would need extensive data on Indigenous uprisings against road construction during the period under study (1930s to 1950s). To our knowledge, however, this data does not exist in Mexico. However, we do not believe that this invalidates the results of our study. The main objective of this study is to improve our understanding of which Indigenous groups oppose large-scale infrastructure projects. To this end, we use differences in the degree of Indigenous institutions that capture the ability of Indigenous groups to coordinate collective action to persuade the state to meet their demands. We provide qualitative evidence of various Indigenous uprisings against road construction during the nation-building period. We document that Indigenous people resisted road building because the nation-building process aimed to override Indigenous identities and traditions by improving connectivity through roads.

We perform different checks to rule out the possibility that supply-side discrimination drives our results. First, we check whether municipalities with non-Indigenous people receive a higher share of road infrastructure than municipalities with Indigenous people. The results show no discrimination of road provision to municipalities with Indigenous people. Second, we built a counterfactual road network, which maximised population connectivity between the state capitals and Mexico City. We show that after controlling for the counterfactual supply-side road network, our variable of interest, the Indigenous Institution index, is still negatively associated with road development between 1920 and 1960. We interpret this result as a lack of demand for road infrastructure by municipalities with centralised Indigenous institutions. Last, we control for different measures that the literature has shown to correlate with the supply of public good provision and ethnic diversity, such as ethnic fractionalisation and polarisation, and the share of Indigenous people in each municipality.

We suggest two possible mechanisms that could explain our main findings. The first mechanism assumes that politically centralised Indigenous groups are better able to coordinate because they have better capacity for collective action and can thus push through their political demands. We illustrate this mechanism with an important land restitution policy from the early 20th century, in which 16 million hectares of ancestral land were redistributed to Indigenous people (Elizalde, 2020a). The redistribution of ancestral land was not an easy process, as there were powerful large landowners in the rural areas of Mexico who regularly blocked land petitions. Therefore, the restitution of land required permanent collective organisation of the Indigenous population (land occupations, protests, barricades, etc.). In line with Elizalde (2020a), we find that in municipalities where the Indigenous population originated from politically centralised societies of the pre-colonial period, a larger share of ancestral land was redistributed. In addition to this evidence, we use recent survey data on Indigenous attitudes and beliefs from the Latin American Public Opinion Project (LAPOP). We find that individuals who descend from politically centralised societies are more likely to engage in collective actions such as participation in demonstrations, seizure of private property or land, and road blockades than individuals from politically fragmented ethnic groups in pre-colonial times.

Our second mechanism states that municipalities with a politically centralised Indigenous population have a higher preference for preserving their Indigenous identity and consequently a lower preference for road infrastructure. To prove this mechanism, we use survey data on individuals' preferences regarding Indigenous identity, taken from the "National Survey on Discrimination" of Mexico (ENADIS). From this survey, we select several questions on attitudes towards an individual's national identity. We find that individuals who come from politically centralised societies of the pre-colonial period are more proud of their Indigenous identity, more likely to preserve their Indigenous culture and traditional practises, and more likely to identify with their Indigenous identity than with national identity than individuals who come from politically fragmented societies.

We exclude alternative mechanisms that could explain our main results. We argue that the differences in road infrastructure are due to the lack of Indigenous demand. However, one could also argue the opposite, that the state provided less road infrastructure in municipalities with a higher proportion of Indigenous population, suggesting discrimination. To rule out this alternative channel, we first compare road infrastructure development between municipalities with and without Indigenous populations in post-revolutionary Mexico and find that road development is virtually the same between these two groups of municipalities. Second, we control for the proportion of Indigenous population in all our specifications to address concerns about discriminatory policies towards Indigenous people. Third, a large body of literature argues that ethnic diversity affects the provision of public goods.[10] To examine ethnic diversity as a factor influencing the provision of roads, we control for measures of ethnic fractionalization and ethnic polarisation in all our specifications according to Alesina et al. (2003) and Montalvo and Reynal-Querol (2005).

Finally, we turn to the economic consequences of the differential development of road infrastructure in post-revolutionary Mexico. We find that the total stock of road infrastructure in the late 1950s is positively correlated with a variety of contemporary economic outcomes (electricity rate, drainage rate, literacy rate, and nighttime lighting density). These findings have important policy implications. According to the United Nations, Indigenous peoples around the world tend to oppose large-scale infrastructure projects. However, our results suggest that successfully rejecting these infrastructure projects can lead to lower development outcomes in the long run.

Our paper relates to several strands in the literature. First, we link to a number of studies that examine the role of ethnicity as an important determinant in the provision of public goods.[11] For example, the seminal work of Alesina, Baqir, and Easterly (1999) argues that areas with greater ethnic diversity provide fewer public goods. Our main contribution is to show a new way of under-provision of public goods: the "outright" rejection by ethnic groups when public goods undermine their identity and traditions. We show that Indigenous groups that are better able to coordinate collective action are more successful in rejecting the provision of large-scale infrastructure projects. We find that the first large-scale road building programme in post-revolutionary Mexico was less successful in municipalities with more politically centralised Indigenous groups. We provide evidence that this rejection may be due to the fact that the provision of road infrastructure was used to override pre-existing ethnic identities and cultural traditions.

Second, our work is related to recent findings on the impact of pre-colonial institutions on contemporary outcomes. Cross-national and within-country studies have found

[10] For a summary of this literature, see: Alesina and La Ferrara (2005)

[11] Alesina et al. (2003); Alesina and La Ferrara (2005); Montalvo and Reynal-Querol (2005); Montalvo and Reynal-Querol (2017), Alesina, Baqir, and Easterly (1999)

a positive relationship between more advanced pre-colonial institutions and contemporary economic outcomes around the world.[12] For Africa, Gennaioli and Rainer (2007) document a strong and positive correlation between the provision of public goods and the degree of political centralisation of pre-colonial ethnic groups. For the Americas, Angeles and Elizalde (2017) argue that more advanced pre-colonial Indigenous groups would have been able to organise and defend their collective interests, leading to better economic outcomes.For Asia, Dell, Lane, and Querubin (2018) compares contemporary economic outcomes between Vietnamese areas that were under centralised and decentralised states in pre-colonial times. They argue that due to centralised administrative forms of organisation in the pre-colonial period, villages under centralised pre-colonial states have developed better local collective actions throughout the colonial period until today. Our main findings contribute to this growing literature by demonstrating that the impact of pre-colonial institutions on policy outcomes can also have unintended negative consequences for long-term development. We show that in municipalities with more politically centralised Indigenous groups, large-scale infrastructure projects are more likely to be rejected.

Our work is also related to a number of studies that examine the impact of large-scale infrastructure projects on economic outcomes.[13] These studies have documented that infrastructure projects produce winners and losers.[14] For example, Duflo and Pande (2007) show that rural poverty in India increased where dams were built. We contribute to this literature by showing that in areas with politically centralised Indigenous groups, road infrastructure development in Mexico was significantly lower, leading to worse long-term economic outcomes. Our findings have important policy implications. Recently, human rights activists have highlighted the negative impacts of large-scale infrastructure projects on Indigenous outcomes. Understanding which types of Indigenous groups are better able to reject these infrastructure projects could help governments adapt their negotiation strategies to be more successful and less threatening to Indigenous communities.

Finally, our findings add to the existing literature analysing the main determinants of cultural persistence.[15] An important question in this literature is why do traditions persist in some settings and not in others? (Tabellini, 2008). In a recent paper, Giuliano and Nunn (2020) addressed this question in the particular context of Indigenous peoples in the US. They show that Indigenous peoples living in areas where the environment has been unstable for a long time are more likely to preserve their traditional languages. We build on this literature with our findings by demonstrating which types of Indigenous

[12] See: Michalopoulos and Papaioannou (2013); Michalopoulos and Papaioannou (2014); Chiovelli (2014) Bandyopadhyay and Green (2016); Elizalde (2020a)

[13] Duflo and Pande, 2007; Lipscomb, Mobarak, and Barham, 2013; Baum-Snow et al., 2017; Donaldson, 2018

[14] See in particular: Duflo and Pande (2007); Baum-Snow et al. (2017); Asher and Novosad (2020)

[15] For a summary see: Nunn (2012)

groups are better able to preserve cultural traditions. We demonstrate that in Mexican municipalities where Indigenous people were better able to reject cultural integration during the nation-building process in the form of better transport access, Indigenous identities and traditions are more likely to survive today.

## 2.2 Background

The first national road network in Mexico was not created until after the Mexican Revolution and had two main objectives. First, it was to serve as an instrument for nation-building. In the words of Waters (2006), the post-revolutionary state saw road infrastructure as *"something that could help bring Mexicans together if controlled by the national government…[it] was a nationalist project that via discourses and practices forged national identity"*. Second, the construction of the road network was intended to connect people and markets in order to promote economic progress (Bess, 2017).

In 1918, Venustiano Carranza, the first post-revolutionary president, stressed the need to begin the "social reconstruction of the nation" through road infrastructure (Bess, 2017). However, the national road building programme did not begin until 1925. From the beginning, the development of the road network served as an instrument of nation-building. The inauguration of roads was often reported nationally and large celebrations were held to convey nationalist messages. In 1936, at the inauguration of a road in the village of Tepoztlan in the southern state of Morelos, an Indigenous resident addressed Mexican President Lazaro Cadenas in Nahualt (the lingua franca of the former Aztec Empire) and Spanish: "*like a doctor, you have identified the cause of our illness [poverty] and administered the right medicine [roads] to alleviate it*". The roads have helped isolated communities develop a new sense of national integration. Waters (2006) expresses this clearly: roads in post-revolutionary Mexico signified both "*a change in individual identities*...[by expanding] *the process of Mexicanidad*" and a way for individuals to "*come into more direct contact with market forces and the institutions of the state*".

Road building generally took place at the regional and local level (Bess, 2014). Roads brought many positive outcomes for communities by connecting them to towns and markets, but also posed significant threats to the natural environment and the social cohesion of communities, including Indigenous villages. To minimise confrontations, local governments relied on board meetings attended by bureaucrats and residents to influence the provision of road infrastructure, including on issues related to construction (Bess, 2017; Bess, 2014).

The community was also involved in the construction of roads. The formation of brigades at the local level was encouraged throughout the country. In some regions, for example, road construction even became a grassroots activity, with people volunteering to do collective days of work for road construction. This not only provided jobs, but also led to communities engaging in the process of state consolidation as part of national integration (Waters, 2006).

The provision of road infrastructure was also supported by a highly centralised state bureaucracy. In 1925, a National Roads Authority (NRA) (or *"Comision de Caminos"*) was created. The NRA was not only responsible for managing revenues for road infrastructure, but also controlled the entire logistical infrastructure of road construction, from the provision of machinery and labour to the supply of technical expertise and construction work (Bess, 2017). The state essentially centralised the provision of roads throughout the Mexican territory.

Between the 1920s and the 1950s, 31,094 kilometres of roads were built, representing 41% of the total road stock in 1990. We focus on this period in our study because this was the phase in which the new state consolidated its power (Hodges and Gandy, 1983). For example, Manuel Avila Camacho, president of Mexico between 1940 and 1946, declared in 1941: "*it is not possible to truly integrate a sense of the nation without an ample road network that facilitates economic exchange [and] connects human groups*" (Bess, 2014).

After the 1950s, modernisation policies became more important than nation building. Public infrastructure was therefore increasingly seen as an important element in promoting local markets and economic growth. This led to a significant change in public policy, particularly in the provision of road infrastructure, which was gradually developed to meet the needs of the private sector and foreign multinational companies (Hodges and Gandy, 1983). Then, in the late 1980s, road infrastructure was substantially taken over by the private sector through private concessions known as Built, Operate and Transfer. International companies thus increasingly took over responsibility for road infrastructure, from technical implementation and investment to construction and maintenance (Blankespoor et al., 2017).

Nevertheless, the road infrastructure in Mexico was developed in different ways. In particular, road development in the southern states was apparently less than in those north of central Mexico. These contrasting patterns are important when considering that these regions often differ not only in their distinct Indigenous heritage, but also in the degree of state presence. In three of the southern states with more entrenched Indigenous practises, Chiapas, Guerrero and Oaxaca, some of their Indigenous communities have managed to develop considerable autonomy in policy implementation through political dialogue or conflict (Harvey, 1998; Díaz-Cayeros, Magaloni, and Ruiz-Euler, 2014)[16]. In these communities, the presence of the state was mostly weak.

Qualitative evidence suggests that Indigenous communities are opposed to road infras-

[16]The most prominent examples of Indigenous autonomy in the southern states of Mexico are the Indigenous Mayan guerrillas in Chiapas and the use of Indigenous practises (*usos y costumbres*) to elect community leaders in Oaxaca.

tructure. Indigenous people viewed road infrastructure not as a public good that would improve their standard of living, but as a "threat" to the continuation of their existing forms of social organisation and traditions (Waters, 2006; Bess, 2017). This means that the provision of road infrastructure came into conflict with the ethnic identity and traditions of Indigenous people, leading to resistance to road construction within their communities.

We argue that resistance to road infrastructure was only possible through the concerted and sustained organisation of Indigenous groups to resist it - an institutional legacy that can be traced back to pre-colonial times. At the time of first contact with Europeans, Indigenous groups in Mexico had very different political structures and institutional complexity, ranging from city-states administered by the Aztec Empire to numerous small tribes whose political organisation did not go beyond the village level (Adams and Macleod, 2000). Colonialism, however, intervened deeply in this established institutional order and led to a radical change in the ethnic structure of the Indigenous population.

However, pre-colonial institutional factors were not erased by the arrival of Europeans and continue to influence outcomes to this day (Angeles and Elizalde, 2017). Indigenous populations remained within their ancestral territories thanks to colonial policies to control them. These Indigenous settlements were referred to as *Pueblos de Indios*. Importantly, they enjoyed self-government throughout the colonial period, so that the Indigenous population could continue to use their traditional forms of organisation to meet their collective needs (e.g. labour and taxes) (Tanck de Estrada, 2005). Indigenous people thus played a key role as the main source and organiser of labour during the colonial period (Lang, 1975). In the words of Lang (1975), "*The Spanish enterprise in the New World rested on the indigenous social order*".

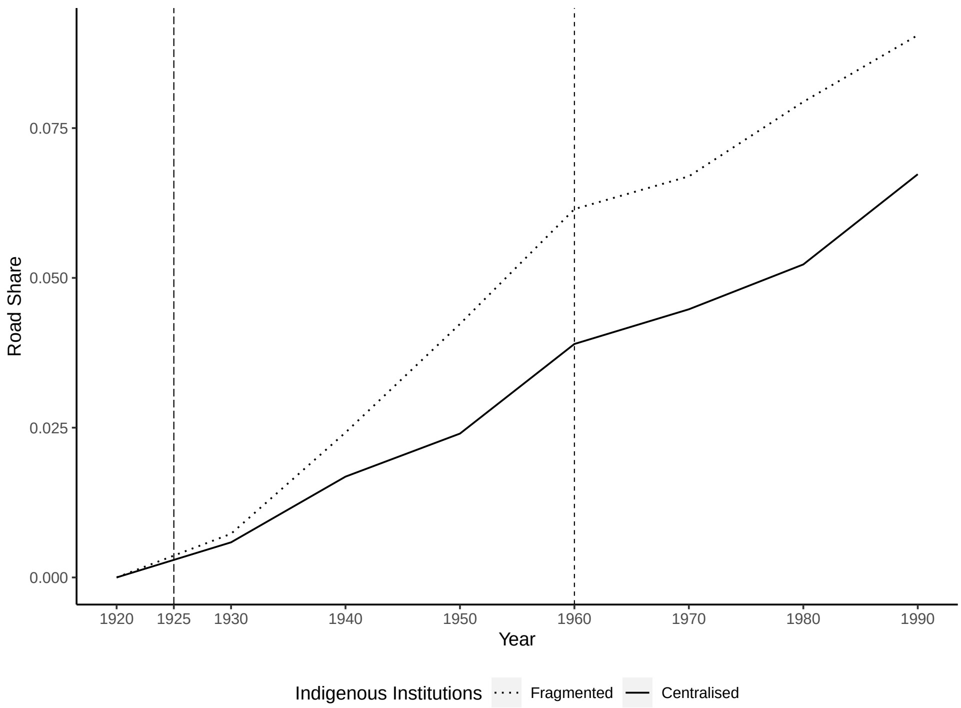



FIGURE 2.1: Road infrastructure and Indigenous institutions in Mexico, 1920-1990

*Notes:* This figure shows the development of road infrastructure in two groups of municipalities with different levels of Indigenous institutions. The "Centralised" group (solid line) are the municipalities with an Indigenous institutions index above 0. The "Fragmented" group (dashed line) are the municipalities with an Indigenous institutions index of 0. The vertical dashed line in 1925 represents the beginning of the nation building period in Mexico, and the vertical dashed line in 1960 represents the end of the nation building period. Road infrastructure is the proportion of the municipal area covered by a 2 km buffer along roads.

Figure 2.1 shows road infrastructure in two groups of municipalities between 1920 and 1990. The solid line includes municipalities with Indigenous populations originating from centralised societies of the pre-colonial period. The dashed line includes municipalities with more Indigenous people originating from fragmented societies. The two vertical lines mark the beginning of the road building programme in 1925 and the end of the nation-building period in 1960 respectively. The differences in road building between these two groups are striking. During the nation-building period from the 1920s to the 1950s, roads were constructed more rapidly in municipalities where the Indigenous population had a legacy of fragmented societies, while in municipalities where the majority of the population can be associated with centralised societies, road construction was much slower. In this study, we aim to examine these contrasting patterns of road infrastructure in more detail, particularly between the 1920s and 1950s, when state power was consolidating in Mexico.

## 2.3 Hypothesis development

To understand why road infrastructure differs in municipalities with different levels of pre-colonial Indigenous institutions, we develop a conceptual framework based on the description of post-revolutionary Mexican history. The expansion of the road network after the Mexican Revolution was essentially a manifestation of state power and nation building. Therefore, road infrastructure in post-revolutionary Mexico was both a public good and a symbol of the adoption of a new national identity for the Indigenous population.

We present a model for the link between road infrastructure and Indigenous institutions. We focus on the demand side for roads because we assume that the supply side, the Mexican state, does not discriminate against certain groups of people because it wants to create the idea of *Mexicanidad* (a unified nation). Therefore, the demand side should determine the allocation of road infrastructure during the period of nation building.

Municipality $i \in I$ can invest part of its resources $0 \leq x \leq 1$ (time, money, etc.) in avoiding roads $r^i(x)$, where $r^i(\cdot)$ is an increasing and concave function. As in the empirical analysis, roads can be viewed as proxies for state dominance. Let us assume that the preferences of this municipality (we abstract from heterogeneity within a municipality) can be represented by the increasing and quasiconcave utility function:

$$u^i(r^i(x), 1-x). \tag{2.1}$$

We denote the partial derivative of a function $f(\cdot)$ w.r.t. to its $k$-th argument by $f_k(\cdot)$ and let

$$x^i \in \arg\max_x u^i(r^i(x), 1-x). \tag{2.2}$$

For an interior solution, $x^i$ is characterized by the FOC

$$r_1^i(x^i) = \frac{u_2^i(r^i(x^i)), 1-x^i)}{u_1^i(r^i(x^i)), 1-x^i)}. \tag{2.3}$$

It follows from this optimality condition that the optimal amount of resources invested $x^i$ - and hence the optimal amount of roads avoided $r^i(x^i)$ - may differ between municipalities for two reasons: First, the efficiency of converting invested resources into avoided roads $r^i(\cdot)$ might differ between municipalities with different levels of Indigenous institutions (e.g. due to differences in the ability to collectively organise resistance to the central state). Second, the willingness to give up resources to avoid more roads might differ $u_2^i(\cdot)/u_1^i(\cdot)$ (e.g. due to a higher preference for preserving Indigenous identity and way of life). From now on, we assume that municipalities can be ordered in a meaningful way, with a higher index $i$ corresponding to "centralised" Indigenous institutions and an index $j$ corresponding to "fragmented" Indigenous institutions.[17]

[17] We follow Gennaioli and Rainer (2007) for this classification

**Hypothesis 1**

$$i > j \quad \Rightarrow \quad \min\left\{r^i(\cdot) - r^j(\cdot), \frac{u_2^i(\cdot)}{u_1^i(\cdot)} - \frac{u_2^j(\cdot)}{u_1^j(\cdot)}\right\} \geq 0. \tag{2.4}$$

This assumption states that a "centralised" municipality can avoid more roads for all the resources invested, or is willing to give up more resources to avoid an additional road (or both). The idea is that the observed heterogeneity in road avoidance can be attributed to differences in the ability to avoid roads and preferences for avoiding roads. Note that empirically we only have data on roads avoided and not on resources invested. This leads to the following mechanisms as to why "centralised" municipalities would avoid a greater central state presence than "fragmented" municipalities.

**Mechanism 1: collective action**

$$i > j \quad \Rightarrow \quad r^i(x^i) \geq r^j(x^j). \tag{2.5}$$

The first mechanism states that a "centralised" municipality will choose to avoid a greater number of roads because it has a lower cost of avoiding roads than a "fragmented" municipality. This would mean that municipalities with "centralised" Indigenous institutions are better able to coordinate to meet their collective demands. The following proposition provides examples of utility functions (in increasing order of generality) that are consistent with the model presented above.

**Mechanism 2: identity preferences**

$$r^i(\cdot) = r^j(\cdot) \quad \Rightarrow \quad \frac{u_2^i(\cdot)}{u_1^i(\cdot)} > \frac{u_2^j(\cdot)}{u_1^j(\cdot)} \tag{2.6}$$

The second mechanism states that if "centralised" and "fragmented" municipalities have a similar cost function for avoiding roads, the marginal rate of substitution is higher for"centralised" municipalities than for "fragmented" municipalities. This implies that "centralised" municipalities receive a higher utility for an additional unit of avoided roads than "fragmented" municipalities due to stronger preferences for self-determination and the preservation of traditional traditions. Consequently, the number of avoided roads would be higher in "centralised" municipalities than "fragmented" municipalities, because "centralised" municipalities would be willing to invest more resources in avoiding roads.

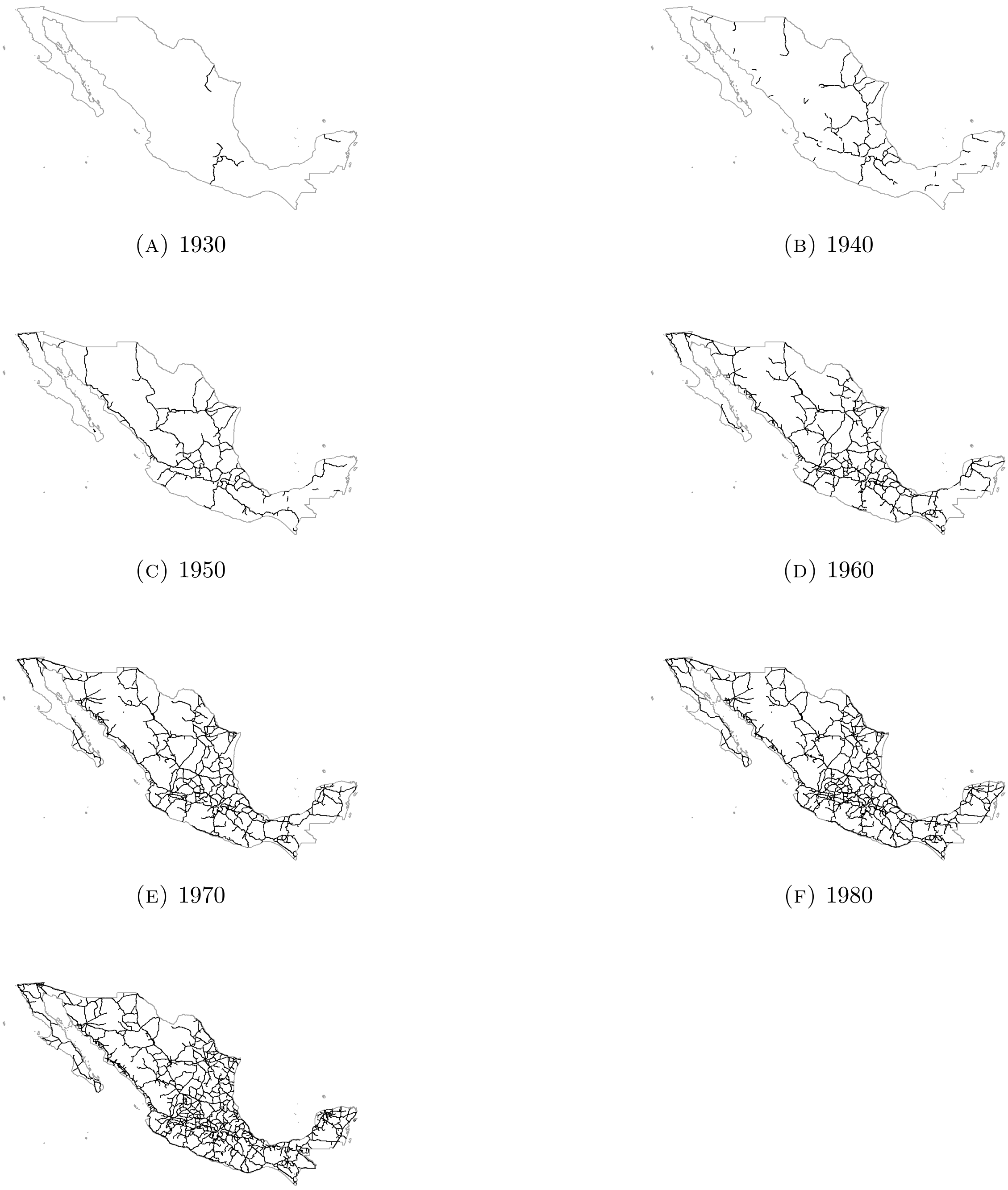
## 2.4 Data and empirical framework

### 2.4.1 Construction of outcome: expansion of road infrastructure

Our dependent variable is road share, which is defined as the proportion of the municipal area covered by a 2 km buffer along roads, following Dalgaard et al. (2018).[18] Our main period of study is the 1920s to 1950s, when roads were used as a tool for nation building. Figure 2.2 shows the spatial development of road infrastructure in Mexico from 1930 to 1990. It can be seen that road construction developed into a large-scale infrastructure project during the 20th century. In 1930, more than 2,000 municipalities had no road infrastructure, but by 1990 this number had halved.

We constructed our dependent variable by relying on unique collections on Mexican transport systems obtained from the Mexican Ministry of Transport. In our digital map, we geocoded the extent of road length in each decade after 1920, when roads were not built in Mexico. To our knowledge, this is the first comprehensive attempt to map the evolution of Mexico's road network over time. The online appendix contains a detailed description of how this variable was constructed.

[18] To capture the adjacent investments along the European road network, Dalgaard et al. (2018) apply a 5 km buffer instead. Since we mainly want to capture the effects of the roads themselves in our study, we reduce the buffer to 1 km on each side. Yet, as robustness checks in section 2.4.6.2 we apply buffers of different sizes. Figure A.4 shows an example of a road lane and the 1 km buffer along both sides.



(A) 1930

(B) 1940

(C) 1950

(D) 1960

(E) 1970

(F) 1980

(G) 1990

FIGURE 2.2: Road infrastructure in Mexico between 1930-1990.

*Notes:* This figure shows the development of road infrastructure in Mexico between 1930 and 1990. The geo-referenced maps were produced from unique collections published by the Mexican Ministry of Transport and Communications, showing the cartography of the Mexican road network since the beginning of the 20th century. The maps were georeferenced using the 2015 national road network. Sources: SCT (1988), SCT (1987) and *Instituto Nacional de Estadistica y Geografia.*

### 2.4.2 Construction of variable of interest: Indigenous institutions

We measure Indigenous institutions following the literature on long-run development.[19]. Specifically, we use the index of Indigenous institutions developed by Elizalde (2020a). The author combines detailed census data on the proportion of each Indigenous group in all Mexican municipalities from the early twentieth century with anthropological information on the political complexity of Indigenous groups as contained in Murdock's Ethnographic Atlas (Murdock, 1967). From this Atlas, the variable *Jurisdictional Hierarchy Beyond the Local Community* was used. This variable classifies ethnic groups based on their different levels of political complexity in pre-colonial times. The variable ranges from 0 to 4, with 0 denoting ethnic groups that have no political authority beyond the local community (stateless societies); 1 denoting small chiefdoms; 2 denoting larger chiefdoms or small states; and finally, 3 and 4 denoting states and larger states (see the online appendix for more details on the sources and construction of this index). The resulting variable is an Indigenous population weighted average of the Murdock's variable on the political complexity of the Indigenous population for each municipality in Mexico. Formally, the index takes the following form:

$$IndInst_i = \sum \frac{IndigPop_{eit=1930}}{IndigPop_{it=1930}} \cdot \mathbf{JH}_e \tag{2.7}$$

In equation 2.7, $IndInst$ is the index of Indigenous institutions in municipality $i$. $IndigPop_{eit=1930}$ represents the total number of Indigenous people of ethnic group $e$ in 1930 in municipality $i$. $IndigPop_{it=1930}$ represents the total number of Indigenous people from municipality $i$ in 1930. $\mathbf{JH}_e$ is the respective degree of political complexity of ethnic group $e$ based on Murdock's Jurisdictional Hierarchy Index. The index ranges from 0 to 2, with lower values denoting municipalities with less complex Indigenous institutions, while higher values represent the opposite. This index captures the political complexity of the Indigenous population within each municipality in Mexico up to 1930.

We use only the index created with data from the 1930 census.[20] This is because while we assume that Indigenous people remain within their historic settlements, we cannot rule out the possibility of Indigenous people moving to avoid road infrastructure. These intentional movements would bias our coefficients (in absolute terms) upwards, as the rejection of road infrastructure may not be due to how effectively Indigenous groups reject the implementation of road infrastructure, but to these movements. By using the different proportions of Indigenous groups from the last census after the Mexican Revolution, we reduce selection into the treatment by possible movements of Indigenous

[19]See: Gennaioli and Rainer (2006); Gennaioli and Rainer (2007); Michalopoulos and Papaioannou (2013); Angeles and Elizalde (2017); and Elizalde (2020a)

[20]We use 1930 because this is the first census with information on the ethnic composition of the different Indigenous populations.

people in subsequent years. However, our results are robust to different forms of the index, such as using the average value between 1930 and 1960, using the index value in each period, or using the total population as the denominator of the index.

We restrict our sample to municipalities with Indigenous populations in 1930, the rationale being that we are interested in capturing the impact of institutional factors on road infrastructure development that are primarily driven by Indigenous populations. Municipalities with no Indigenous population show no value for Indigenous institutions. Figure A.9 shows the spatial distribution of municipalities with Indigenous populations (grey) and non-Indigenous populations (lightest grey) in Mexico.

Using our continuous $IndInst$ index, we construct a binary indicator of pre-colonial political *centralisation*, following Dippel (2014). We label a municipality as politically *centralised* if municipality $i$ has an $IndInst$ index greater than 0, and we label a municipality as politically $fragmented$ if municipality $i$ has an $IndInst$ index equal to 0.[21] Formally, our *centralisation* dummy variable has the following form:

$$centralisation_i = \begin{cases} Centralised\ if\ IndInst > 0 \\ Fragmented\ if\ IndInst = 0 \end{cases} \tag{2.8}$$

### 2.4.3 Construction of covariates

To rule out alternative mechanisms influencing our results, we control for the following confounding factors:[22].

*Population.* We control for population density, as the provision of road infrastructure (and public goods in general) may have been more costly in less densely populated regions. In addition, we control for the proportion of Indigenous population to rule out the mechanism of discriminatory policies towards municipalities with a high proportion of Indigenous population.

*Geography.* A large number of studies have shown that geographical endowment is an important source for the long-term development of a state (Engerman and Sokoloff, 1997; Bruhn and Gallego, 2012; Herbst, 2014). For example, Herbst (2014) argues that the

[21]Gennaioli and Rainer (2007) were the first to empirically measure pre-colonial institutions in the context of African countries. They defined politically centralised societies as those groups with a value of 2 and above according to Murdock's Jurisdictional Hierarchy Index, while those with a value of 0 and 1 were classified as politically fragmented societies. We do not follow this classification and instead use the approach of Dippel (2014). Dippel (2014) focuses on the Native American context, specifying only Indigenous groups with a value between 0and 2 of Murdock's Jurisdictional Hierarchy Index. Since the majority of Indigenous groups in the United States report a value of zero, the Dippel (2014) study classifies as politically centralised those Indigenous groups with a value of 1 and 2 of Murdock's Jurisdictional Hierarchy Index, and those with a value of zero as politically fragmented. Since our data is more similar to Dippel (2014)'s study, we follow his approach. Nevertheless, our main specifications are subject to Gennaioli and Rainer (2007)'s classification of pre-colonial political centralisation

[22]For the detailed description and corresponding sources for all controls, see the online appendix.

complexity of African geography is a major reason for the persistent lack of state centralisation. We add a set of variables that measure some of the geographical characteristics of municipalities. These are: Latitude, longitude, altitude, land area of the municipality and a land suitability index.

*Pueblos de Indios.* We are interested in the impact of Indigenous institutions among the Indigenous population. However, there is evidence that the original location of Indigenous population has lasting long-term impacts (Arteaga, 2018). Therefore, we control for the original location of Indigenous people in Mexico (Tanck de Estrada, 2005).

*Ethnic diversity.* There is ample evidence in the literature that the provision of public goods is associated with ethnic diversity, especially in developing countries (Alesina and La Ferrara, 2005; Miguel and Gugerty, 2005; Montalvo and Reynal-Querol, 2017). To rule out the possibility that ethnic heterogeneity within municipalities may also have influenced variation in road provision in post-revolutionary Mexico, we construct two measures of ethnic diversity according to Alesina et al. (2003) and Montalvo and Reynal-Querol (2005): ethnic fractionalisation and ethnic polarisation of the Indigenous population, respectively.[23]

[23]The measures were calculated for each municipality in Mexico from 1920 to 1960. Since no data are available for the period before 1930, the index for 1920 is extrapolated using the observed data for 1930

TABLE 2.1: Descriptive Statistics

| | All Municipalities | | | Centralised | | | Fragmented | | |
|---|---|---|---|---|---|---|---|---|---|
| Statistic | N | Mean | St. Dev. | N | Mean | St. Dev. | N | Mean | St. Dev. |
| Road share | 6,297 | 0.02 | 0.06 | 3,115 | 0.02 | 0.05 | 3,182 | 0.02 | 0.07 |
| IndInst | 6,297 | 0.39 | 0.55 | 3,115 | 0.78 | 0.56 | 3,182 | 0.00 | 0.00 |
| Population | 6,297 | 9,108 | 19,008 | 3,115 | 6,741 | 12,865 | 3,182 | 11,426 | 23,286 |
| Pop Density | 6,297 | 42.91 | 103.33 | 3,115 | 37.81 | 60.89 | 3,182 | 47.90 | 132.10 |
| Indigenous pop | 6,297 | 1,559 | 2,701 | 3,115 | 2,441 | 3,293 | 3,182 | 697 | 1,523 |
| Share of Indigenous pop | 6,297 | 0.33 | 0.34 | 3,115 | 0.54 | 0.30 | 3,182 | 0.13 | 0.23 |
| Pueblos Indios | 6,297 | 2.41 | 2.83 | 3,115 | 2.56 | 3.14 | 3,182 | 2.25 | 2.48 |
| Area sq km | 6,297 | 716 | 2,260 | 3,115 | 374 | 915.64 | 3,182 | 1,051 | 3,010 |
| Soil suitability | 6,297 | 79.87 | 24.64 | 3,115 | 87.43 | 12.23 | 3,182 | 72.47 | 30.73 |
| Elevation km | 6,297 | 1.35 | 0.88 | 3,115 | 1.22 | 0.88 | 3,182 | 1.48 | 0.87 |
| Ancestral Land | 6,297 | 0.06 | 0.22 | 3,115 | 0.07 | 0.24 | 3,182 | 0.04 | 0.19 |
| Indigenous Polarization | 6,285 | 0.31 | 0.39 | 3,115 | 0.56 | 0.37 | 3,182 | 0.06 | 0.20 |
| Indigenous Fractionalisation | 6,297 | 0.16 | 0.20 | 3,115 | 0.29 | 0.19 | 3,182 | 0.03 | 0.10 |
| Fractionalisation | 6,297 | 0.28 | 0.24 | 3,115 | 0.45 | 0.19 | 3,182 | 0.12 | 0.18 |

*Notes:* This table shows the descriptive statistics of the variables used for the analysis. The panel *All Municipalities* includes all municipalities used in the analysis. The panel *Centralised* includes all municipalities labelled as centralised. The panel *Fragmented* includes all municipalities that are marked as fragmented. *Centralised* municipalities are defined as those with an Indigenous institution index greater than zero, and *Fragmented* municipalities are defined as those with an Indigenous institution index equal to zero.

Table 2.1 shows the descriptive statistics of our main variables and controls. The first three columns show the statistics of our full sample of municipalities. The mean road share between 1920 and 1960 is 0.02 (0.07 sd). Comparing the mean values between municipalities within the centralised and fragmented groups in the following columns, it is clear that municipalities classified as centralised have a lower road share than fragmented municipalities. Furthermore, there are significant mean differences in our set of confounding factors between these two groups of municipalities, demonstrating the importance of controlling for these confounding factors.

### 2.4.4 Empirical strategy

#### 2.4.4.1 Specifications: Difference-in-Differences

To examine which types of Indigenous groups were better able to reject road infrastructure in Mexico, we use the difference-in-differences (DiD) approach embedded in equation 2.9:

$$Road_{it} = \alpha_i + \tau_t + \delta_{s,t} + \beta(IndInst_i * postPR_t) + (X_i' * postPR_t)\gamma + Z_{it}'\rho + \epsilon_{it} \quad (2.9)$$

In equation 2.9, $Road_{it}$ is the road share in municipality $i$ at time $t$. $\alpha_i$, $\tau_t$, and $\delta_{t,s}$ are municipality, time and state-by-time fixed effects, respectively. $IndInst_i$ is our index of Indigenous institutions as defined in equation 2.7, while $postPR$ is the treatment indicator that takes the value of 1 after the start of the first national road construction programme (1925) and 0 otherwise. The coefficient of interest is $\beta$, which captures the differential evolution of road share between municipalities with different levels of $IndInst$, before and after the start of the road programme. $X_i'$ is a vector of time-invariant covariates: *Pueblos de Indios*,latitude, longitude, municipality area, elevation and land suitability. $Z_{it}'$ represents a vector of time-variant covariates that includes: Population density, proportion of Indigenous population and the two indices of ethnic heterogeneity (polarisation and fractionalisation) in municipality $i$ at time $t$. Note that time-invariant controls interact with the treatment indicator $postPR$. In all specifications, standard errors are clustered at the municipality level, and our sample includes only municipalities with Indigenous population in 1930.

We use a second specification in which we use the *centralisation* indicator that classifies municipalities as politically centralised or politically fragmented. Our second diff-in-diff strategy is embedded in equation 2.10, in which $\beta$ captures the differences in road share between politically *centralised* and politically *fragmented* municipalities before and after the start of the road programme. The rest of the model resembles equation 2.9.

$$Road_{it} = \alpha_i + \tau_t + \delta_{t,s} + \beta(centralisation_i * postPR_t) + \Gamma(X_i' * postPR_t) + \rho Z_{it}' + \epsilon_{it} \quad (2.10)$$

The diff-in-diff setting allows us to mitigate potential endogeneity problems that might bias our estimator. First, one might worry that time-invariant factors such as geography or culture affect our results, as they are correlated with both the IndInst index and the road share. To counter this, we control for time-invariant confounders by adding municipality fixed effects. Second, we add time fixed effects to control for time-varying factors that affect all Mexican municipalities equally (e.g. economic shocks). Third, we could still think about time-varying factors that affect municipalities within a given state.[24] For example, the southern state of Guerrero enacted a law in the 1930s that provided for deductions from civil servants' salaries to raise revenue for road infrastructure. This law only affected municipalities within that state. To control for such shocks, we therefore include state-by-time fixed effects. Finally, we interact our various time-invariant covariates with our treatment indicator *postPR*, to control for differences in the evolution of the road share across municipalities with different geographical levels and historical factors before and after the start of the road construction programme.

One might also worry that confounding factors that vary at the time level of municipality $x$ are correlated with both our outcome and the main variable of interest. To address this concern, we include a series of controls for municipality $x$ time. First, we control for population density in municipality $i$ and time $t$ to rule out the possibility of population distribution effects on road infrastructure provision (Herbst, 2014). Second, previous research has documented lower provision of public goods in areas with a higher proportion of Indigenous population (Montenegro and Stephens, 2006; The World Bank, 2015). We control for the proportion of Indigenous population to rule out the hypothesis that the state discriminates against municipalities with a higher proportion of Indigenous population. Finally, there is evidence in the literature that economic outcomes are associated with ethnic diversity, especially in developing countries (Alesina and La Ferrara, 2005; Miguel and Gugerty, 2005; Montalvo and Reynal-Querol, 2017).To rule out that ethnic heterogeneity also influenced road infrastructure in post-revolutionary Mexico, we construct and control for two measures of ethnic diversity according to Alesina et al. (2003); and Montalvo and Reynal-Querol (2005): ethnic fractionalisation and ethnic polarisation, respectively, in municipality $i$ at time $t$.

[24]Mexico has 32 states, which are the largest administrative units below the national level, one administrative level above the municipalities.

#### 2.4.4.2 Main results

Column (1) of table 2.2 shows a pooled OLS regression of road share on Indigenous institutions. The result shows a negative and significant correlation between Indigenous institutions and road share between 1920 and 1960. Figure A.12 illustrates the graphical representation of this correlation. The causal effect of this correlation should be interpreted with caution due to possible endogeneity problems. However, the sign of this estimator suggests that municipalities with a higher share of Indigenous institutions are correlated with lower road infrastructure, which is in line with our main hypothesis.

Columns (2) and (3) of table 2.2 show the results of the difference-in-differences specification from equation 2.9. In column (2), we control for municipality fixed effects, time fixed effects and state-by-time fixed effects. In column (3) we add our set of time-invariant and time-varying controls.

The results of columns (2) and (3) in table 2.2 confirm the main hypothesis of the paper. The interaction term between $IndInst_i$ and *postPR* is negative and statistically different from zero. Furthermore, the coefficients are stable to the inclusion of our set of controls. The results suggest that after the start of the road construction programme, road infrastructure was lower in municipalities with a higher index of Indigenous institutions. The coefficient of interest is -0.014, so a one standard deviation increase in the IndInst index leads to a 0.008 decrease in road share. The effect is of non-negligible magnitude, as the average road share between 1920 and 1960 is 0.02. Therefore, in 1960, a one standard deviation increase in the IndInst index resulted in a 40% lower road share compared to the mean.

In column (4) our indicator *centralisation* is used. The result shows that in post-revolutionary Mexico, road infrastructure was significantly lower in municipalities where a larger proportion of the Indigenous population came from politically centralised societies of the pre-colonial period. This suggests that Indigenous people coming from pre-colonial societies with state characteristics (e.g. *Aztecs* or *Tarascos*) were better able to reject road infrastructure than those coming from politically *"fragmented"* societies (e.g. *Chinantec*, *Cocopa* or *Huichol*).

Our results are particularly interesting in light of the evidence provided by Gennaioli and Rainer (2007). They find that regions with pre-colonial centralised institutions are associated with better provision of public goods in Africa, including road infrastructure. Our results differ from those of this study in one important respect. We find that in regions with *centralised* Indigenous institutions, road infrastructure was lower in Mexico. Nevertheless, this need not be understood as a counter-position to the positive impact of pre-colonial institutions on economic development, but as a relevant complement to the seminal work of Gennaioli and Rainer. As these authors point out for the African case, ethnic groups in regions with a legacy of politically centralised pre-colonial states were better able to organise themselves by developing better accountability mechanisms for local leaders. In our study, we interpret our findings in a similar way: Indigenous groups with *centralised* Indigenous institutions were better able to coordinate collective action to oppose a large infrastructure project. Road infrastructure in Mexico was not seen by indigenous communities as a public “good", but rather as a public “bad" which threatened their ethnic identity and traditions (Waters, 2006; Bess, 2017).

TABLE 2.2: Road share and Indigenous institutions (DiD)

| | Road Share | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| IndInst | $-0.010^{***}$ | | | |
| | (0.002) | | | |
| IndInst x postPR | | $-0.014^{***}$ | $-0.014^{***}$ | |
| | | (0.003) | (0.003) | |
| Centralisation x postPR | | | | $-0.011^{***}$ |
| | | | | (0.004) |
| Municipality FE | | ✓ | ✓ | ✓ |
| Time FE | | ✓ | ✓ | ✓ |
| State x time FE | | ✓ | ✓ | ✓ |
| Time Invariant Controls | | | ✓ | ✓ |
| Time Variant Controls | | | ✓ | ✓ |
| $N$ | 6,297 | 6,297 | 6,297 | 6,297 |
| $R^2$ | 0.010 | 0.046 | 0.069 | 0.067 |
| Mean Road Share | 0.02 | 0.02 | 0.02 | 0.02 |
| SD Road Share | 0.06 | 0.06 | 0.06 | 0.06 |

*Notes:* This table presents the estimates used to examine the impact of Indigenous institutions on road share. The dependent variable in all columns is our measure of road share. Column (1) shows the correlation between Indigenous institutions and road share. In columns (2) and (3), the independent variable is the interaction between the continuous Indigenous institutions index, IndInst, and the post-treatment variable, postPR. Column (2) contains municipality, time and state by time fixed effects. Column (3) contains all fixed effects, time-varying control variables such as population density, proportion of Indigenous population, ethnic fractionalisation and ethnic polarisation, and time-invariant control variables such as the number of *Pueblos Indios*, latitude, longitude, municipality area, altitude and land suitability. In column (4), our independent variable of interest is the interaction term between our binary dummy variable for indigenous institutions and the indicator for time after treatment, postPR. Standard errors are clustered at the municipality level (in parentheses). Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

#### 2.4.4.3 Flexible Difference-in-Differences

In this section, we use an additional specification to test whether differential road infrastructure development was an immediate phenomenon shortly after the road programme started or whether it took a few years to become relevant. Therefore, we estimate a flexible difference-in-differences approach embedded in equation 2.11:

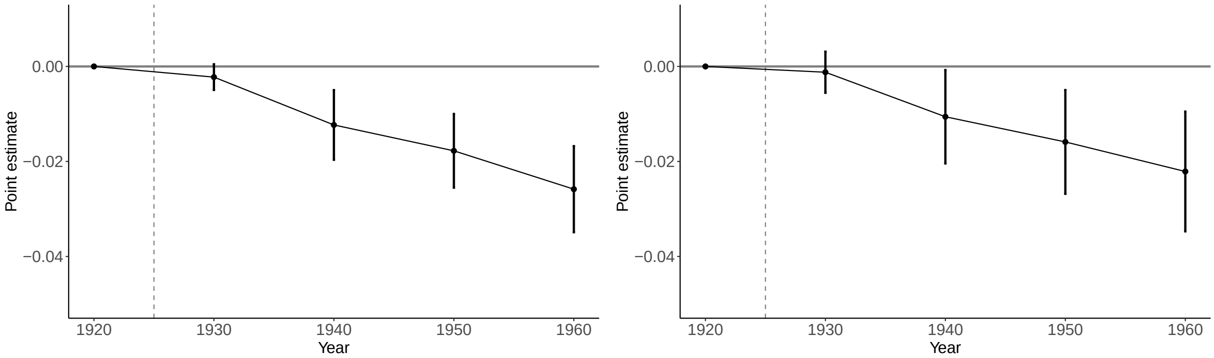
$$Road_{it} = \alpha_i + \tau_t + \delta_{t,s} + \sum_{t=1920}^{1960} \beta_t (IndInst_i * T_t) + \sum_{t=1920}^{1960} \Gamma(X_i' * T_t) + Z_{it}'\rho + \epsilon_{it} \quad (2.11)$$

In equation 2.11 we interact our treatment variable, $IndInst_i$ or $centralisation_i$, with our time indicator $T_t$. With this specification, we can observe whether the road share varies between different levels of Indigenous institutions or between politically fragmented and politically centralised municipalities in each time period, and we compare this to our base year (1920). The rest of the model remains the same as in equation 2.9.

Figure 2.3 shows the coefficient of interest ($\beta_t$) of the flexible difference-in-differences approach from equation 2.11. The vertical line marks the year when road construction started (1925). The error bars show the 95% confidence intervals. Table 2.11 shows the results of this regression.

Subfigure 2.3a shows the coefficient for the interaction term between $IndInst_i$ and $T_t$. The coefficients are negative and increase in absolute value with time. These results show a negative relationship between road share and our IndInst index in each year compared to 1920. Thus, the result shows that differences in road infrastructure between municipalities with different IndInst increased in the post-revolution period, which was arguably characterised as a nation-building period. Moreover, the differential effect (in absolute terms) increases over time and is significant at least at the 5% level in 1940, 1950 and 1960, but not in 1930. Since road construction started in 1925, it may have taken some time for the investment to come to fruition, which is reflected in the zero differences from the 1930 census.[25] Similarly, most of the investment in road infrastructure was made by the state in the 1940s (see figure A.7), which would likely have been documented in the 1950 and 1960 censuses. Comparing the coefficients for 1940, 1950 and 1960, we find that they increase in absolute terms over time, indicating zero convergence in terms of road infrastructure between municipalities with different IndInst index levels.

[25]For example, a section of the famous Panamericana was completed in 1936 after more than a decade of planning and construction (Bess, 2017).The section of this road runs from the northern city of Nuevo Laredo in the state of Tamaulipas to Mexico City.




(A) Indigenous Institutions: IndInst

(B) Fragmented vs Centralised

FIGURE 2.3: Flexible-DiD estimates

*Notes:* This figure shows the point estimates for the interaction between $IndInst_i$ and $T_t$ for the flexible difference-in-differences embedded in equation 2.11. The vertical line marks 1925, the year when road construction began in Mexico. The error bars show the 95% confidence intervals. Subfigure 2.3a shows the coefficients using our continuous measure of Indigenous institutions. Subfigure 2.3b shows the coefficients for comparing the "centralised" and "fragmented" groups. All regressions control for time-varying control variables such as population density, proportion of Indigenous population, ethnic fractionalisation and ethnic polarisation; and time-invariant control variables such as number of *Pueblos Indios*, latitude, longitude, municipality area, altitude and land suitability; municipality, time and state-by-time fixed effects. The standard errors are clustered at the municipality level to construct our confidence intervals
.

Subfigure 2.3b shows the results using our *centralisation* indicator, which also support the main hypothesis of the paper. The figure shows that in municipalities with a larger proportion of Indigenous population coming from politically centralised societies in pre-colonial times, road infrastructure during the nation-building period was significantly lower than municipalities where the majority of the Indigenous population came from a politically fragmented society.

In summary, these results show that during the nation-building period (1920sto 1950s), the state's ability to provide road infrastructure was more limited in municipalities with "centralised" Indigenous institutions. Differences in the proportion of roads between municipalities with different levels of Indigenous institutions are significantly different from zero in the 1940s and 1950s, when road infrastructure was significantly developed. Interestingly, these differences are mainly caused by municipalities with "centralised" Indigenous institutions.

### 2.4.5 Robustness checks

#### 2.4.5.1 Parallel trend assumption: pre-treatment period

The validity of our identification strategy rests on the assumption that road infrastructure in municipalities with different levels of indigenous institutions would have developed in parallel in the absence of the nation-building period. In this subsection we provide evidence to support this assumption.

As mentioned in the Background section, there was no road infrastructure in Mexico until the beginning of the 20th century. Roads (or fully paved highways) did not exist until after 1925, when the first large-scale road building programme was launched by the post-revolutionary governments (Chias, Reséndiz, and García, 2010; Bess, 2016). Before roads existed, Mexico had two types of transport infrastructure: colonial roads (dirt roads) and railways (Hernán, 1994). The colonial roads consisted of a small network of bridleways that were already built during the colonial period (1521-1810). During the colonial period, these roads were essentially used to transport mineral resources - especially silver and gold - to the most important ports in Mexico at the time (e.g. Veracruz), from where they were shipped to the Spanish Empire. The colonial roads in central Mexico connected the most productive agricultural plantations with the large urban areas, while the southern colonial roads connected areas of high Indigenous population density, including some areas in Guatemala (Hernán, 1994).

During the dictatorial regime of Porfirio Diaz (1884-1911) (Bess, 2016), the railways became a symbol of national power. Under Diaz's regime, the railway network was expanded from 893 km to 19,205 km by 1920 (Coatsworth, 1979). The railways became the engine of growth for transporting raw materials and primary products produced on the large plantations in central and northern Mexico to the United States, thus promoting the economic progress of Mexico's northern and central regions (Coatsworth, 1979).

To test whether pre-revolutionary transport infrastructure correlates with our Indigenous institution index, we digitised and georeferenced i) the total length of roads built during the colonial period and ii) the total length of railways up to 1920.[26] We calculate the share of colonial roads and railways up to 1920 at the municipality level and use the following cross-sectional specification:

$$Infrastructure_i = \alpha + \beta IndInst_i + X_i'\Gamma + Z_i'\lambda + \mu_s + \epsilon_i \tag{2.12}$$

$Infrastructure_i$ is either the share of colonial roads up to 1821 or the share of railways up to 1920. $IndInst_i$ is our Indigenous institutions index as defined in equation 2.7. In another specification, we replace $IndInst_i$ with our indicator for *centralisation*. $X'$ represents a vector of geographical characteristics, such as altitude, latitude, longitude, land suitability, municipal area and *Pueblos de Indios*. $Z'$ is a vector of demographic characteristics, such as population density in 1800 if we use colonial roads as the dependent variable; if we use railways as the outcome, we include population density in 1920 and the

[26]Our main sources are the Mexican Ministry of Transport and the Mexican National Archives. More details on the description and sources of these variables can be found in the online appendix.

proportion of indigenous population in 1920. $\mu_s$ represents state fixed effects. We report standard errors that are robust to heteroskedasticity.

Table 2.3: Pre-revolutionary infrastructure

| | Colonial road share | | Railway share 1920 | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| IndInst | −0.005* | | −0.004 | |
| | (0.003) | | (0.003) | |
| centralisation | | −0.001 | | 0.0005 |
| | | (0.004) | | (0.005) |
| State FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| $N$ | 1,166 | 1,166 | 1,147 | 1,147 |
| $R^2$ | 0.070 | 0.068 | 0.269 | 0.268 |
| Mean DV | 0.02 | 0.02 | 0.04 | 0.04 |

*Notes:* This table presents the estimates used to examine the impact of Indigenous institutions and the binary variable *centralisation* on public infrastructure before the revolution. The dependent variable in columns (1) and (2) is the colonial road share. In columns (3) and (4), the dependent variable is the share of railways by 1920. All columns include state fixed effects, geographic controls such as latitude, longitude, municipal area, altitude and land suitability, and demographic controls such as the number of *Pueblos Indios*, and colonial population density. In addition, we add population density in 1920 and the proportion of indigenous population as control variables in columns (3) and (4). Robust standard errors in parentheses. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table 2.3 shows the results of the cross-sectional regression in equation 2.12. In columns (1) and (2), we use colonial roads as the dependent variable. Columns (1)-(2) show that the share of colonial roads is statistically insignificant, especially between municipalities with different levels of Indigenous institutions. In the following two columns (3)-(4) we use the share of railways up to 1920 as the dependent variable. The results are broadly similar to those in the previous columns. We find that the railway network in the pre-revolutionary period is not correlated with either measure of Indigenous institutions.

In light of these findings, we are confident that without the Mexican Revolution and the process of nation-building, the construction of large-scale public transport infrastructures, including roads, would have followed parallel trends between municipalities with different levels of Indigenous institutions.

#### 2.4.5.2 Parallel trend assumption: post-treatment period

In this section we will take further steps to prove the assumption of parallel trend. We use the change in road infrastructure policy in the mid-1950s when the state adopted a modernisation policy vis-à-vis nation building. After the mid-1950s, the road construction programme became more diversified across the Mexican territory to meet private sector demand (Hodges and Gandy, 1983; Bess, 2017). We therefore hypothesise that after the nation-building period, road infrastructure would show parallel trends between municipalities with different Indigenous institutions, as Indigenous people would no longer see roads as a threat to their identity.

The turning point in the politics of public infrastructures is clear from the qualitative data of Hodges and Gandy (1983). They argue that while the state provided fuel oil and dams, railways and roads in the late 1950s, private companies were after the profits. Then, in the late 1980s, road infrastructure was substantially taken over by the private sector through private concessions known as Built, Operate and Transfer. Companies thus increasingly assumed responsibility for road infrastructure, from technical implementation and investment to construction and maintenance. One of the most ambitious road building programmes in the world was carried out in Mexico at the time, when some 3,600 kilometres of roads were built in just one presidential term (6 years), which was about a third of the time it took in Western Europe (Foote, 1997).

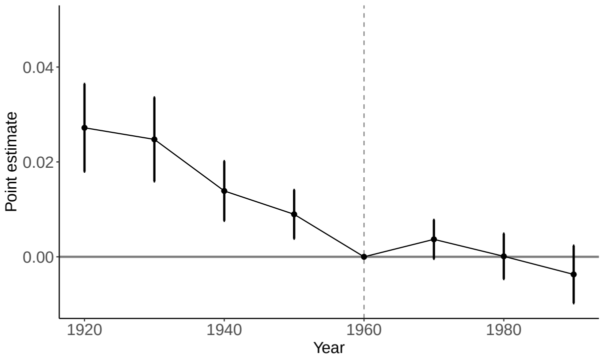
Figure 2.1 shows that after 1960, policies on the provision of road infrastructure changed between municipalities with different levels of Indigenous institutions. In the 1950s, the Mexican state changed its nation-building policy to one of promoting economic growth. Therefore, the 1960 census was designed to capture roads built during the last decade of the nation-building phase. Graphical inspection of figure 2.1 supports our hypothesis that road infrastructure followed parallel trends between centralised and fragmented municipalities, when roads were no longer used as a tool for nation building. To test this more formally, we use the following flexible diff-in-diff specification:

$$Road_{it} = \alpha_i + \tau_t + \delta_{t,s} + \sum_{t=1920}^{1990} \beta_t(IndInst_i * T_t) + \sum_{t=1920}^{1990} \Gamma(X_i' * T_t) + Z_{it}'\rho + \epsilon_{it} \quad (2.13)$$

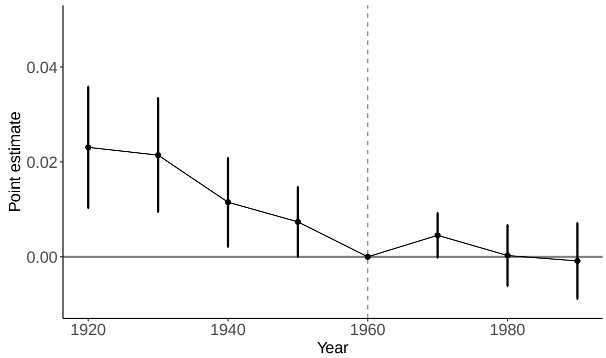
Equation 2.13 follows a flexible difference-in-differences approach, similar to equation 2.11. However, in this specification, our base period is 1960. We use 1960 as the base period because this year captures all roads built in the last decade of the nation-building process in the 1950s. In this specification, the coefficients of interest ($\beta_t$) capture the differential impact of Indigenous institutions on road share in each year compared to our base year, i.e. 1960.

Figure 2.4 shows the coefficient of interest of the flexible difference-in-differences approach from equation 2.13. The error bars indicate the 95% confidence intervals. The vertical line separates the two periods of interest: the nation-building period (1920-1960) and the subsequent period of political change (1960-1990). Figure 2.4 follows a similar structure as figure 2.3. The subfigures 2.4a-2.4b show the results using our continuous

*IndInst* Index and the political *centralization* indicator, respectively. Table A.3 shows the estimates for these specifications. The results support our hypothesis by showing that after the termination of the state's nation-building policy, road development follows parallel trends between centralised and fragmented municipalities.




(A) IndInst Index




(B) Fragmented vs Centralised

FIGURE 2.4: Flexible-DiD estimates for parallel trend

*Notes:* Figure 2.4 shows the point estimates for the interaction between $IndInst_i$ and $T_t$ for the flexible difference-in-differences embedded in equation 2.11. The vertical line marks 1960, the end of the nation-building period. The error bars show the 95% confidence intervals. Subfigure 2.4a shows the coefficients using our continuous measure of Indigenous institutions. Subfigure 2.4b shows the coefficients for comparing the "centralised" and "fragmented" groups. All regressions control for time-varying control variables, such as population density, proportion of Indigenous population, ethnic fractionalisation and ethnic polarisation; and time-invariant control variables, such as number of *Pueblos Indios*, latitude, longitude, municipality area, altitude and land suitability; municipality, time and state-by-time fixed effects. The standard errors are clustered at the municipality level to construct our confidence intervals.

#### 2.4.5.3 Supply Side: discrimination

Our results show that during the nation-building period in Mexico, municipalities with more complex Indigenous institutions received less road infrastructure. We argue that the differences in road infrastructure are due to the fact that Indigenous populations were able to reject the construction of roads, especially during the nation-building period. We interpret this "rejection of public infrastructure" as a lack of demand for road infrastructure. However, one could also argue the opposite. That the state lacks interest in developing road infrastructure in regions with Indigenous populations. In this section, we perform several checks to rule out the possibility that our results are caused by supply-side discrimination.

First, we examine whether there is a lack of road infrastructure by the state in the form of discrimination against municipalities with more Indigenous populations. We address this question by comparing the difference in road provision between municipalities with and without Indigenous populations in 1930, and with different proportions of Indigenous populations. The basic idea is that if the state discriminated against municipalities because of their high proportion of Indigenous population, we should observe a significantly lower development of roads in municipalities with Indigenous population than in municipalities without Indigenous population. Figure A.8 shows the difference in the development of road provision between municipalities with Indigenous population (dashed line) and municipalities without Indigenous population (solid line) in 1930. This figure shows that road infrastructure in municipalities with and without Indigenous population was parallel.

Second, we control for a potential supply of a road infrastructure network. To do so, we built a counterfactual road network with the same length in km as the original road network between 1920 and 1960. This counterfactual network would have been built if the Mexican state had maximised population connectivity between the state capitals and Mexico City. Figure A.11 in the appendix shows maps of the counterfactual road network between 1930 and 1960.[27]

We use the specification in equation 2.9 and include the counterfactual road network as a control variable. The idea behind this is that the supply side biases our main results. We would expect that our variable of interest, the Indigenous institution index, does not affect the evolution of the real road network if the supply side is a confounding variable.

Table 2.4 is a replication of our main results in table 2.2, but includes the counterfactual road network as a control variable. The estimates are smaller in (absolute) size, but still negative and statistically different from zero. These results show that after controlling for potential road infrastructure on the supply side, the demand side through the index of Indigenous institutions influenced the development of road infrastructure in Mexico between 1920 and 1960.

Last, the literature has shown that regions with ethnic higher levels of ethnic diversity have lower levels of public goods provision due to a lack of supply of public goods (Alesina et al., 2003; Alesina and La Ferrara, 2005; Montalvo and Reynal-Querol, 2005). Therefore, to mitigate potential concerns that ethnic diversity rather the indigenous institution index is driving our main results. We control for ethnic fractionalisation, ethnic polarisation and the share of Indigenous population in all our specifications to exclude other mechanisms that could potentially affect our main results.

[27] In appendix A.1 see a description of how we built the counterfactual network.

TABLE 2.4: Indigenous institutions index Vs Counterfactual road network

| | Road Share | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| IndInst | −0.006*** | | | |
| | (0.002) | | | |
| IndInst x postPR | | −0.010*** | −0.010*** | |
| | | (0.003) | (0.002) | |
| Centralisation x postPR | | | | −0.007** |
| | | | | (0.003) |
| Counterfactual road network | 0.308*** | 0.268*** | 0.256*** | 0.257*** |
| | (0.029) | (0.034) | (0.035) | (0.035) |
| Municipality FE | | ✓ | ✓ | ✓ |
| Time FE | | ✓ | ✓ | ✓ |
| State x time FE | | ✓ | ✓ | ✓ |
| Time Invariant Controls | | | ✓ | ✓ |
| Time Variant Controls | | | ✓ | ✓ |
| *N* | 6,297 | 6,297 | 6,297 | 6,297 |
| $R^2$ | 0.206 | 0.150 | 0.162 | 0.160 |
| Mean Road Share | 0.02 | 0.02 | 0.02 | 0.02 |
| SD Road Share | 0.06 | 0.06 | 0.06 | 0.06 |

*Notes:* This table presents the estimates used to examine the impact of Indigenous institutions on road share. The dependent variable in all columns is our measure of road share and we control in all columns for the counterfactual road network. Column (1) shows the correlation between Indigenous institutions and road share. In columns (2) and (3), the independent variable is the interaction between the continuous Indigenous institutions index, IndInst, and the post-treatment variable, postPR. Column (2) contains municipality, time and state by time fixed effects. Column (3) contains all fixed effects, time-varying control variables such as population density, proportion of Indigenous population, ethnic fractionalisation and ethnic polarisation, and time-invariant control variables such as the number of *Pueblos Indios*, latitude, longitude, municipality area, altitude and land suitability. In column (4), our independent variable of interest is the interaction term between our binary dummy variable for indigenous institutions and the indicator for time after treatment, postPR. Standard errors are clustered at the municipality level (in parentheses). Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

### 2.4.6 Rejection of roads: qualitative evidence

To further confirm that Indigenous people in the "centralised" municipalities rejected roads during the nation-building period (1920s to 1950s), we would need to augment our dataset with extensive historical data on Indigenous uprisings. However, to our knowledge, this data is not available. Although this analysis could have supported our findings, we do not believe that this is crucial for the main objective of this paper, which is to investigate which types of Indigenous groups rejected roads.

In any case, as we will show from the evidence of mechanisms in the next section, Indigenous groups in "centralised" municipalities tend to have better capacities for collective action that enable them to persuade the state to meet their demands. This would suggest that road development was lower in "centralised" municipalities because Indigenous groups had greater capacity to coordinate collective action. We argue that Indigenous people demanded less road development because it was used as a nation-building tool to establish the notion of *Mexicanidad*. Road infrastructure was therefore not seen as a public “good" for Indigenous communities, but rather as a public “bad" that threatened their ethnic identity and traditions.

However, in the absence of quantitative data on Indigenous uprisings, we provide qualitative evidence of Indigenous peoples’ rejection of roads. The analysis draws on disputes related to the construction of roads in Mexico between the 1930s and 1950s, when roads were used as a means of nation-building.

An example of Indigenous rejection of roads comes from Los Altos, a region of 17 municipalities in the southern state of Chiapas. In this region, policies implemented since the mid-1930s aimed at integrating the Indigenous population into the process of "Mexicanidad" met with significant resistance from Indigenous communities. Thus, through decentralised state institutions such as the National Indigenous Institute, the state promoted the assimilation of Indigenous culture into the national identity of the *Mexicanidad* through the provision of various public goods, including roads. This led to an Indigenous uprising in the late 1940s, opposing the construction of a road to connect the region to the main cities in the state of Chiapas. In the early 1950s, the Indigenous people even invaded the construction site in large numbers and threatened direct action if the construction of the road continued (Kulher, 1975). It was not until the 1960s that road construction could be resumed.

Other qualitative evidence suggests that opponents of road building were prepared to use violence to prevent road building in their communities. Bess (2017) documents the claims of a supporter of road construction from the southern state of Veracruz in the late 1920s, who expressed that "*...uneducated peasants who opposed [road construction] went so far as to take up arms. These people ... were ready to use violence in their opposition to progress"*. In the 1940s, for example, the Tzotzil in a community in Chiapas opposed road construction because they believed that roads would give access to *ladinos* (of mixed Spanish and Indigenous descent), leading to a deterioration of Indigenous autonomy and a potential threat to their traditions (Waters, 2006).

Opposition manifested itself in clear forms of organisational capacity. Harvey (1998) documented anecdotal evidence from the southern state of Chiapas, where Indigenous communities stopped the construction of a road in the Lancandon region. This construction was intended to demarcate several communities from this region. The organisation made several trips to the affected communities to inform them about the road construction. This led to the mobilisation of hundreds of people who even took on the army. As the movement grew rapidly, negotiations soon took place with the government, which was only able to carry out the technical studies for the construction of the road.

#### 2.4.6.1 Validating hypothesis: bureaucrats

In this section we develop an alternative analysis to further validate the main hypothesis of the paper. An important argument of our hypothesis is that Indigenous people may be more inclined to reject the provision of large-scale infrastructure projects if the state appears to use such projects as a tool for nation-building. This would mean that in the post-revolutionary period, Indigenous people did not see roads as a public good that could bring economic progress to their communities, but rather as a public "bad" that threatened their identity and traditions. Thus, if our hypothesis is correct, we would expect not only a reduced provision of road infrastructure in post-revolutionary Mexico, but also a limited state presence in politically centralised municipalities as a whole.

To explore this question, we examine whether the development of the new state after the Mexican Revolution differed between politically centralised and politically fragmented municipalities. In doing so, we draw on recent work by Garfias (2018) and Acemoglu, García-Jimeno, and Robinson (2015), which use bureaucrats as proxies for state capacity in Mexico and Colombia, respectively.

We therefore collect data on bureaucrats per thousand inhabitants for each municipality in Mexico between 1900-1940, taken from Garfias (2018). Figure A.14 illustrates the unconditional mean of bureaucrats per 1000 inhabitants in 1900, 1930 and 1940 between politically centralised (solid line) and politically fragmented (dashed line) municipalities. We find that the number of bureaucrats around 1900 is greater in fragmented municipalities than in centralised municipalities, suggesting that prior to the Mexican Revolution, in municipalities with Indigenous populations originating from politically centralised societies of the pre-colonial period, Indigenous populations were better able to keep the pre-revolutionary state out of their communities. Moreover, the graph shows that the increase in bureaucrats is significantly greater in fragmented municipalities than in centralised municipalities. Thus, from this graphic inspection, it is clear that the presence of the state in these two groups of municipalities differed greatly during the nation-building process, implying that the new state was able to make greater inroads in the politically fragmented municipalities.

To formally examine the above patterns of state presence, in table A.4 we repeat our basic specification from table 2.2, but use bureaucrats per 1000 inhabitants as the dependent variable. The results from table A.4 support our main hypothesis of this paper. During the nation-building process, state presence, as a proxy for bureaucrats, was significantly lower in municipalities where Indigenous institutions are more centralised.

#### 2.4.6.2 Specification

In this section we show that our results are robust to different specifications. First, we use alternative forms of $IndInst_i$. Table A.5 shows the results. In column (1), we use the total population instead of the total Indigenous population in the denominator of the equation 2.7. In column (2) we use the average $IndInst_i$ between 1930 and 1960. In column (3) we use the time-varying form of our index. The results show that the coefficients (in absolute terms) are slightly larger than our baseline models in column (3) in table 2.2, indicating a potential upward movement of the Indigenous population, as described in section 2.4.

Second, in table A.6 we show that our results are robust to different forms of our dependent variable. In column (1), we use road density instead of road share. To calculate road density, we divide the length of roads in km by the total $km^2$ in each municipality. In columns (2) and (3) we use road shares with buffers of 10 $km^2$ and 500 m respectively. Our main results are not challenged in any way. We are therefore confident that the results are robust to our arbitrary decision to choose a buffer of 2 $km^2$.

Finally, we check whether outliers affect our main results. To do this, we first remove all observations that have a road share greater than 0.4. As figure A.10 shows, our result is heavily skewed to the left. We then transform our dependent variable using its inverse hyperbolic sine form (ihs). The advantage of using the ihs transformation instead of logs is that we can keep the zeros in the data. As table A.7 shows, our results are robust to these two transformations of our main result, suggesting that outliers do not affect our main estimates.

## 2.5 Mechanisms

Our main findings show that Indigenous peoples coming from politically centralised societies of the pre-colonial period were better able to reject road infrastructure in Mexico after the Mexican Revolution. In the following subsections, we propose two possible mechanisms to explain our findings. As highlighted in section 2.3, mechanism (I) assumes that because of the legacy of pre-colonial centralised institutions, better capacities for collective action were developed in municipalities with "centralised" Indigenous institutions. Therefore, Indigenous peoples in politically centralised municipalities coordinate more effectively to get the state to meet their collective demands than in municipalities with "fragmented" Indigenous institutions. The mechanism (II) states that in municipalities with "centralised" Indigenous institutions, Indigenous peoples derive higher benefits from avoiding roads than in municipalities with "fragmented" Indigenous institutions. Therefore, a municipality with "centralised" Indigenous institutions would have a higher preference for maintaining its Indigenous identity and consequently a lower preference for removing its Indigenous identity.

### 2.5.1 Mechanism (I): collective action

In this section we provide evidence that supports mechanism (I). To do so, we follow Elizalde (2020a) and use the redistribution of ancestral land carried out after the Mexican Revolution.

Between 1917 and 1992, the Mexican state implemented an extensive land restitution policy. In the early twentieth century, a significant proportion of the Indigenous population was dispossessed of their ancestral lands. One of the main features of the Mexican Revolution at the beginning of the twentieth century was therefore the restitution of Indigenous land. The post-revolutionary governments therefore enforced a reform of land ownership with Article 27 of the Mexican Constitution of 1917. In a period of about 75 years, the state was able to redistribute 16 million hectares of ancestral land to the Indigenous population.

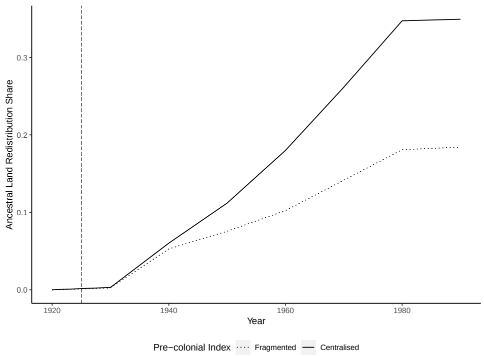
However, the redistribution of ancestral lands was not a simple process and required sustained collective organisation of the Indigenous population. The process of restitution of ancestral lands not only proved administratively cumbersome, but also posed a major challenge, especially in the demarcation of ancestral territories, which was crucial to the extent of land to be transferred (Harvey, 1998). Powerful landowners in rural Mexico opposed such policies, for example by making the demarcation work a lengthy process (Harvey, 1998). Faced with these constraints, Indigenous people had to resort to a wide range of collective methods to make their demands more successful: from land occupations to barricades to organising short-term guerrillas and developing grassroots events to make their demands heard by the state (Elizalde, 2020a).

Figure 2.5a shows the evolution of redistribution of ancestral land in politically "centralised" and politically "fragmented" municipalities between 1920 and 1990.[28] Figure 2.5a highlights several important observations. First, redistribution of ancestral land occurred in both groups of municipalities throughout the reform. Second, ancestral land was redistributed more in municipalities with "centralised" Indigenous institutions than in politically "fragmented" municipalities. These trends show that Indigenous people in municipalities with "centralised" Indigenous institutions were better able to coordinate collective action to achieve (or reject) the provision of collective goods believed to be of greater (or lesser) interest to their own communities.

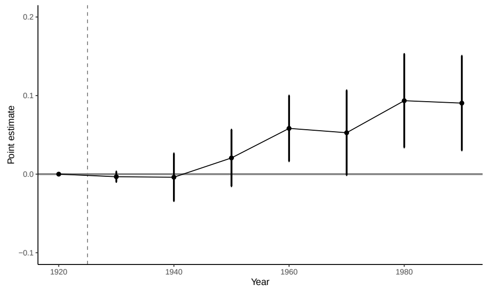
To formally investigate whether there are differences in ancestral land redistribution conditioned by the level of Indigenous institutions, we follow Elizalde (2020a). We estimate a flexible difference-in-differences approach similar to the equation 2.11. However, in this specification, the dependent variable is *Ancestral Land*, which measures the cumulative share of ancestral land redistributed in municipality $i$ at time $t$.

[28] Our measure of ancestral land is the ratio of total $km^2$ redistributed ancestral land in municipality $i$ per total $km^2$ area of municipality $i$

Figure 2.5b shows the estimates when comparing the different evolution of redistribution of ancestral land between municipalities with "fragmented" and "centralised" Indigenous institutions. Consistent with Elizalde (2020a), the estimates show that municipalities with "centralised" Indigenous institutions receive a larger share of ancestral land over time than politically "fragmented" municipalities. This evidence supports mechanism (I).




(A) Ancestral land redistribution




(B) Ancestral land Redistribution: centralised vs fragmented

FIGURE 2.5: Mechanism: collective action

*Notes:* Subfigure 2.5a shows the evolution of redistribution of ancestral land between the two groups of municipalities with different levels of Indigenous institutions. The "centralised" group (solid line) includes municipalities with an index of Indigenous institutions above 0. The "fragmented" group (dashed line) includes municipalities with an index of Indigenous institutions equal to 0. Subfigure 2.5b shows the coefficients comparing redistribution of ancestral land between the groups: "fragmented" vs. "centralised'" following Elizalde (2020a).

### 2.5.2 Mechanism (I): evidence from surveys

In this section, we provide further evidence for mechanism (I) by using contemporary survey data to show that centralised Indigenous groups have a higher propensity for cultural attitudes towards collective action.

We collect individual-level data on collective action preferences from the Latin American Public Opinion Project (LAPOP). As far as we know, this is the only survey that collects information on individuals' ethnicity and preferences for participating in collective action. If our results are due to mechanism (I), individuals with "centralised" indigenous institutions would have a higher preference for collective action than individuals with "fragmented" Indigenous institutions. For the question *In the last 12 months, have you participated in a demonstration or protest march?* , the proportion of people with "centralised" Indigenous institutions answering *yes* should be higher than for people with "fragmented" Indigenous institutions.

We use the survey waves for 2010, 2012 and 2014, where information on Indigenous ethnicity was only collected by LAPOP. Ideally, we would like to use a dataset for the years of our analysis (1920-1960); however, to the best of our knowledge, this information/ dataset does not exist. We are therefore aware that these individual preferences may be the result of years of state indoctrination; the results of this section should therefore be interpreted with caution.

In the survey we find eight different Indigenous groups: Aztecs, Zapotecs, Mixtecs, Mayas and Others. We match each Indigenous group with data from Murdock (1967) and classify them according to their respective degree of pre-colonial political centralisation. Following Dippel (2014), we classify as politically “centralised" those individuals who described themselves as part of an ethnic group with values 1 ( petty chiefdoms) and 2 (larger chiefdoms) of Murdock’s *Jurisdictional Hierarchy beyond the Local Community* variable. Those who self-report belonging to an ethnic group in category zero of Murdock’s index (stateless societies) are classified by us as politically “fragmented". Therefore, Aztecs, Zapotecs, Mixtecs and Mayas are classified as politically ”centralised", while all Indigenous people who described themselves as “other" Indigenous are classified as politically “fragmented".[29] We are then able to compare preferences for collective action between individuals with “fragmented” and “centralised” Indigenous institutions.

It is worth pointing out other disadvantages of the data. We first acknowledge that there may be selection in terms of “being Indigenous", as only people who described themselves as Indigenous stated their ethnic group. In addition, only 138 people reported their ethnicity; therefore, with such a small sample, we could not run a regression analysis as this may have resulted in poorly estimated coefficients with very large standard errors. In view of these shortcomings, we could only perform a mean comparison of the questions we selected.

From the LAPOP surveys, we select five different questions that denote individual commitment to collective action. Table A.9 shows the questions selected for this analysis. The questions are: Q1: (*Prot3*) Have you participated in a demonstration or protest march in the last 12 months? Q2: (*cp5**) In the last 2 weeks, have you tried to help solve a problem in your community or neighbourhood? Q3: (*E5*) Of people taking part in legal demonstrations. How much do you approve or disapprove of this? Q4: (*E14*) Of people seizing private property or land to protest. How much do you approve or disapprove of this? Q5: (*E15*) People taking part in road blockades to protest. How much do you approve or disapprove?.[30]. Note that questions Q1 and Q2 are yes/no questions. Therefore, the mean shows the proportion of people who answered *yes*, broken down by ethnic group. Questions Q3, Q4 and Q5, on the other hand, asked people how much they agreed or disagreed on a scale of 1 to 10. Therefore, the mean value shows how much an ethnic group agrees or disagrees with the respective statement.

Figure A.15 shows the mean and 95% level confidence interval (error bars) of responses

[29]The political complexity of the Zapotecs and Mixtecs was recorded in Murdock’s atlas as being integrated into the Mexican state. These Indigenous groups were among the most politically centralised societies before colonisation. A detailed description of our variables can be found in the appendix A.1.

[30]In brackets we give the original names of the selected questions from LAPOP

for each question by ethnic group, while Table A.10 presents the comparison of means more formally. The results show that individuals who come from politically centralised societies in pre-colonial times are 3.8 and 9.9 percentage points more likely to answer *yes* than people who come from politically fragmented societies in terms of their preferences for participation in demonstrations and community support. Similarly, mean comparisons show that politically centralised individuals have a higher preference for participating in legal demonstrations, seizing private property or land and blocking roads as a form of protest than individuals with a heritage of politically fragmented societies.

This section, taking into account the above drawbacks, provides evidence that individuals with centralised Indigenous institutions are more likely to engage in collective actions than individuals with fragmented Indigenous institutions, which explains why centralised individuals were more likely to reject roads in post-revolutionary Mexico.

### 2.5.3 Mechanism (II): identity

In this section we provide evidence that supports the mechanism (II). This mechanism states that individuals with "centralised" indigenous institutions have a higher utility for preserving their identity and therefore have a higher interest in rejecting road infrastructure than individuals with "fragmented" Indigenous institutions.

To prove the mechanism (II), we collect data on individual attitudes towards ethnic identity from the "National Survey on Discrimination" in Mexico (ENADIS, Spanish acronym). This survey was only conducted in 2005 and 2010. As mentioned above, the use of current survey data should be treated with caution due to possible bias in individuals' responses.

We consider the "Ethnicity module" from the 2010 survey, where we find about 500 respondents who self-report their Indigenous identity and answer our selected questions. Therefore, we construct our indicator on "centralisation" as in section 2.5.2.[31] We select three questions that ask about different aspects of attitudes towards individual ethnic identity and create three indicators as follows. From question *mi3*: "How proud are you of your tradition and heritage?" we create the variable $proud_i$ which takes the value 1 if the respondent answered "A lot", and 0 if the respondent answered "a little" or "nothing". From question *mi2*: "Do you keep some of the customs or traditions of your ancestors?" the variable $custom_i$ is formed, which takes the value 1 if the respondent answered "yes" or "yes, partly" and 0 if the respondent answered "no". Finally, from question *mi4*: "Do you feel...", we create the variable $feel_i$, which takes the value 1 if the respondent answered feeling "...more [*ethnicity i*] than Mexican" or "...as [*ethnicity i*] as Mexican", and 0 if the respondent answered "...more Mexican than [*ethnicity i*]".

[31] ENADIS reported a slightly larger list of Indigenous groups than LAPOP.

Using all three questions, we estimate the impact of belonging to a politically centralised ethnic group compared to a politically fragmented ethnic group on respondents' attitudes towards their Indigenous identity. For this purpose, we use an individual-level cross-sectional regression. In all regressions, we control for demographic characteristics such as age, age$^2$, and gender. In addition, we include municipality fixed effects to account only for individual differences within the municipality. In all our regressions, standard errors are clustered at the municipality level to correct for possible autocorrelation of respondents within the same municipality.

Table A.11 in the appendix shows the results. Column (1) shows that individuals who descend from politically centralised societies in pre-colonial times are more proud of their Indigenous customs and traditions. Individuals descending from "centralised" societies are 10.6 percentage points more likely to say they are proud of their Indigenous tradition and heritage than individuals descending from "fragmented" societies. These results are noteworthy. With 88% of respondents stating that they are proud of their traditions and heritage, this means that (almost) all "centralised" people are proud of their Indigenous traditions and heritage.

Column (2) of table A.11 shows that "centralised" respondents are 16.7 percentage points more likely to preserve their customs and traditions than "fragmented" respondents. Column (3) shows that "centralised" respondents are 10.3 percentage points more likely to feel that their national identity is Indigenous (or as much as Mexican) than "fragmented" respondents.

This section provided survey evidence for mechanism (II). We found that individuals with centralised Indigenous institutions are more proud of their Indigenous identity, more likely to preserve their customs and traditions, and more likely to identify as Indigenous, than individuals with fragmented Indigenous institutions. If identity attitudes are passed on over time and individuals receive better preferences from their ancestors, this would mean that "centralised" individuals would have lower preferences for maintaining road infrastructure, as these roads were arguably used as a tool to create a national sense of *Mexicanidad*. As a result, Indigenous people would have resisted the construction of large-scale infrastructure projects in their communities, especially at a time when roads were a symbol of the adoption of a new national identity.

## 2.6 On the economic consequences of road infrastructure

Numerous studies show that transport infrastructure is an important determinant of economic development.[32] In this section, we provide evidence on the economic consequences of differential road infrastructure development in Mexico. To this end, we regress the

[32]See: Lipscomb, Mobarak, and Barham (2013); Hornung (2015); Baum-Snow et al. (2017); Donaldson (2018);

road share in 1960 on a variety of present-day economic outcomes using the following specification:

$$Y_i = \alpha + \beta Road_{i,1960} + X_i'\Gamma + Z_i'\lambda + \mu_s + \epsilon_i \tag{2.14}$$

In equation 2.14, our dependent variable $Y_i$, is our measure of current economic outcomes in municipality $i$. We consider four outcomes: the proportion of the population with access to electricity at home in 1990, the proportion of the population with access to a sewage system at home in 1990, the literacy rate in 1990, and the satellite light density at night in 2010.

Our main explanatory variable is $Road_{i,1960}$ which measures the total proportion of roads in municipality $i$ in 1960, when the nation-building period was arguably over. $X_i'$ represents a set of control variables that are time-invariant, such as. the municipality's land area, altitude, land suitability, latitude, longitude and number of *Pueblos de Indios*. $Z_i'$ represents socio-economic variables in 1990, such as population density and the proportion of Indigenous population in municipality $i$. $\mu_s$ represents state fixed effects to account only for variation withinthe state. Note that all variables are measured at the municipality level and we use robust standard errors in all specifications.

Table 2.5 shows that road infrastructure up to 1960 is positively correlated with better economic outcomes today. The results show that a higher road share in 1960 is strongly correlated with better access to electricity, a sewage system and higher literacy rates in 1990, as well as with night-time satellite lighting density in 2010. Our main findings in section 4.3.3 show that Indigenous people in municipalities with centralised Indigenous institutions were more likely to reject the implementation of road infrastructure during the nation-building period, leading to worse economic outcomes today. We interpret these findings not against the evidence of the positive impact of pre-colonial institutions on economic outcomes, but as a complementary finding regarding the ability of politically centralised groups to influence policy at the local level.

These findings have important political implications. Indigenous peoples around the world have resisted major infrastructure projects.[33] The impact of large-scale infrastructure projects on the rights of Indigenous peoples is indeed significant. Nevertheless, such projects are important sources of economic growth in the long term.[34] Our results show that in areas where road infrastructure has been less successful, economic outcomes are worse today. We find that an important predictor of road infrastructure development in post-revolutionary Mexico was the legacy of political complexity among Indigenous

[33] See UN Human Rights Council reports A/HRC/39/17/Add.2 for Mexico; A/HRC/39/17/Add.3 for Guatemala; A/HRC/33/42/Add.2 for Honduras; A/HRC/36/46/Add.1 for the United States.

[34] See: Lipscomb, Mobarak, and Barham (2013); Hornung (2015); Baum-Snow et al. (2017); Donaldson (2018)

groups. More recently, these pre-existing institutions have been shown to persist in many parts of the world, particularly in developing countries.[35] Thus, by understanding which types of Indigenous groups are better able to reject large-scale infrastructure projects, governments and developers could design their negotiation strategies to benefit both sides in the implementation of these projects.

TABLE 2.5: Correlation between present outcomes and road share by 1960

| | Electricity 1990 | Drainage 1990 | Literacy rate 1990 | lnlight density 2010 |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Road share 1960 | 0.451*** | 0.495*** | 0.230*** | 2.089*** |
| | (0.040) | (0.067) | (0.022) | (0.144) |
| State FE | ✓ | ✓ | ✓ | ✓ |
| Socio Controls | ✓ | ✓ | ✓ | ✓ |
| Geo Controls | ✓ | ✓ | ✓ | ✓ |
| *N* | 2,085 | 2,070 | 2,086 | 2,085 |
| $R^2$ | 0.365 | 0.514 | 0.645 | 0.514 |

*Notes:* This table shows the estimates of the cross-sectional regressions embedded in equation 2.14. In columns (1)-(4) we use the road share up to 1960 as independent variables. Column (1) uses the share of the population with access to electricity in 1990 as the dependent variable. In column (2), the dependent variable is the share of population with access to sewerage in 1990. In column (3), the dependent variable is literacy rate in 1990. Finally, in column (4), we use nighttime satellite light density in 2010 as the dependent variable. All specifications include state fixed effects as well as demographic and geographic controls. Robust standard errors are in parentheses. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

## 2.7 Conclusion

This paper presents evidence that the low provision of public goods can be partly explained by the lack of demand from Indigenous groups with high preference for Indigenous identity and high coordination capacity. Using the first large-scale road construction programme in post-revolutionary Mexico at the beginning of the 20th century, we show that Indigenous people who descend from politically centralised societies in pre-colonial times were more able to resist the construction of road infrastructure in their municipality.

We find large differences in road infrastructure between municipalities with different levels of Indigenous institutions, before and after the start of the road construction programme. We confirm our identification strategy by showing that pre-revolution public infrastructure does not differ between politically centralised and politically fragmented municipalities: Colonial roads and the railway network up to 1920 do not correlate with our measures of Indigenous institutions. Similarly, we show that after post-revolutionary governments stopped using roads as a means to create a national sense of *Mexicanidad*

[35] textcitegennaioli2007modern; Michalopoulos and Papaioannou (2013); Michalopoulos and Papaioannou (2014); Bandyopadhyay and Green (2016); Angeles and Elizalde (2017); Elizalde (2020a); Heldring (2020).

in the mid-1950s, road development follows parallel trends between politically centralised and politically fragmented municipalities.

We then address two potential mechanisms that explain our results: collective action and higher preferences for preserving Indigenous identity. We test the first mechanism using data from a major ancestral land redistribution reform from the early 20th century. Due to strong resistance from powerful landowners and cumbersome administrative procedures, the success of land restitution required sustained collective organisation of Indigenous people. Consistent with Elizalde (2020a), we show that in municipalities where indigenous people descended from politically centralised societies of the pre-colonial period, a better distribution of ancestral land was achieved. We test the second mechanism by using contemporary survey data on individuals' preferences regarding their Indigenous identity. We show that individuals descended from politically centralised pre-colonial ethnic groups are more proud of their Indigenous customs and traditions, more likely to preserve their Indigenous culture and traditions, and more likely to identify as Indigenous.

Finally, we document the economic consequences of road infrastructure. We show that the rejection of road infrastructure resulted in unintended negative consequences for economic outcomes. In municipalities where road infrastructure was lower, economic outcomes are worse today. These results have important policy implications. Qualitative evidence suggests that Indigenous people are not against large-scale infrastructure projects *per se* but in the imposition of "development" models that undermine their Indigenous identity and traditions within their communities. Understanding which types of Indigenous groups are better able to reject large-scale infrastructure projects may ensure better forms of project implementation, thus facilitating successful projects that are essential for economic growth.

# Appendix

## A.1 Variable description

**Road share**. Our dependent variable was created based on unique collections on the Mexican transport system from the National Library of the Mexican Ministry of Transport. Specifically, we used the following publications: "Volúmenes de tránsito local de 1987. Red carretera troncal. República Mexicana" (SCT, 1987); and "El Transporte en México. Pasado, presente, futuro." (SCT, 1988). These collections contain a series of images depicting the development of the road network in Mexico since the early 20th century, allowing us for the first time to chart its changes over time. Although detailed geospatial data is available on transport systems in Mexico, to our knowledge there has been a lack of comprehensive mapping of the development of the road network. We have achieved this as follows.

The main objective was to signpost each road section at 10-year intervals on a georeferenced base map. This was a very laborious approach which enabled us to increase accuracy. Although the maps from the above publications were produced by transport experts, their drawings were not developed using geographic information systems (GIS). Therefore, digitising the roads would not have accurately represented the exact geographical length and location of all roads built during our study period.

For the signage of the individual road sections in each period, we relied on the visual location and information about the individual road sections contained in the maps. This was possible because all the nodes connecting road sections were usually marked on the images. Since the nodes were labelled with the place names, we were able to accurately label the road sections on the georeferenced base map. Figure A.1 shows the nodes and corresponding place names for some road sections built in 1930 (in red) and 1940 (in orange) in northeastern Mexico. While the images on the left refer to the maps extracted from the above publications, the image on the right shows how these road sections were subsequently drawn into the georeferenced base map. Our georeferenced base map was the 2015 National Road Network (NRN), which consists of a series of shapefiles containing detailed geospatial data on all the different elements of Mexico's infrastructure and transport system, including roads. We used the shapefile for the 2015

NRN as this was the most recent version at the time of data processing. The shapefile for our georeferenced base map can be accessed online via the following link: here.


FIGURE A.1: Example of the cartography of roads and georeferenced map in Mexico by 1930 and 1940.

*Notes:* This figure shows an example of the cartography of roads in Mexico in 1930 and 1940. The maps show the nodes and corresponding names of the localities for some of the road sections that were built in 1930 (in red) and 1940 (in orange) in northeastern Mexico. The two images on the left relate to the cartography prepared as reported in SCT (1988). The image on the right shows how road sections were signposted into the 2015's National Road Network (NRN). Source: SCT (1988) and *Instituto Nacional de Estadistica y Geografia.*

We labelled the road sections manually, individually and per period. We started with the map showing all road sections built until 1930, as shown in figure A.2. Using the visual position and information of the nodes from this map, we were able to draw them into our georeferenced base map. This was done by simply generating an additional attribute for the time period, in this case 1930, into the shapefile of the base map itself. Observations that were not signposted between the printed maps and our georeferenced base map were omitted, so that this shapefile contained all road sections (or more precisely polylines) built up to 1930. We used the same strategy for the remaining periods. In the end, we obtained a number of different shapefiles, each containing the specific universe of roads built in each time period. Figure A.3 shows all the maps used for our main period of analysis (1920-1960) and the corresponding signage results in the georeferenced base map.[36] We then simply merged all the road sections into a single master shapefile that

[36]Please note that there is no map for 1920, as no paved roads (or highways) were built by the state in Mexico in 1920.

contains all the relevant geospatial data, including - and this is important - the period in which they were built.


FIGURE A.2: Cartography or roads in Mexico by 1930

*Notes:* This figure shows the cartography of roads that were built by 1930. The red lines represent the roads, while some of the white dots denote the nodes connected to the road system. The star symbols depicts the capitals of the federal states. Source: SCT (1988) and *Instituto Nacional de Estadistica y Geografia.*

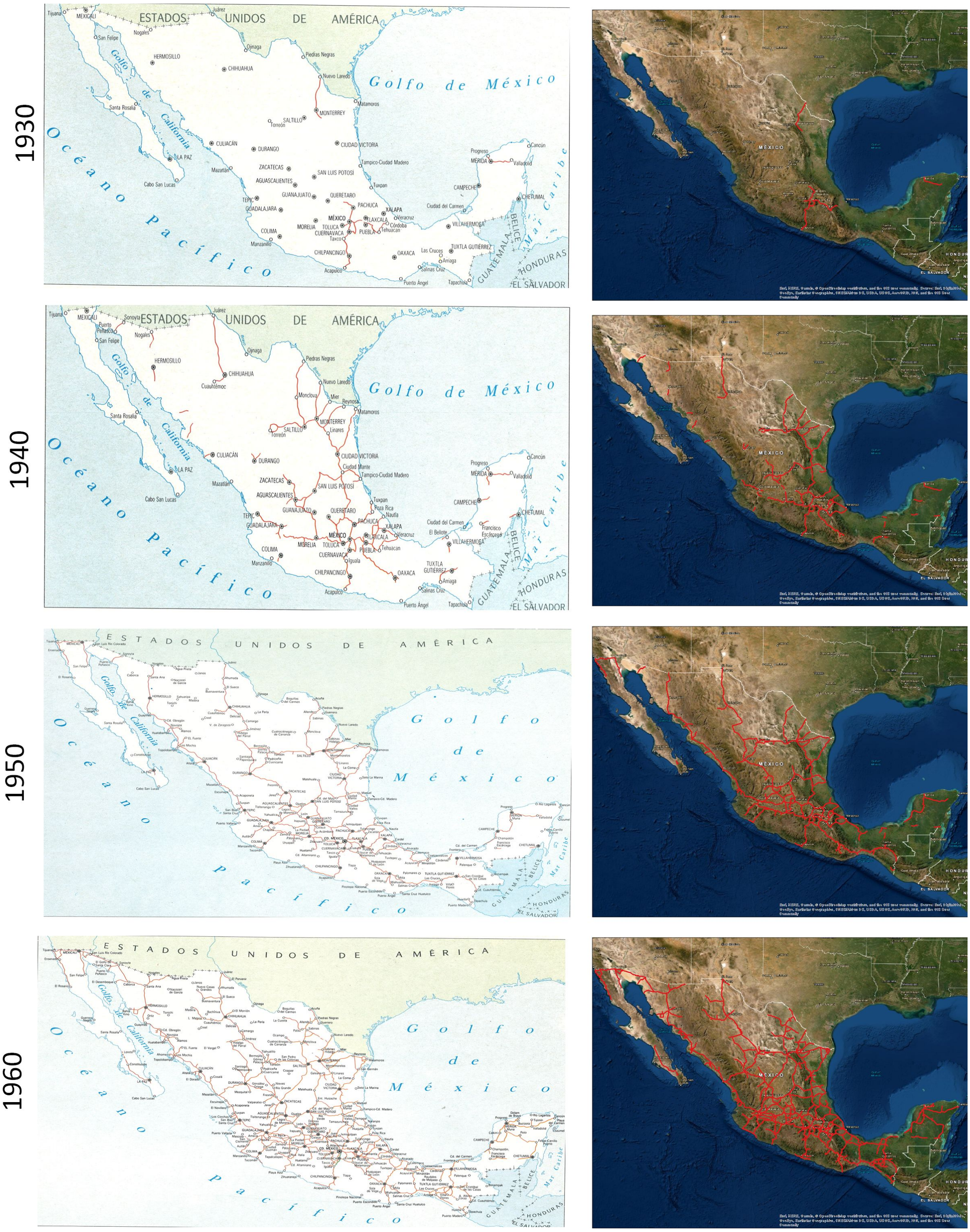



FIGURE A.3: Road network expansion in Mexico between 1930-1960

*Notes:* This figure shows the development of the road network in Mexico between 1930 and 1960. The images on the left show the cartography of roads as reported in SCT (1988). The images on the right illustrate the georeferenced map using the 2015's National Road Network (NRN). Source: SCT (1988) and *Instituto Nacional de Estadistica y Geografia*.

Finally, to obtain our main dependent variable, road share, we followed Dalgaard et al. (2018) and calculated buffers of 1 km along each side of the roads using ArcGIS. Then, using ArcGIS, we simply calculated the area fraction of the buffers within each municipality. Figure A.4 shows a graphical example of how the road share variable is calculated. The figure shows a Mexican municipality crossed by a road (orange line) and the corresponding 2 km buffer.

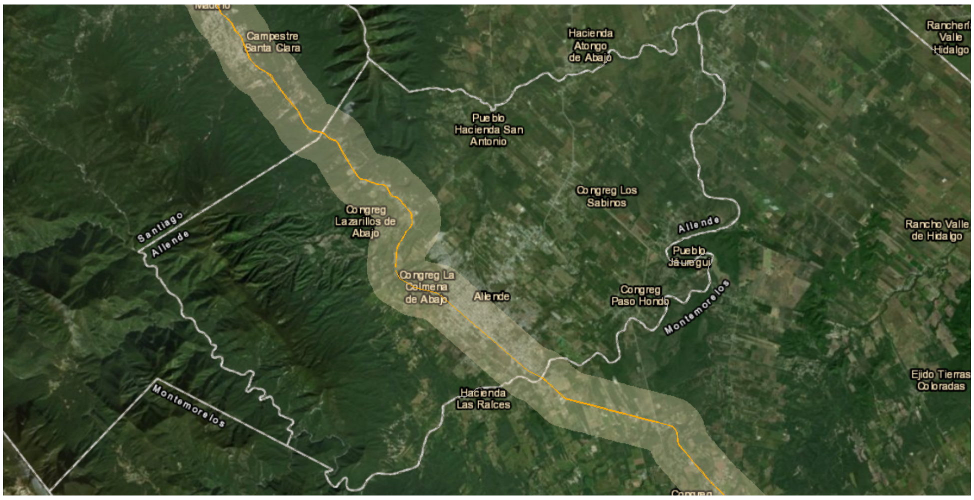



FIGURE A.4: Buffer 2km$^2$

*Notes:* The figure shows a Mexican municipality that is crossed by a road (orange line), we calculated a 2km buffer along the road. To calculate the road share we compute the share of the area of the municipality that is covered by the 2km buffer. We use this procedure to calculate our dependent variable for each road in each Mexican municipality from 1920 to 1990.

**Indigenous institutions index**. We follow Elizalde (2020a) to create an index of indigenous institutions. The index combines anthropological data on the political complexity of Indigenous populations with census data on the different proportions of Indigenous groups in Mexico. To this end, he digitised and collected detailed census data on the proportions of Indigenous groups over the total population by Indigenous language at the municipality level since the 1930 census. The anthropological data came from Murdock's Ethnographic Atlas (Murdock, 1967). In particular, the variable Jurisdictional Hierarchy Beyond the Local Community (v32) was used, which breaks down ethnic groups according to their respective levels of political complexity. The variable varies between 0 and 4, where 0 denotes ethnic groups that have no political authority beyond the local community (stateless societies); 1 represents small chiefdoms; 2 denotes larger chiefdoms or small states; and finally, 3 and 4 represent states and larger states.

In using Murdock's data, the first step was to match the groups between the anthropological and census data. According to the published version of the Atlas (Murdock, 1967), 19 ethnic groups had some information on v32, using the censuses from 1930 to 1960. It is important to note that a total of 15 groups gave an exact level of jurisdiction hierarchy. 3 groups were instead classified by Murdock himself as part of the Mexican or Guatemalan state when referring to their respective level of the jurisdictional hierarchy without giving a specific value. These groups are: Mam (Guatemalan state), Mixtec (Mexican state) and Zapotec (Mexican state). Similarly, another group (Mazatec) has been identified as a “mutual” group without specifying a precise jurisdictional hierarchy level: Chinantec-Mazatec. The jurisdictional hierarchy levels for these groups were then supplemented with ethnographic sources, particularly for Mixtecs, Zapotecs and Mams. In particular, it was documented that the Mixtecs and Zapotecs were organised into large kingdoms before contact with the Europeans, while the Mams belonged to a group that had characteristics of a small chiefdom (Adams and Macleod, 2000). The Mixtecs and Zapotecs were therefore assigned a level two jurisdiction hierarchy and the Mams level one. The Mazatecs were simply given the same jurisdictional hierarchy level (zero) as their corresponding “mutual” group (Chinantec).

Finally, groups reported in the census data but not found in the Murdock Atlas were classified as "Other". The total populations belonging to these groups were assigned a jurisdiction hierarchy level of zero. This strategy was used because all major groups that were categorised (Aztec, Tarasco, Totonac, Mixe and Maya) or partially categorised (Mixtec, Zapotec and Mams) could be found between Murdock’s Atlas and the census data.

Our preferred form of the index is based on Murdock’s categorisation of the political complexity of the Indigenous groups as given in column 1 of the table A.1. Nevertheless, following Angeles and Elizalde (2017) and Elizalde (2020a) in table A.8, we tested the robustness of our results by applying a more conservative categorisation to the groups reported in Panel B of table A.1 that Murdock classified as part of the Mexican or Guatemalan state. Specifically, we downgraded the Mixtecs, Zapotecs and Mams one level in their jurisdictional hierarchy. Thus, the Mixtecs and Zapotecs were downgraded from level two (larger chiefdoms or smaller states) to level one (small chiefdoms), while the Mams were downgraded from level one (small chiefdoms) to level zero (stateless societies) (see categorisation in column 2 of the table A.1). Second, these three groups were assigned the average level of the jurisdictional hierarchy from the 15 groups in Panel A of table A.1, which have a full categorisation based on Murdock’s Atlas (see categorisation in column 3 of table A.1). Finally, we excluded the entire population corresponding to the Indigenous groups categorised as ’other’ and whose category was assigned level zero in the jurisdictional hierarchy. We excluded this population because their groups were not found in Murdock’s Atlas and in the census data.

Table A.8 shows the results using the conservative approach in Murdock’s classification to the jurisdictional hierarchy, as given in table A.1. Column (1) shows the results using $IndInst$, with the groups with a partial classification downgraded by one level in their jurisdiction hierarchy as described above. Column (2) shows the results using the

TABLE A.1: Categorisation for Jurisdictional Hierarchy by Indigenous groups

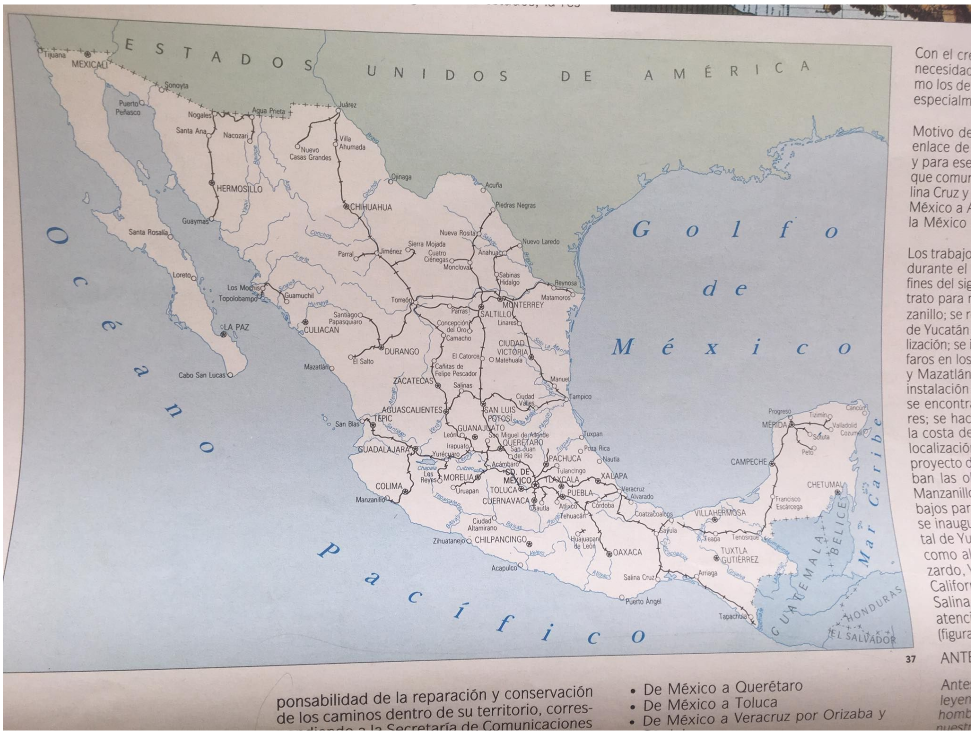
| Indigenous group | (1) | (2) | (3) |
|---|---|---|---|
| *Panel A: Complete categorisation on Jurisdictional Hierarchy* | | | |
| Aztec | Level two | Level two | Level two |
| Chichimeca | Level zero | Level zero | Level zero |
| Chinantec | Level zero | Level zero | Level zero |
| Cocopa | Level zero | Level zero | Level zero |
| Huichol | Level zero | Level zero | Level zero |
| Mixe | Level one | Level one | Level one |
| Papago | Level zero | Level zero | Level zero |
| Pima | Level zero | Level zero | Level zero |
| Popoluca | Level zero | Level zero | Level zero |
| Seri | Level zero | Level zero | Level zero |
| Tarahumara | Level zero | Level zero | Level zero |
| Tarasco | Level two | Level two | Level two |
| Totonac | Level one | Level one | Level one |
| Yaqui | Level zero | Level zero | Level zero |
| Maya | Level one | Level one | Level one |
| *Panel B: Partial categorisation on Jurisdictional Hierarchy* | | | |
| Mam | Level one | Level zero | Average level |
| Mazatec | Level zero | Level zero | Level zero |
| Mixtec | Level two | Level one | Average level |
| Zapotec | Level two | Level one | Average level |

*Notes:* This table shows all the Indigenous groups with some Jurisdictional Hierarchy information (v32), as reported by Murdock's Atlas (Murdock, 1967). Column (1) provides the categorisation that was used to compute IndInst. Columns (2)-(3) show the categorisation applied for the groups in which partial information on Jurisdictional Hierarchy was provided by Murdock's Atlas. Specifically, in column (2), Mam, Mixtec and Zapotec were downgraded to one level in their Jurisdictional Hierarchy, whereas in column (3), these same 3 groups were assigned the average Jurisdictional Hierarchy from the groups in Panel A that reported a complete categorisation.

Indigenous Institutions measure, where the same groups with a partial categorisation were assigned to the average level of the jurisdictional hierarchy of the 15 groups that reported a full categorisation (see Panel A of table A.1). Finally, column (3) shows the results using $IndInst$ excluding the Indigenous population corresponding to the groups classified as "other". The results in table A.8 show that our baseline specifications are robust to these different categorisations.

**Railways**. To construct a variable on railways, we used a similar strategy to that used to construct our main dependent variable. We therefore signposted the railways using the publications SCT (1988) and "Carta de Comunicaciones de 1895-1896 (Von Glümer, 1896), which together contain maps showing the entire extension of the railways until 1920 (see figures A.5a and A.5b). For our georeferenced base map we used another source, which comes from the National Statistics Office INEGI: "Conjunto de datos vectoriales de la serie topográfica".


(A) Railways in Mexico in the 1890s




(B) Railways in Mexico by 1911

FIGURE A.5: Cartography of the railway network in Mexico by the 1910s

*Notes:* Subfigure A.5a shows the cartography of the railway network in Mexico by 1895-96, while subfigure A.5b shows the map of the network by 1911. Sources: SCT (1988) and Von Glümer (1896)

**Colonial roads**. Unlike today's roads and railways, colonial roads were digitised. For this purpose, the map originally created by Alexandre Von Humboldt and published by Arrowsmith in 1810 (see Figure A.6) was used, which originated from the National Archives of Mexico (Arrowsmith, 1810). The georeferencing of the roads was done by drawing polylines across the intersections of all the places the roads crossed.


FIGURE A.6: Cartography of colonial roads in Mexico

*Notes:* This figure shows the main towns and roads of Mexico during the colonial period. Specifically, the map shows the main roads before the Mexican Independence in 1810. Colonial roads are represented in thin dark continuous lines. Source: Arrowsmith (1810)

**Ancestral land (%)**.Ancestral land is the total ancestral land redistributed divided by the land area of municipalities between 1930 and 1990. Source: *Padron e Historial de Nucleos Agrarios* (PHINA), *Registro Agrario Nacional* (RAN).

**Population density** was calculated by dividing the total population at each period by the municipality land area in square km. Source: *Instituto Nacional de Estadistica y Geografia*

**Share of Indigenous population**.The share of Indigenous population measures the percentage of the Indigenous population with respect to the total population at each period. Source: *Instituto Nacional de Estadistica y Geografia*

**Latitude**. Absolute latitude for each municipality. Source: Elizalde (2020a).

**Elevation**. Average elevation (km) for each municipality. Source: Elizalde (2020a)

**Municipalitiy area**. Total surface land area (sq. km.) for each municipality.Source: Elizalde (2020a).

**Suitability**.Index measuring land suitability in each municipality. The index takes values between 0 and 1, capturing the probability of the land area that is served for cultivation. Higher values denote better. Source: Elizalde (2020a)

**Pueblos de Indios**. Total number of Pueblos within each municipality as reported in Atlas Ilustrado de los Pueblos de Indios. Source: *Atlas Ilustrado de los Pueblos de Indios* (Tanck de Estrada, 2005).

**Ethnolinguistic fractionalisation**. The index measures the probability that two randomly selected persons from a given municipality do not belong to the same ethnolinguistic group, according to Alesina et al. (2003). Two different indices were created. One index considered all Indigenous groups and the non-Indigenous group, and the other considered only Indigenous groups. Source: *Instituto Nacional de Estadistica y Geografia*

**Polarisation index**. This index measures how widely individuals in a population are distributed among different ethnic groups, following Montalvo and Reynal-Querol (2005). The index predicts the prevalence of inter-group disputes. Source: *Instituto Nacional de Estadistica y Geografia*

## A.2 Figures

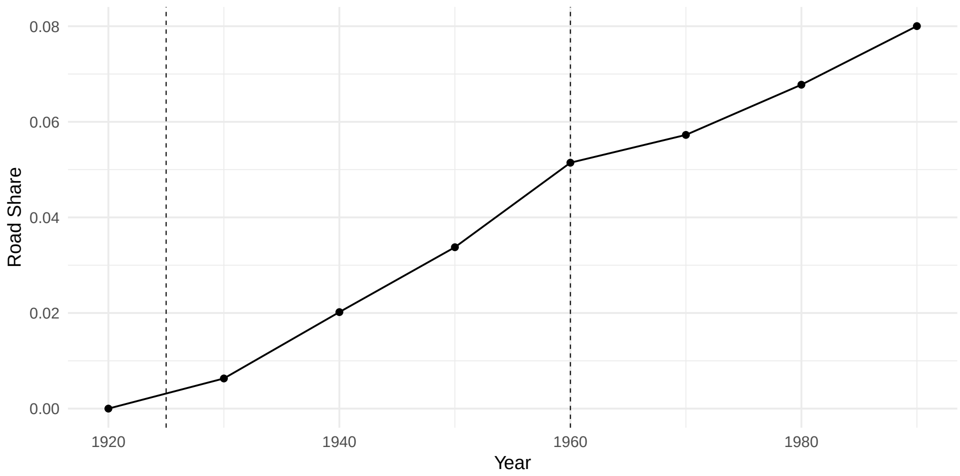



Figure A.7: Road share in Mexico, 1920-1990

*Notes:* This figure shows road share evolution between 1920 and 1990

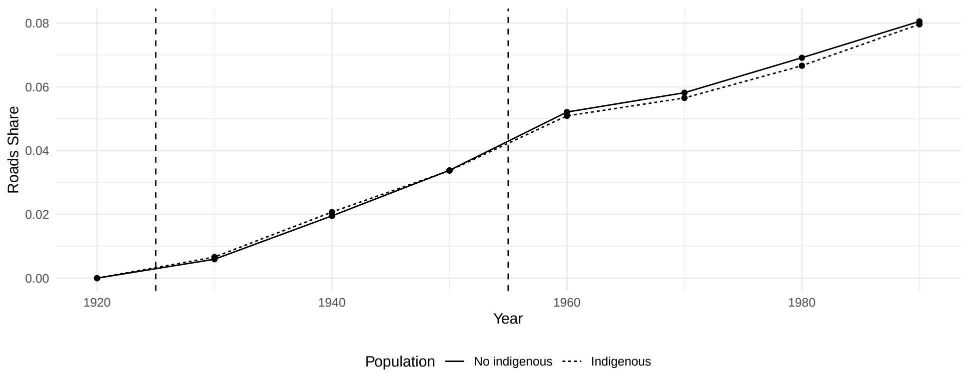



Figure A.8: Road share: Indigenous vs Non-Indigenous

*Notes:* This figure shows the development of road share from 1920 to 1990 in municipalities with Indigenous and non-Indigenous population in 1930.


Figure A.9: Spatial distribution of Indigenous Institutions Index

*Notes:* This figure shows the spatial distribution of our *IndInst* index in Mexico. Darker colors represent a higher Indigenous institution index at the municipality level.

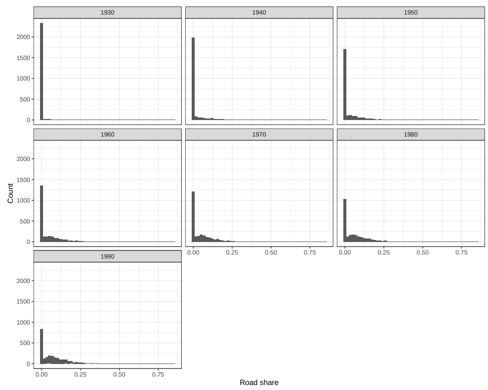



Figure A.10: Histogram on road share by year

*Notes:* This figure shows the histogram of road share in each time period from 1920 to 1990

(A) 1930

(B) 1940

(C) 1950

(D) 1960

Figure A.11: Counterfactual Road infrastructure in Mexico between 1930-1960.

*Notes:* This figure shows the development of counterfactual road infrastructure in Mexico between 1930 and 1960.

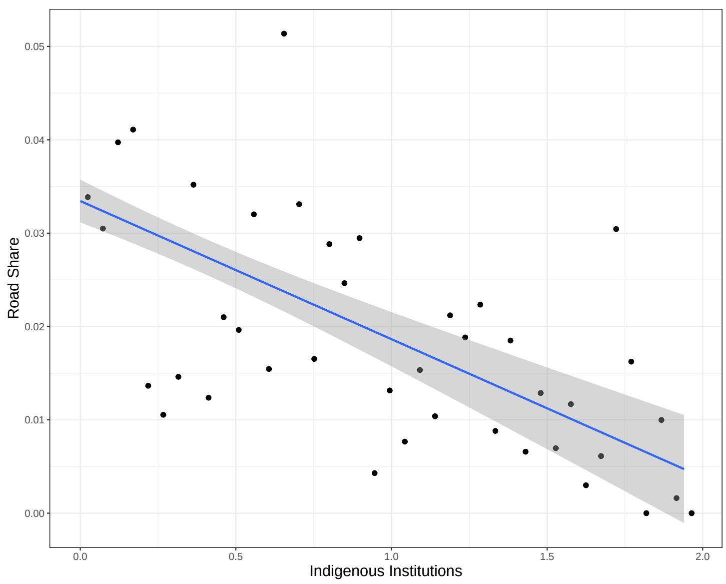



Figure A.12: Road Share and Indigenous institutions

*Notes:* This figure shows the correlation between Indigenous institutions and road share during the nation-building process in Mexico (1920-1960).

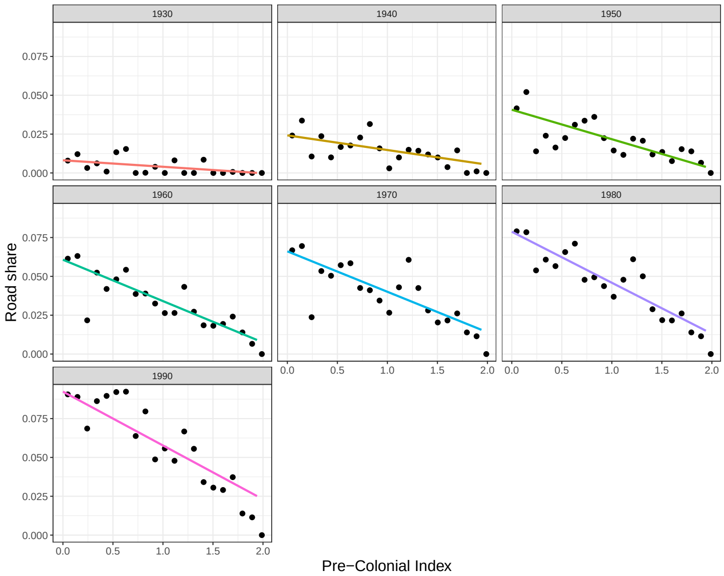



FIGURE A.13: Road Share and Indigenous institutions by year

*Notes:* This figure shows the correlation between Indigenous Institutions and road share by year from 1920 to 1990.

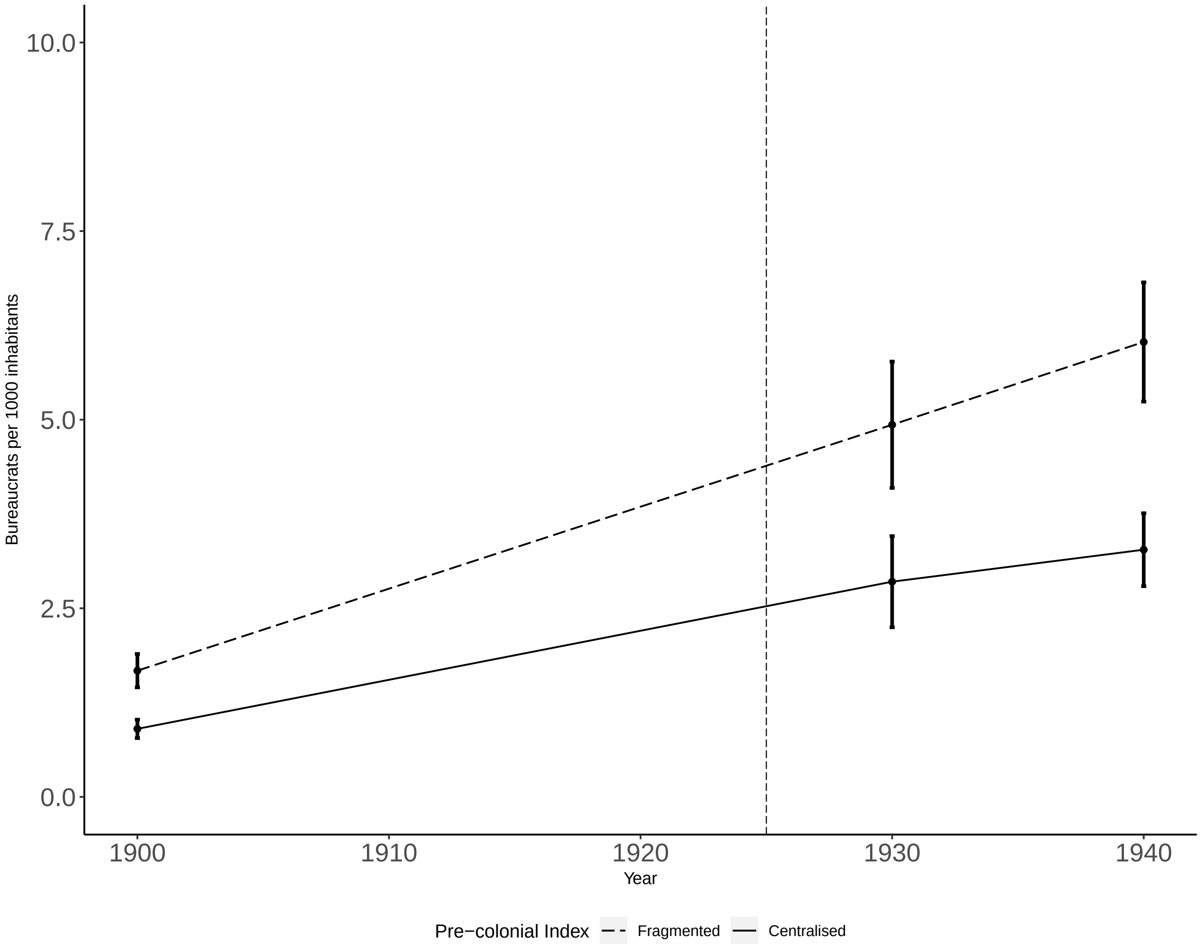
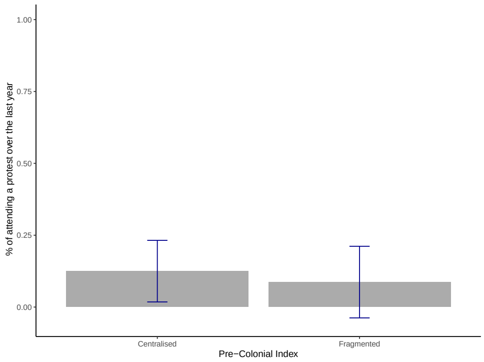
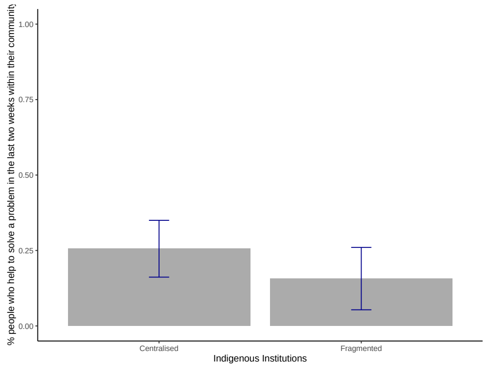



Figure A.14: Bureaucrats per 1000 inhabitants: fragmented vs centralised

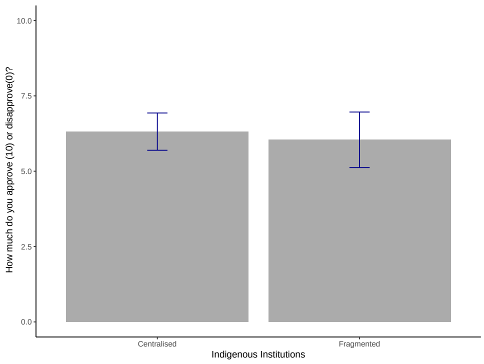
*Notes:* This figure shows the unconditional mean of bureaucrats per thousand inhabitants (from:Garfias, 2018) in politically centralised and politically fragmented municipalities from 1990 until 1940. The error bars show the confidence intervals at the 95% level.



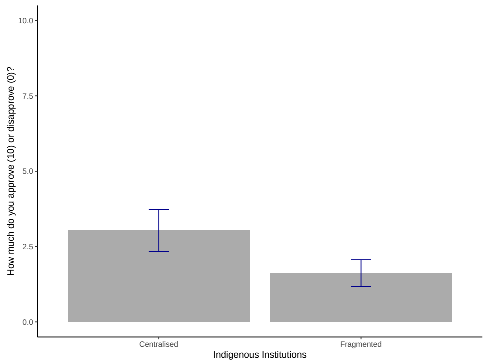

(A) In the last 12 months, have you participated in a demonstration or protest march?



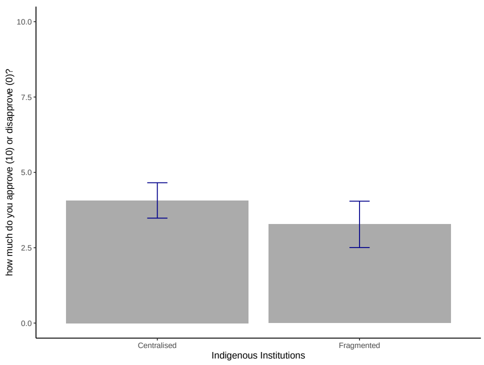

(B) In the last 2 weeks have you tried to help to solve a problem in your community or your neighbourhood?




(C) Of people participating in legal demonstrations. How much do you approve or disapprove?




(D) Of people participating in an organization or group to try to solve community problems. How much do you approve or disapprove?




(E) Of people seizing private property or land in order to protest. How much do you approve or disapprove?

FIGURE A.15: Survey evidence: collective action

*Notes:* These figures show the mean comparison between centralised and fragmented individuals regarding their attitudes towards collective action. The bars show the mean with the 95% confidence intervals. Subfigure A.15a shows the answer to question Q1: (*Prot3*) In the last 12 months, have you participated in a demonstration or protest march?. Subfigure A.15b shows the answer to question Q2: (*cp5**) In the last 2 weeks have you tried to help to solve a problem in your community or your neighbourhood?. Subfigure A.15c shows the answer to question Q3: (*E5*) Of people participating in legal demonstrations. How much do you approve or disapprove?. Subfigure A.15d shows the answer to question Q4: (*E14*) Of people seizing private property or land in order to protest. How much do you approve or disapprove?. Subfigure A.15e shows the answer to question Q5: (*E15*) Of people participating in the blocking of roads to protest. Using the same scale, how much do you approve or disapprove?. Questions Q1 and Q2 are Yes/No questions. Therefore, the mean shows the share of individuals who answered *yes*, by ethnic group. Whereas questions Q3, Q4, and Q5 asked individuals how much they disapprove or approve in a scale from 1 to 10. Hence, the mean shows how much an ethnic group agrees or disagrees with each statement. Survey data corresponds to waves 2010, 2012, 2014 drawn from the Latin American Public Opinion Project (LAPOP).

Table A.2: Flexible DiD

| | Road Share | |
|---|---|---|
| | (1) | (2) |
| IndInst x 1930 | −0.002 | |
| | (0.001) | |
| IndInst x 1940 | −0.012*** | |
| | (0.004) | |
| IndInst x 1950 | −0.018*** | |
| | (0.004) | |
| IndInst x 1960 | −0.026*** | |
| | (0.005) | |
| Centralisation x 1930 | | −0.001 |
| | | (0.002) |
| Centralisation x 1940 | | −0.011** |
| | | (0.005) |
| Centralisation x 1950 | | −0.016*** |
| | | (0.006) |
| Centralisation x 1960 | | −0.022*** |
| | | (0.006) |
| Municipality FE | ✓ | ✓ |
| Time FE | ✓ | ✓ |
| State x time FE | ✓ | ✓ |
| Time Invariant Controls | ✓ | ✓ |
| Time Variant Controls | ✓ | ✓ |
| *N* | 6,297 | 6,297 |
| $R^2$ | 0.083 | 0.076 |

*Notes:* This table shows the coefficients of the flexible difference-in-differences approach embedded in equation 2.11. Further, the table shows the coefficients plotted in figure 2.3. The dependent variable is road share. In column (1), we use a continuous measure of Indigenous institutions. In column (2), we use our binary dummy discrete variable on Indigenous institutions: *centralisation*. All regressions include time-variant control variables, such as population density, share of Indigenous population, ethnic fractionalisation and ethnic polarisation; and time-invariant control variables, such as the number of *Pueblos Indios*, latitude, longitude, municipality area, elevation, and soil suitability; and municipality, time and state-by-time fixed effects. Standard errors are clustered at the municipality level are in parentheses. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

TABLE A.3: Parallel trend: pre-treatment period

| | Road Share | |
|---|---|---|
| | (1) | (2) |
| IndInst x 1920 | 0.027*** | |
| | (0.005) | |
| IndInst x 1930 | 0.025*** | |
| | (0.005) | |
| IndInst x 1940 | 0.014*** | |
| | (0.003) | |
| IndInst x 1950 | 0.009*** | |
| | (0.003) | |
| IndInst x 1970 | 0.004* | |
| | (0.002) | |
| IndInst x 1980 | 0.0001 | |
| | (0.002) | |
| IndInst x 1990 | −0.004 | |
| | (0.003) | |
| Centralisation x 1920 | | 0.023*** |
| | | (0.007) |
| Centralisation x 1930 | | 0.021*** |
| | | (0.006) |
| Centralisation x 1940 | | 0.012** |
| | | (0.005) |
| Centralisation x 1950 | | 0.007** |
| | | (0.004) |
| Centralisation x 1970 | | 0.005* |
| | | (0.002) |
| Centralisation x 1980 | | 0.0003 |
| | | (0.003) |
| Centralisation x 1990 | | −0.001 |
| | | (0.004) |
| Municipality FE | ✓ | ✓ |
| Time FE | ✓ | ✓ |
| State x time FE | ✓ | ✓ |
| Time Invariant Controls | ✓ | ✓ |
| Time Variant Controls | ✓ | ✓ |
| *N* | 10,258 | 10,258 |
| $R^2$ | 0.082 | 0.076 |

*Notes:* This table shows the coefficients of the flexible difference-in-differences approach embedded in equation 2.13. Further, the table shows the coefficients plotted in figure 2.4. The dependent variable is road share. In column (1), we use a continuous measure of Indigenous institutions. In column (2), we use our binary dummy discrete variable on Indigenous institutions: *centralisation*. All regressions include time-variant control variables, such as population density, share of Indigenous population, ethnic fractionalisation and ethnic polarisation; time-invariant control variables, such as the number of *Pueblos Indios*, latitude, longitude, municipality area, elevation, and soil suitability; and municipality, time and state-by-time fixed effects. Standard errors are clustered at the municipality level are in parentheses. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

## A.3 Tables

TABLE A.4: Bureaucrats and Indigenous institutions (DiD)

| | Bureaucrats per 1000 inhabitants | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| IndInst | $-2.147^{***}$ | | | |
| | (0.224) | | | |
| IndInst x postPR | | $-0.986^{***}$ | $-1.437^{***}$ | |
| | | (0.324) | (0.357) | |
| Centralisation x postPR | | | | $-1.342^{***}$ |
| | | | | (0.465) |
| Municipality FE | | ✓ | ✓ | ✓ |
| Time FE | | ✓ | ✓ | ✓ |
| State x time FE | | ✓ | ✓ | ✓ |
| Time Invariant Controls | | | ✓ | ✓ |
| *N* | 2,758 | 2,758 | 2,758 | 2,758 |
| $R^2$ | 0.020 | 0.124 | 0.142 | 0.141 |

*Notes:* This table presents the estimates examining the effects of Indigenous institutions on Bureaucrats per 1000 inhabitants from Garfias (2018). Column (1) shows the correlation between Indigenous institutions and Bureaucrats per 1000 inhabitants. In columns (2) and (3) the independent variable is the interaction between the continuous Indigenous institutions index and the post treatment variable, postPR. Column (2) includes municipality, time and state-by-time fixed effects. Column (3) includes all fixed effects and time-invariant control variables. In column (4), our independent variable of interest is the interaction term between our binary dummy variable of Indigenous institutions and the post treatment indicator, postPR. Standard errors are clustered at the municipality level are in parentheses. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table A.5: Different specification: IndInst

| | Road Share | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| IndInst all pop x postPR | $-0.022^{***}$ | | |
| | (0.003) | | |
| IndInst avg X postPR | | $-0.018^{***}$ | |
| | | (0.004) | |
| IndInst time-variant x postPR | | | $-0.013^{***}$ |
| | | | (0.003) |
| Municipality FE | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ |
| State x time FE | ✓ | ✓ | ✓ |
| Time Invariant Controls | ✓ | ✓ | ✓ |
| Time Variant Controls | ✓ | ✓ | ✓ |
| $N$ | 6,297 | 6,297 | 6,281 |
| $R^2$ | 0.072 | 0.070 | 0.070 |

*Notes:* This table presents the estimates examining the effects of different Indigenous institutions index on road share. In column (1), we use as denominator of equation 2.7, the total population in municipality $i$. In Column (2), we use the average $IndInst$ between 1930 and 1960. Finally, in column (3), we use the time-variant $IndInst_{it}$ index. All columns include municipality, time and state-by-time fixed effects and time-variant and time-invariant controls. Standard errors are clustered at the municipality levels in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table A.6: Different specification: dependent variable

| | Road density | Buffer 10km$^2$ | Buffer 0.5 km$^2$ |
|---|---|---|---|
| | (1) | (2) | (3) |
| IndInst x postRP | $-0.013^{***}$ | $-0.044^{***}$ | $-0.004^{***}$ |
| | (0.005) | (0.009) | (0.001) |
| Municipality FE | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ |
| State x time FE | ✓ | ✓ | ✓ |
| Time Invariant Controls | ✓ | ✓ | ✓ |
| Time Variant Controls | ✓ | ✓ | ✓ |
| $N$ | 6,297 | 6,297 | 6,297 |
| $R^2$ | 0.048 | 0.116 | 0.056 |

*Notes:* This table presents the estimates examining the effects of $IndInst$ on different specification of our dependent variable. In column (1), the dependent variable is road density. In column (2) and (3), the road share buffer is 10 km$^2$ and 0.5 km$^2$, respectively. All columns include municipality, time and state-by-time fixed effects and time-variant and time-invariant controls. Standard errors are clustered at the municipality levels in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

TABLE A.7: Outliers

| | Road share < 0.4 | ihs Road share |
|---|---|---|
| | (1) | (2) |
| IndInst x postPR | −0.009*** | −0.013*** |
| | (0.002) | (0.003) |
| Municipality FE | ✓ | ✓ |
| Time FE | ✓ | ✓ |
| State x time FE | ✓ | ✓ |
| Time Invariant Controls | ✓ | ✓ |
| Time Variant Controls | ✓ | ✓ |
| *N* | 6,282 | 6,297 |
| $R^2$ | 0.081 | 0.071 |

*Notes:* This table presents the estimates examining the effects of *IndInst* on road share. In column (1), we remove all municipalities with a road share larger than 0.4. In column (2), we use the inverse transformation sine of our dependent variable. All columns include municipality, time and state-by-time fixed effects and time-variant and time-invariant controls. Standard errors are clustered at the municipality levels in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table A.8: Specifications using different forms of IndInst

| | Road share | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| IndInst dropping JH level x postRP | −0.015*** | | |
| | (0.003) | | |
| IndInst avg x postRP | | −0.013*** | |
| | | (0.004) | |
| IndInst dropping others x postRP | | | −0.014*** |
| | | | (0.003) |
| Municipality FE | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ |
| State x time FE FE | ✓ | ✓ | ✓ |
| Time Invariant Controls | ✓ | ✓ | ✓ |
| Time variant Controls | ✓ | ✓ | ✓ |
| *N* | 6,297 | 6,297 | 6,297 |
| $R^2$ | 0.068 | 0.067 | 0.069 |

*Notes:* This table presents the estimates examining the effects of *IndInst* on road share using a more conservative approach in the construction of *IndInst* when groups had a partial Murdock's categorisation on jurisdictional hierarchy. Column (1) shows the coefficient on *IndInst* in which the groups with a partial categorisation in their jurisdictional hierarchy were downgraded one level. Column (2) presents the results using an index of Indigenous institutions in which these same groups with a partial categorisation were assigned the average level of jurisdictional hierarchy from the groups that reported a complete categorisation. Column (3) shows the coefficient on *IndInst* in which the Indigenous population corresponding to the groups classified as "others" was excluded. All columns include municipality, time and state-by-time fixed effects and time-variant and time-invariant controls. Standard errors are clustered at the municipality levels in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table A.9: Questions: collective action preferences

| Column | Question | Answer and coding |
|---|---|---|
| Q1 | In the last 12 months, have you participated in a demonstration or protest march? | 0=No / 1=Yes |
| Q2 | In the last 2 weeks have you tried to help to solve a problem in your community or your neighbourhood? | 0=No / 1=Yes |
| Q3 | Of people participating in legal demonstrations. How much do you approve or disapprove? | 0=dissaprove; ... ; 10=approve |
| Q4 | Of people seizing private property or land in order to protest. How much do you approve or disapprove? | 0=dissaprove; ... ; 10=approve |
| Q5 | Of people participating in the blocking of roads to protest. how much do you approve or disapprove? | 0=dissaprove; ... ; 10=approve |

Table A.10: Mean comparison

| | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| centralisation | 0.038 | 0.099 | 0.272 | −0.318 | 1.411*** | 0.794* |
| | (0.080) | (0.070) | (0.553) | (0.399) | (0.408) | (0.408) |
| Constant | 0.087 | 0.157*** | 6.042*** | 8.432*** | 1.622*** | 3.275*** |
| | (0.060) | (0.051) | (0.457) | (0.277) | (0.216) | (0.216) |
| Mean dependent variable | 0.111 | 0.219 | 6.214 | 8.235 | 2.500 | 3.775 |
| SD dependent variable | 0.317 | 0.415 | 2.954 | 2.045 | 2.369 | 2.766 |
| $N$ | 63 | 137 | 131 | 98 | 98 | 138 |
| $R^2$ | 0.003 | 0.013 | 0.002 | 0.006 | 0.084 | 0.019 |

*Notes:* This table presents the unconditional mean comparison between centralised and fragmented individuals of our selected questions in table A.9. In all columns, constant shows the unconditional mean of fragmented groups and centralisation the differences between politically centralised and politically fragmented individuals. Standard errors are clustered at the municipality level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table A.11: Identity preferences: survey evidence

| | proud | costum | feel |
|---|---|---|---|
| | (1) | (2) | (3) |
| centralisation | 0.106* | 0.167** | 0.103* |
| | (0.056) | (0.077) | (0.058) |
| Demographic controls | ✓ | ✓ | ✓ |
| Municipality FE | ✓ | ✓ | ✓ |
| Mean dependent variable | 0.884 | 0.565 | 0.843 |
| SD dependent variable | 0.320 | 0.496 | 0.364 |
| $N$ | 484 | 497 | 472 |
| $R^2$ | 0.194 | 0.223 | 0.293 |

*Notes:* This table presents mean comparison between politically centralised and politically fragmented individuals of our selected questions in section 2.5.3. In column (1), we use question *mi3*: "How proud do you feel of your tradition and heritage". In column (2), we use question *mi2*: "Do you preserve some of the custom or traditions of your ancestors?". In column (3), we use question *mi4*: "Do you feel... []...more [*ethnicity i*] than Mexican or []...as [*ethnicity i*] as Mexican"[...] or "...more Mexican than [*ethnicity i*]" . In all columns we control for demographic controls and municipality fixed effects. Standard errors are clustered at the municipality level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

# Chapter 3

# NAFTA and drug-related violence in Mexico

This chapter is based on Hidalgo, Hornung, and Selaya (2022).[1]

## 3.1 Introduction

Trade agreements can create economic shocks that result in criminal activity and violence. The literature proposes two potential channels for this: an opportunity cost effect and a rapacity effect.[2] If shocks negatively affect labor markets, the resulting worker displacement can lead to increases in violence because of a decline in the opportunity costs of criminal activity. If shocks are favorable to natural resources, the resulting increase in income derived from their ownership is associated with violent conflict because of increased returns to appropriating the resources.

This paper advances a novel mechanism according to which positive trade shocks trigger the rapacity effect. The proposed mechanism is rooted in the complementarity between trade in legal and illegal goods that results from the clandestine cross-border transportation of illegal goods hidden between legal goods. We suggest that trade liberalization agreements, by facilitating the unchecked exchange of legal goods, unintentionally increase the profits from smuggling illicit goods. When experiencing such positive shocks, firms in the illicit goods sector compete over profits by using violence, due to the absence of legally enforced property rights. Therefore, according to our hypothesis, trade-induced

[1] We thank Rod Abouhard (discussant), Christian Ambrosius (discussant), Anna Bindler, Claudio Ferraz, Kai Gehring, Markus Ludwig, Susanne Prantl as well as seminar audiences at HU Berlin, Cologne, Copenhagen, Nottingham, and the University of South Florida, conference audiences at BCDE UPSA 2021, EEA Copenhagen 2021, GLAD Göttingen 2021, 14th RGS Doctoral Conference in Economics, SAEe Barcelona 2021, and audiences at the workshops 'The dynamics of conflict and change' at UCL and 'Instrumentalizing Economics for Political Goals, Instrumentalizing Politics for Economic Goals' at WZB Berlin for comments. We are grateful to Antonia Weddeling for her excellent research assistance. Pablo Selaya thanks the Criminal Entanglements project, CRIMTANG ERC-2016-COG ERC Consolidator Grant, for funding this research. Erik Hornung acknowledges that his research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2126/1 - 390838866.

[2] For important contributions to this literature see, e.g., Becker (1968), Collier and Hoeffler (2004a), Dal Bó and Dal Bó (2011), Dube and Vargas (2013), and Dell, Feigenberg, and Teshima (2019).

positive shocks to illicit markets incentivize firms to invest in conflict and capture strategically important locations such as production sites or smuggling routes using violence.

We provide evidence for our hypothesis by exploiting the positive income shock to the drug-trafficking sector induced by Mexico's accession to the North American Free Trade Agreement (NAFTA). NAFTA came into force in 1994 and eliminated most barriers to trade between Mexico and the United States. By massively increasing the cross-border flow of legal goods while lowering inspection rates at the border, NAFTA lowered the cost of smuggling illegal drugs into the U.S. (andreas1996us,andreas2012border). This increased profits in the Mexican drug-trafficking sector and thereby the value of controlling this illicit sector.

NAFTA increased profits of Mexican drug-trafficking organizations (DTOs) because their main expenses derive from the cost of transporting illegal drugs across the U.S. border into consumer markets. Next to trafficking locally produced cannabis and opium, from the mid-1980s, Mexican DTOs provided cocaine trafficking services to cartels in Colombia. Due to increased interdiction efforts, Andean cartels shifted from Caribbean maritime and air routes to Mexican overland routes. Mexican cartels thus were able to negotiate a 50% cut of the transported cocaine from Colombia and became major players in the cocaine business themselves (Cockburn and Clair, 1998).

Empirically, we test the hypothesis that Mexican regions traversed by drug-trafficking routes saw higher increases in drug-related violence after the introduction of NAFTA than other regions. To test this hypothesis, we combine municipal-level panel data on drug-related homicides with predicted optimal drug-trafficking routes. Optimal drug-trafficking routes are predicted by connecting locations of major drug eradication and seizures of illegal drugs in Mexico with all U.S. land ports of entry using Dijkstra's algorithm (Dijkstra, 1959). Using the full extent of the road network, we predict which Mexican municipality is located on a drug-trafficking route.

Our empirical strategy exploits the fact that the introduction of NAFTA in 1994 increased the value of controlling the corridors for transporting illegal drugs into the U.S. for DTOs. We thus expect that municipalities located on a drug-trafficking route experience an increase in drug-related violence after the implementation of the trade agreement. To analyze the consequences of NAFTA for drug-related violence, we use difference-in-differences models that compare the number of drug-related homicides per 100.000 inhabitants in municipalities with and without a drug-trafficking route before and after 1994.

The results confirm that the introduction of NAFTA is associated with an increase in drug-related homicides by approximately 2.3 per 100.000 inhabitants in municipalities with a predicted drug-trafficking route. The increase in homicides is economically sizable, i.e., NAFTA is associated with 27% more homicides compared to the pre-NAFTA mean.

Furthermore, the results show a positive association between the length of routes within a municipality and drug-related homicides, suggesting that violence concentrated in municipalities with longer segments of trafficking routes. When studying violence alongside the routes, we find that violence increased more in upstream segments of routes after NAFTA's introduction.

We corroborate the validity of our identification strategy in several ways. First, using an event-study design, we show that trends in drug-related homicides did not differ across municipalities with and without routes prior to NAFTA. Second, using falsification tests, we show that the introduction of NAFTA is neither associated with changes in homicides of demographics that are typically not involved in the trafficking business, such as females and older people, nor with other causes of deaths, such as suicides and traffic fatalities. Third, using regions that predominantly produced maize that arguably suffered the strongest from import competition due to NAFTA as origins to generate placebo routes, we find no change in drug-related homicides in municipalities traversed by such a route after 1993. These checks confirm that the estimated increase in drug-related homicides in municipalities with a predicted drug-trafficking route is triggered by NAFTA and is not confounded by the detrimental effects of import competition in maize.

Clearly, NAFTA affected the Mexican economy in ways that potentially constitute alternative explanations for our estimated effects on violence. For example, Mexican producers faced increased trade competition leading to job losses, especially in the agricultural sector and particularly for maize farmers. This may have reduced the opportunity costs of using violence for farmers due lower agricultural incomes. Hence, if there is a spatial correlation between our predicted drug-trafficking routes and the local exposure to a negative trade shock, our results might not be driven by an increase in profits for Mexican cartels.

To rule out that our results are explained by the shock to legal trade directly, we add several control variables that aim to capture trade competition. Following the trade literature, we include variables that interact municipal-level employment shares with aggregate imports by sector. Other specifications control for changes in unemployment after NAFTA, an interaction term between the local maize suitability and the national maize price, and an interaction term between small farms focusing on maize production (ejidos) and the national maize price. The results suggest that the opportunity cost channel might indeed be present in the manufacturing sector but that this effect does not confound a rapacity effect of violent competition over rents from trafficking routes.

To study changes in the spatial distribution of drug-related violence after NAFTA's introduction, we run a local polynomial regression to inspect differential changes depending on distance to trafficking routes. We find that violence was diverted from regions further away to those in close proximity of routes. Rather than increasing aggregate violence, NAFTA seems to have had displacement effects.

Finally, using homicides from the period 2007–2010, we inspect whether NAFTA had lasting consequences on the geography of violence and whether homicides on trafficking routes were indeed a result of inter-cartel competition. In cross-sectional regressions we show that municipalities traversed by routes still experienced higher levels of drug-related homicides in the 2000s. Furthermore, route location is significantly related to homicides resulting from inter-cartel conflict but not to homicides resulting from confrontations or aggression between cartels and the military or police forces. This aligns with our hypothesis that DTOs use violence to compete over trafficking routes.

We contribute to several branches of the literature. In its broadest sense, this paper contributes to the literature on income shocks and civil conflict. This literature studies how exactly income shocks are related to violence and conflict, i.e., by changing the opportunity costs of using violence and insurrection or by increasing the value of resources usually owned by the state and thereby increasing the incentives to seize the state (Angrist and Kugler (see, e.g., 2008), Bazzi and Blattman (2014), Berman and Couttenier (2015), Berman et al. (2017), Collier and Hoeffler (1998, 2004a), Dal Bó and Dal Bó (2011), Dube and Vargas (2013), and Miguel, Satyanath, and Sergenti (2004)). We add to this literature by highlighting that the complementarity between legal and illegal trade proposed by Russo (2014) results in positive income shocks for sectors engaged in illegal goods trade, thereby increasing the incentives to seize such sectors using violence in the absence of enforced property rights.[3]

Furthermore, our paper relates to a growing literature studying the relationship between trade liberalization and crime. This literature focuses on analyzing how trade shocks change labor market conditions, which ultimately induces violence and property crimes (see Deiana, 2016; Dix-Carneiro, Soares, and Ulyssea, 2018; Iyer and Topalova, 2014). Dell, Feigenberg, and Teshima (2019) show that job losses in the manufacturing industry in Mexico induced by trade competition with China, increased cocaine trafficking and violence. The authors argue that changes in local labor markets lower the opportunity cost of criminal activity, which led to an increase of violence and drug-trafficking. We contribute to this literature by providing evidence that next to the effect on labor markets, trade liberalization may affect illegal markets by changing profits for trafficking organizations due to a complementarity between legal and illegal trade.

More narrowly, the intensity of the Mexican drug war has drawn the attention of economists. Between 2007 and 2010, the National Security Council registered over 50,000 drug-related homicides. In 2016 alone, Mexico registered 23,000 intentional homicides, equivalent to 17 intentional homicides per 100.000 inhabitants (IISS, 2017). Most of

[3]When studying the smuggling of illicit goods, the literature often implicity assumes such complementarities between legal and illegal trade. For examples, see Fisman and Wei (2009) on cultural property or Dube, Dube, and García-Ponce (2013) on illegal weapons.

these homicides were arguably caused by conflicts between DTOs competing for the control of territories and by government interventions (Calderón et al., 2015; Castillo and Kronick, 2020; Dell, 2015; Osorio, 2015), an increase in the drug-profits for drug trafficking organizations (Castillo, Mejía, and Restrepo, 2020; Sobrino, 2020), or increases in unemployment (Dell, Feigenberg, and Teshima, 2019).

There are few studies that explore the early development of the drug industry in Mexico during the 1990s, the golden years of Mexican cartels that may have paved the path for current conflicts.[4] Dube, García-Ponce, and Thom (2016) analyze how international commodity-price fluctuations, driven by the introduction of NAFTA in 1994, affect the illegal drug market in Mexico. Their findings imply that a decrease in maize prices increased the cultivation of cannabis and opium, leading to more intense activity of Mexican cartels and violence. Trejo and Ley (2018) show that Mexican cartels increasingly resorted to violence after they lost government protection due the increase in political competition starting in the 1990s. We contribute to this literature by providing evidence that the introduction of NAFTA is associated with a lasting increase of drug-related violence in places that were of strategic importance for drug trafficking.

The consequences of NAFTA's introduction have been largely discussed in the economic literature. By combining trade data with post-NAFTA survey studies, Burfisher, Robinson, and Thierfelder (2001) find that both the U.S. and Mexico benefited from the trade agreement, with much larger relative benefits for Mexico.[5] By inspecting the consequences of NAFTA for an illegal market, we add a hitherto neglected perspective to this literature.

## 3.2 Background

### 3.2.1 The rise of DTOs and violence in Mexico

Throughout the 1970s and 1980s illegal drugs, especially refined cocaine and heroin, were typically shipped from Latin American producers to U.S. consumer markets via maritime routes through the Caribbean.[6] Mexico was the world's largest producer of cannabis (more than 50% of worldwide production) and also a source country for opium and methamphetamine consumed in the U.S. The U.S. was the world's largest market for the consumption of cocaine, which assumes approximately two thirds of U.S. total expenditures on illicit drugs (The White House, 1992, p. 78).

[4] Murphy and Rossi (2020) show that the location of cartels in the 2000s can be traced to Chinese immigration of the early 20th century.

[5] Studies of NAFTA's effects on labor markets include Robertson (2000) who finds that the U.S.-Mexican labor market was highly integrated prior to NAFTA. Juhn, Ujhelyi, and Villegas-Sanchez (2013) study the relationship between trade liberalization and gender equality, and provide evidence that the introduction of NAFTA increased relative wages and employment of women in blue-collar jobs, but not in white-collar jobs.

[6] Colombia, Peru, and Bolivia account for virtually the total worldwide coca leaf cultivation.

When authorities increased interdiction efforts in the mid-1980s, Colombian drug-lords shifted their operations from maritime to Mexican overland routes. Colombian cartels started to increasingly rely on Mexican DTOs, experienced in trafficking cannabis into the U.S., for the trafficking services in cocaine. The extensive land border with many ports of entry into the U.S. allowed Mexican DTOs to cross the border with small amounts at high frequency, thereby hedging the risk of detection and seizure as compared to large maritime shipments. Trafficking service became extremely profitable for Mexican cartels that exploited the bargaining power bestowed on them by their geographic location. In the early 1980s, they negotiated a 50% cut of the transported cocaine and moved from providing pure logistical services to becoming a major supplier of cocaine to U.S. markets (Cockburn and Clair, 1998, p. 361). The shift to Mexican routes is apparent in the data of the U.S. State Department, i.e., the percentage of cocaine entering the U.S. from Mexico shifts from negligible in the mid-1980s to 70% in 1995 (Andreas, 1996).[7] From 1992, the Southwest border also accounted for the majority of cocaine seized annually (DEA, 1995, p. 6).[8]

In the 1980s, the Mexican trafficking market was dominated by the Guadalajara Cartel under the leadership of Miguel Ángel Félix Gallardo, only competing with the Gulf cartel on the east coast. In 1987, Gallardo created the so-called *Federation* and divided his territory into "plazas", i.e., specified territories and trafficking corridors to the U.S. border, each controlled by individual drug lords that operated independently but were loyal to Gallardo. Following the arrest of Gallardo in 1989, the Federation was dismantled and broke into independent DTOs. Disputes and competition among the plazas led to an increasing use of violence. Table 3.1 shows the development of homicides and drug-related homicides (i.e. homicides of males age 15–39) in Mexico in the 1990s. This period marks the emergence of violent conflicts among drug cartels whose levels were surpassed only later during the massive outbursts of violence starting in 2006.

### 3.2.2 NAFTA and drug trade

Mexican drug traffickers primarily use private and commercial land vehicles to transport illegal drugs into the United States (DEA, 1995, p. 6). In 1990, a White House report acknowledged that such vehicles are "all but lost in the tremendous volume of legitimate trade and commerce between the two countries" (The White House, 1990, p. 69). From

[7] This number is based on 'intelligence estimates' and can be found in the National Drug Control Strategy reports as early as 1994 (The White House, 1992, p. 55). Such estimates are largely based on the amount of seizures which declines after 1994, likely to the lower frequency of inspections under the free trade agreement (Dermota, 1999, p. 17). Today the share of cocaine entering from Mexico is estimated at 90 percent.

[8] The UNODC (2010, p. 103) estimates that gross profits in the U.S. cocaine market amounted to 35 billion USD in 2008. 0.5 billion (1.5% of gross profits) went to farmers in the Andean regions. 2.9 billion USD (8% of gross profits) went to traffickers moving cocaine across the southern U.S. border, mostly Mexican cartels. U.S. wholesalers and U.S. mid-level dealers capture 85% of gross profits. Due to the fact that Mexican cartels also participate in the selling of cocaine in the U.S., they hold substantial stakes in the U.S. cocaine market.

Table 3.1: The development of trade and violence in the 1990s

| | Mexico to U.S. exports to GDP ratio | Homicides | | | |
|---|---|---|---|---|---|
| | | all | | male, ages 15-39 | |
| | | total | per 100k | total | per 100k |
| 1990 | 7.1% | 11,475 | 14.04 | 5,414 | 6.63 |
| 1991 | 6.0% | 12,646 | 15.17 | 6,290 | 7.54 |
| 1992 | 10.3% | 13,760 | 16.18 | 7,103 | 8.35 |
| 1993 | 8.6% | 13,510 | 15.57 | 6,862 | 7.91 |
| 1994 | 9.7% | 15,656 | 17.69 | 7,974 | 9.01 |
| 1995 | 18.4% | 15,386 | 17.10 | 7,733 | 8.60 |
| 1996 | 19.4% | 14,350 | 15.67 | 7,276 | 7.95 |
| 1997 | 18.6% | 13,363 | 14.34 | 6,598 | 7.08 |
| 1998 | 19.4% | 13,490 | 14.23 | 6,396 | 6.75 |
| 1999 | 20.1% | 12,068 | 12.51 | 5,593 | 5.80 |

*Notes:* The table shows intentional homicides and intentional homicides of males (ages 15–39) per year as total sum and per 100,000 population, linearly interpolating population censuses in 1990, 1995, and 2000. Column 5 shows the worth of Mexican exports to United States in (2016) USD as a fraction of Mexican GDP in the respective years.

January 1st, 1994, the North American Free Trade Agreement between Canada, the United States and Mexico entered into force and eliminated most tariff and non-tariff barriers to free trade between the three countries. Negotiations started in 1991 and ended in December 1992 with the signing of the treaty. As shown in Table 3.1, the value of exports to the U.S. as percentage of Mexican GDP more than doubled with the introduction of NAFTA.

According to Andreas (1996), NAFTA facilitated and encouraged the exports of illegal drugs via Mexico into the United States. Dermota (1999, p. 15) confirms that anything NAFTA did to promote regional trade also encouraged the trade of illicit drugs. For example, NAFTA led to a decline in inspections to avoid hampering commerce while at the same time the number of trucks crossing with cargo went from 1.9 million in 1993 to 2.8 million in 1994 to 3.5 million in 1996 (Andreas, 1996, p. 58; Dermota, 1999, p. 17). Indeed, the DEA confirms that detection rates at the ports of entry decline in the volume of trade (DEA, 2016). Thus, NAFTA most likely decreased the cost of concealing drug smuggling. Furthermore, NAFTA massively increased capital flows which made it easier to route drug profits out of the U.S. via Mexico (Dermota, 1999, p. 21).

Anecdotal evidence suggests that policy makers were aware of these potential unintended consequences. Assistant U.S. Attorney Glen MacTaggart said in 1993, “If Nafta provides opportunity for legitimate businesses, it may clearly provide opportunities for illegitimate businessmen," [...] “It’s almost common sense” (Andreas, 1996, p. 57). A U.S. official involved in the fight against drug traffickers stated: “The free-trade agreement makes the United States more accessible and convenient for traffickers”[...] “It gives these people better opportunities to smuggle drugs” (Weiner and Golden, 1993, p. 1). However, in the NAFTA agreements there is no section addressing potential concerns about the impacts of the trade agreement on illegal markets. “This was in the ‘too hot to handle’ category” according to Gary Hufbauer (Weiner and Golden, 1993, p. 1).

In Dermota (1999, p. 15), a Colombian trafficker explains that the free trade agreement between Colombia and Venezuela that emerged in 1993 allowed him to increase profits by about 75%, just by reducing shipping costs to the U.S. He estimates that shipping costs per kilogram of cocaine decline from 8,000 USD to 2,000 USD when shipping the product from Venezuela instead of going directly from Columbia to the U.S. Shipping costs from Mexico to the U.S. ranged between 1,000 and 1,500 USD per kilogram of cocaine during the late 1990s. He thus expected to cut costs even further once being able to ship through Mexico.[9]

### 3.2.3 Conceptual considerations

Following the depiction of events described above, we conclude that the introduction of NAFTA increased profits of Mexican drug lords. There are at least four ways how NAFTA may have increased profits for Mexican DTOs. First, revenues may have increased because the amount of drugs trafficked into the U.S. increased. However, Table B.1 in the Appendix shows that, except for methamphetamine, volumes of U.S. drug consumption are stable or slightly declining in the 1990s, while prices are mostly declining. Second, trafficking revenues may have increased because the share of drugs entering the U.S. via Mexico increased. While there is some evidence for this (see, e.g., Watt and Zepeda, 2012, p. 105), estimates of trafficking market shares are hard to obtain and we cannot conclude with certainty that this is the main channel. Thirdly, costs of trafficking declined because Mexico-U.S. ports of entry were less policed and seizures became less likely. Following the above mentioned example by the Colombian trafficker, NAFTA clearly lowered the unit costs of trafficking due to a decline in the risk of detection and kickbacks paid to officials. Combining two and three, the overall net profits of trafficking and selling drugs accruing to Mexican DTOs may have increased. Fourth, the cost of money laundering declined because it was easier to conceal drug money in the increased capital flows from the U.S. to Mexico. Examples of the Mexican president Carlos Salinas using Citibank branches in New York and Mexico to transfer drug money to branches in London and Switzerland support this notion.

Following theoretical considerations in the literature (Castillo and Kronick, 2020; Sobrino, 2020), we develop the hypothesis that the positive drug profit shock introduced by NAFTA was accompanied by violent competition over trafficking routes into the U.S. among DTOs in Mexico because controlling the routes guaranteed a constant stream of

[9] According to Scott Stewart, Vice President of Tactical Analysis of Stratfor, the price of cocaine "increases considerably once it leaves the production areas and is transported closer to consumption markets". The same kilogram of cocaine that can be purchased in Colombia's jungle for \$2200 will cost between \$5500 and \$7000 in Colombian maritime ports, \$10.000 in Central America, \$12.000 in southern Mexico, \$16.000 in northern Mexico, and ultimately between \$24.000 and \$27.000 in U.S. wholesale markets (Stewart, 2016).

income.[10] In the absence of interference by a formal legal system, illegal markets such as the drug market induce participants to use violence to compete for profits or territory (Goldstein, 1985; Jacques, 2010; Reuter, 2009). Following Sobrino (2020) and Castillo and Kronick (2020), we argue that the increase in profits due to NAFTA led to more violent competition among DTOs. This altered the geography of drug-related violence in Mexico by concentrating the conflict on strategically important drug-trafficking routes. Castillo and Kronick (2020) develop a (repeated) contest model in which increases in drug-related profits break low violence agreements and fuel violence among traffickers. Here, interdiction such as seizures of large drug shipments lead to violence if drug profits increase in the presence of an inelastic demand. If revenues increase by more than costs, i.e., the revenues generated from owning trafficking routes increase by more than the cost of acquiring the routes, cartels will invest in conflict. Similarly, in the theoretical framework of Sobrino (2020) positive demand shocks in illegal markets increase violence because they increase the value of controlling drug production and trafficking routes. In her model, the demand shock incentivizes cartels to invest into military capacity and to enter into violent competition over more valuable production sites.

We combine the evidence accumulated in this section with the recent theoretical consideration in the literature to argue that NAFTA's open border policy increased profits of Mexican DTOs, resulting in increased returns to owning trafficking routes, leading to more violent competition over territories containing these routes.

## 3.3 Data

To estimate the effect of the NAFTA-induced increase in drug-trafficking profits on violence in Mexico, we exploit panel data at the municipality level. Municipalities are the second-administrative level beneath the 31 states. Our sample includes all 2,398 municipalities that were part of the 1990 census. If a municipality was divided after this year, we aggregate the data to the administrative boundaries as of 1990.

We restrict the panel to the period 1990–1999 for two reasons. First, data on homicides at the municipality level are available only from 1990. Second the geography of violence in Mexico may have changed after 2000 when the National Action Party (PAN) formed the new government after 70 years of the Institutional Revolutionary Party (PRI) in power. Summary statistics are presented in Table B.2 in the Appendix.

**Drug-related homicides.** In the absence of a direct measure of drug-related homicides, we use the number of male homicides between the age of 15 and 39 per 100,000 inhabitants for the period 1990–1999 from the Mexican *Instituto Nacional de Estadística y Geografía*

[10] Note that DTOs controlling a specific plaza were able to impose a tax on drugs moved on their routes by other DTOs.

(INEGI) as a proxy.[11] In doing so, we follow Calderón et al. (2015, p. 1462) who argue that homicides in this gender-age cohort group in Mexico best resemble drug-related homicides.[12]

The map in Figure B.1 in the Appendix shows the change in drug-related homicides comparing the periods 1990–93 and 1994–99, before and after the introduction of NAFTA. Changes in drug-related homicides have substantial spatial variation and are not concentrated in specific regions.

**Drug-trafficking routes.** We predict the location of drug-trafficking routes using Dijkstra's algorithm (Dijkstra, 1959) similar to Dell (2015). Using this algorithm, we identify optimal paths based on distance between origins and destinations within the Mexican network of main roads and highways drawn from The Digital Chart of the World (DCW). For simplicity we assume that each origin ships one unit of "drug" to the closest destination following the road network.

Destinations consist of all 22 Mexico-U.S. land border-crossings since the majority of drug-traffic into the U.S. occurs at land ports of entries (see DEA, 2016). To select origins, we expand the approach of Dell (2015) who focuses only on drug-producing municipalities and add drug-trafficking municipalities. Drug-producing are municipalities that are above the 95th percentile of cannabis and opium poppy eradication in hectares per area between 1990 and 1993. Drug-trafficking are municipalities in which there was a positive amount of cocaine seized between 1990 and 1993. These amount to 76 origin municipalities. We rely on pre-NAFTA eradication and seizures, to avoid endogenous changes in routing and policing due to open borders. For the same reason, we abstain for using time-varying routes. The data on eradication and seizure are obtained from Dube, García-Ponce, and Thom (2016).

Figure B.2 in the Appendix shows the predicted optimal drug-trafficking routes. This information is used to create an indicator variable that assumes the value one if a municipality is traversed by at least one predicted drug-trafficking route. Alternatively, we measure the length of the route in a municipality in km. This allows us to test whether NAFTA changed violence when the segment of a route located within a municipality is longer. Furthermore, we create a variable that counts the number of tributaries that had flown into the route upstream from a traversed municipality. Following the concept of a stream order to indicate the branching of a river system introduced by Shreve (1966), this

[11] The only available data that specifically distinguishes drug-related homicides was collected by the Federal Mexican Government and measures the "Deaths presumably related to Drug Trafficking Organizations (DTOs)" from December 2006 to December 2010. We use these data in Section 3.6 to distinguish between conflict parties. The spatial correlation between post-NAFTA male homicides between the age of 15 and 39 (1994–99) and drug-related homicides (2006–2010) is 0.15.

[12] The authors reach this conclusion after comparing the minimum mean squared error of drug-related homicides in the period 2006–2010 and of all homicides in combinations of 5-year age cohorts between 15 and 64 years in the period 2006–2010.

variable adds the accumulated tributaries at any confluence of two routes. This allows us to test if NAFTA changed violence by more in upstream or downstream locations.

Since we do not have information on actual trafficking routes, we cannot assess the quality of our predictions. We expect measurement error to be classical, leading to attenuation bias. However, it is conceivable that cartels systematically used more remote highways to avoid violent conflict, leading to systematic measurement error in our predicted routes. This concern is again mollified by the use of routes predicted with pre-NAFTA data that are not prone to changes in errors due to the introduction of NAFTA.

**Baseline control variables.** In our preferred specification, the empirical analysis includes several control variables. These aim to exclude potentially confounding factors at the municipality level that may affect their location on a predicted drug-trafficking route but also deferentially affect violence after the introduction of NAFTA. All baseline control variables are time invariant but will be allowed to have time varying effects in the analysis.

A first set of controls aims at excluding confounding geographical characteristics. These include temperature and precipitation (Fick and Hijmans, 2017), the soil pH (IGBP-DIS, 1998), and the agro-climatically attainable maize yields (FAO).

We further add controls for the potential cultivation of cannabis and opium poppies. For legal reasons, FAO-GAEZ crop suitability indices are not available for illegal crops. Therefore, we create separate suitability measures based on the optimal conditions for the cultivation of *papaver somniferum* (opium poppy) and *Cannabidaceae* (cannabis) based on the FAO EcoCrop database. This procedure follows Sviatschi (2022) and Daniele, Le Moglie, and Masera (2020) and measures the optimal conditions for cultivating illegal crops. We define optimal suitability in terms of precipitation, temperature, and soil pH.[13]

Further, our set of baseline control variables includes the geographic distance to the U.S. border. Finally, we include the number of inhabitants in 1990 (INEGI) to account for differences in demographic characteristics at the municipality level.

**Map** Figure 3.1 depicts the main variables and the source of variation in our analysis in a single map. The fact that changes in drug-related homicides after 1993 seem to cluster alongside our predicted trafficking routes provides visual support of our subsequent econometric analysis. Figure B.3 in the Appendix shows average drug-related homicides

[13] According to EcoCrop, the optimal temperature to grow opium poppies (cannabis) is between 15 and 24 (15–28) degrees Celsius; annual precipitation should be between 800 and 1200 (600–1200) mm; and soil pH between 6.5 and 7.5 (6–7). To define which areas of Mexico are suitable for growing opium poppies, we collect temperature and precipitation data from the WorldClim database (Fick and Hijmans, 2017) and soil pH data from the Atlas of the Biosphere (IGBP-DIS, 1998). Second, we divide Mexico's area into grid cells of 0.05 x 0.05 degrees of latitude by longitude and create an indicator variable that takes the value one if cell $i$ falls within the optimal intervals for growing opium poppy and 0 otherwise. Finally, we calculate the share of cells within each municipality suitable for cultivating poppies and cannabis.

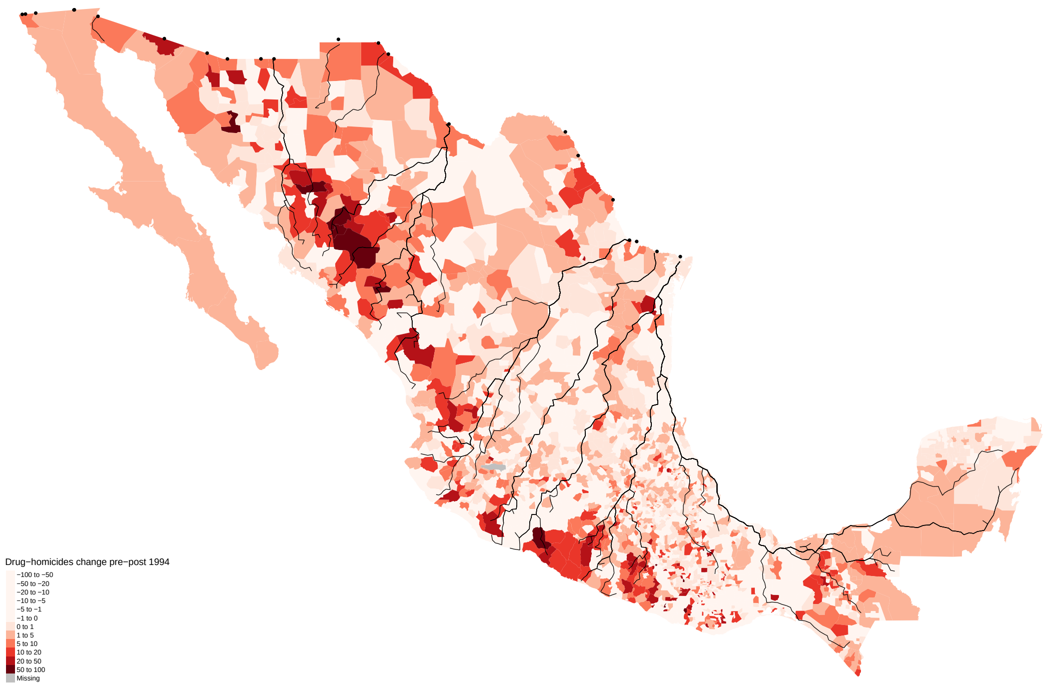



Figure 3.1: Predicted drug trafficking routes and changes in drug-related homicides

*This figure relates changes in drug-related homicides across Mexican municipalities to predicted drug trafficking routes. Darker shades of red indicate higher positive changes in drug-related homicides, i.e., homicides of males aged 15–39 per 100,000 inhabitants, comparing the periods 1990–93 and 1994–1999. Optimal predicted drug-trafficking routes are shown as black lines. Black dots depict the 22 land ports of entry on the Mexico-U.S. border.*

by municipalities with and without routes over time. A discernible jump in violence occurs in treated municipalities right after the introduction of NAFTA in 1994 and 1995. The resulting gap persists until the end of our study period despite the decline in both types of regions.

## 3.4 Empirical analysis

### 3.4.1 Empirical framework

Our main hypothesis is that NAFTA's open border policy resulted in higher profits for DTOs, leading to an increase in violent competition over trafficking routes. To test our hypothesis, we use a difference-in-differences (DiD) strategy and compare the change in drug-related homicides per 100.000 inhabitants after 1994 between municipalities with and without a predicted optimal drug-trafficking route. We apply the following specification:

$$Drug\,homicides_{it} = \alpha_i + \tau_t + \beta(Route_i \times post\,NAFTA_t) + \Gamma(X_i \times post\,NAFTA_t) + \epsilon_{it} \tag{3.1}$$

where the dependent variable $Drug\,homicides_{it}$ is the number of drug-related homicides per 100.000 inhabitants in municipality $i$ during years $t$ ($t \in 1990-99$). $\alpha_i$ are municipality-fixed effects that control for time-invariant characteristics. $\tau_t$ are year-fixed effects that control for common shocks to all municipalities in a specific year $t$. $Route_i$ is an indicator variable that takes the value one if municipality $i$ is traversed by a predicted drug-trafficking route and 0 otherwise. In an alternative specification, this measure is replaced by the length of the route in a given municipality.[14] $post\,NAFTA_t$ is an indicator, which assumes the value one for all years after 1994 and 0 otherwise. The time-invariant baseline control variables captured in the vector $X_i'$ are allowed to have differential effects following NAFTA's introduction via the inclusion of an interaction with the indicator $post\,NAFTA$. In all of our regressions standard errors are clustered at the municipality level.

The coefficient of interest $\beta$ captures differences in the change in drug-related homicides between municipalities with and without a predicted drug-trafficking route after the introduction of NAFTA. The validity of our identification strategy relies on the assumption that in the absence of NAFTA, drug-related homicides would have followed parallel trends between municipalities with and without drug-trafficking routes. We provide evidence for the absence of diverging trends prior to NAFTA using an event-study type specification following equation 3.2:

$$Drughomicides_{it} = \alpha_i + \tau_t + \sum_{t=1990}^{1999} \beta_t(Route_i \times year_t) + \sum_{t=1990}^{1999} \Gamma_t(X_i \times year_t) + \epsilon_{it}. \tag{3.2}$$

Equation 3.2 expands equation 3.1 by replacing the simple $post\,NAFTA_t$ indicator with time indicators ($year_t$). This specification allows us to observe whether drug-related homicides vary between municipalities with and without routes each year relative to the omitted baseline year 1993, the year before NAFTA's introduction.

### 3.4.2 Main results

#### 3.4.2.1 Simple DiD results

Table 3.2 presents the results from estimating equation 3.1. Column 1 shows results when including only municipality- and time-fixed effects whereas column 2 adds the baseline control variables described in Section 3.3. Our coefficient of interest is positive and statistically different from zero in both specifications. The $\beta$ coefficient in our preferred

[14] In Section 3.5 we also explore how the geography of violence changed in distance to the route.

Table 3.2: Drug trafficking routes and drug-related homicides after NAFTA

| Dep. var.: | Drug-related homicides per 100.000 inhabitants | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Route × post NAFTA | 2.378*** | 2.253*** | | 3.174*** |
| | (0.574) | (0.588) | | (0.776) |
| Length of route × post NAFTA | | | 1.245*** | |
| | | | (0.366) | |
| Number of tributaries × post NAFTA | | | | −0.194*** |
| | | | | (0.056) |
| Municipality FE | ✓ | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ | ✓ |
| Baseline controls | | ✓ | ✓ | ✓ |
| Mean homicides pre-Nafta | 8.218 | 8.218 | 8.218 | 8.218 |
| Observations | 23,980 | 23,980 | 23,980 | 23,980 |

*Notes:* The table shows results from estimating equation 3.1. The unit of observation is a municipality. *Length of route* is standardized with zero mean and unit standard deviation. The variable *Number of tributaries* counts the number of tributaries that had flown into the route upstream from a traversed municipality. Baseline controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors clustered at the municipality level in parenthesis. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

specification in column 2 shows that the introduction of NAFTA is associated with an increase of approximately 2.3 homicides per 100,000 inhabitants in municipalities on a drug-trafficking route. This reflects a substantial increase relative to the pre-NAFTA mean of approximately 8.2 homicides per 100,000, i.e., an increase of 27.42% with respect to the pre-NAFTA mean.

In column 3, we replace the route indicator with the length of the route. This measure is standardized with zero mean and unit standard deviation. The coefficient thus implies that a one standard deviation increase in the length of the route is associated with an increase in approximately 1.2 homicides per 100.000 inhabitants, i.e., 15% with respect to the pre-NAFTA mean. This result suggests that violence increased by more if a municipality contains longer segments of routes.

Column 4 tests whether violence increased by more in upstream or downstream municipalities among those traversed by a route following NAFTA. We add the Shreve stream order, i.e., a count of the number of tributaries that flow into the route upstream from the traversed municipality, interacted with the $post\,NAFTA$ dummy. Higher numbers reflect more downstream locations. We find a small but significant negative coefficient that can be interpreted to show that violence declines the further a municipality is located downstream. An increase in the stream order by one is associated with a decline of 0.194 homicides per 100.000.

#### 3.4.2.2 Event-study results

The validity of our DiD identification strategy relies on the common trends assumption. This requires that in the absence of the free trade agreement, violence would have followed parallel trends in treated and untreated municipalities. To inspect the validity of this assumption, we estimate the relationship of interest using an event-study design. This further allows us to study the dynamics of violence after the policy was introduced.

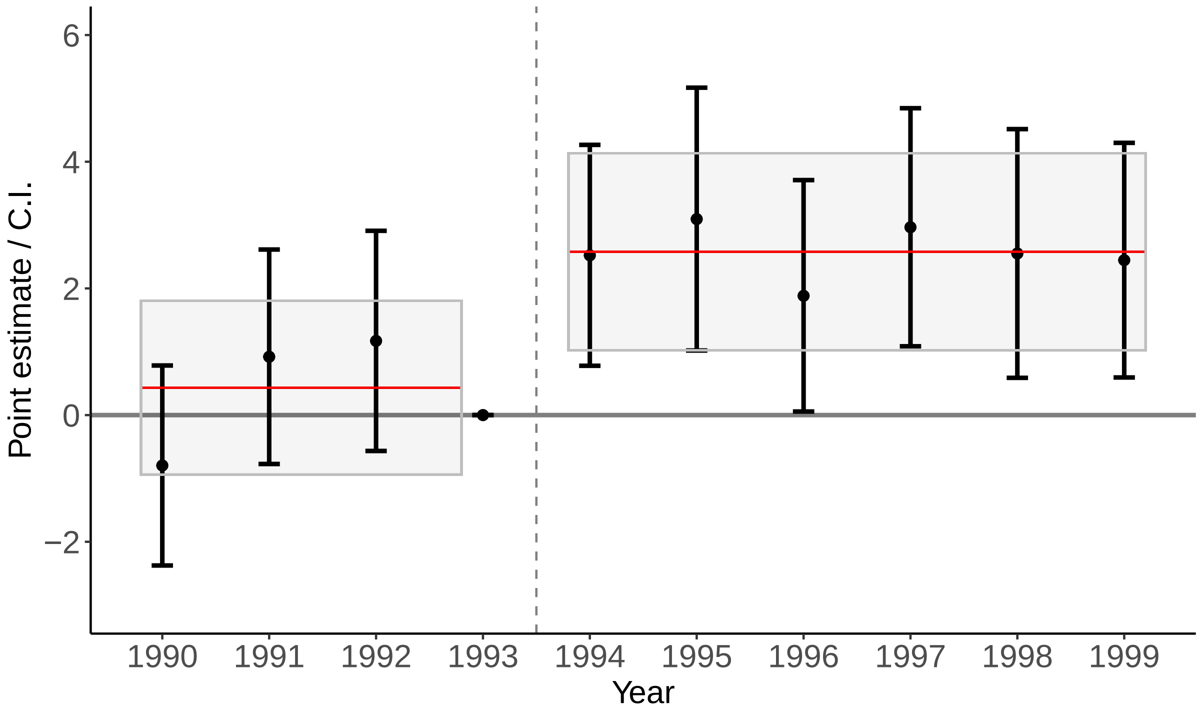



FIGURE 3.2: Main results: dynamic effects of drug-trafficking route location on violence

*The figure plots $\beta_\tau$ coefficients estimated from equation 3.2 with 95% confidence intervals. The omitted year is 1993. The dependent variable measures drug-related homicides per 100,000 inhabitants. The main explanatory variables are indicators that assume the value one if a municipality is traversed by a predicted optimal drug-trafficking route interacted with year dummies. Control variables include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors are clustered at the municipality level.*

Figure 3.2 plots the $\beta_t$ coefficients from an estimation of equation 3.2 over time. The figure shows differences in drug-related homicides between municipalities with and without a drug-trafficking route for each year with respect to 1993. Coefficients for all years after NAFTA's introduction are positive and significantly different from zero. This jump in violence occurs immediately in 1994 and remains of roughly the same magnitude (ca. 2.6 homicides per 100.000 inhabitants) until the end of our study period in 1999. The three pre-treatment coefficients are small and statistically indistinguishable from zero. Estimates comparing the average number of homicides between municipalities with and without routes before and after NAFTA with respect to their difference in

1993 (point estimates are indicated by red lines) confirm these findings. The absence of diverging trends corroborates the common trends assumption and thereby the validity of our identification strategy.

Figure B.4 in the Appendix inspects the other measures of drug-trafficking route exposure using the event-study design. Similar to the established pattern in Figure 3.2, violence is increasing in the length of the route immediately after 1993, again without significant differences in the years leading up to NAFTA (subfigure B.4a). Violence is also decreasing in the number of tributaries immediately after 1993, with no discernible pre-trends (see Figure B.4b). These graphs confirm findings from Table 3.2 and suggest that violence increases in locations with longer segments of a route but decreases in more downstream locations.

Results in this section support our findings and suggest that violence jumped to another level immediately after the introduction of NAFTA, whereas the absence of diverging pre-trends mollifies the concern of anticipatory effects.

### 3.4.3 Robustness checks

#### 3.4.3.1 Alternative transformations and sample restrictions

In this section, we show that our main results, estimated using equation 3.1, are robust to alternative transformations of the dependent variable and sample restrictions. The results of these exercises are presented in Table B.3 in the Appendix.

In columns 1–3, we use the logarithmic, the inverse hyperbolic sine, and the Castillo, Mejía, and Restrepo (2020) transformation of the dependent variable, respectively.[15] These transformations aim to remove skewness in our dependent variable that occurs due to a large number of municipalities with zero homicides. The coefficient of interest is positive and statistically significant across all specifications suggesting that skewness is not driving our results.

In column 4, we add an indicator variable for origin municipalities interacted with the *post* $NAFTA$ indicator as a control variable. This aims to rule out that our results are in fact driven by violence in drug-producing regions rather than regions on trafficking routes. Indeed, we find that origins experience a significant increase in homicides of about 3.2 per 100,000. However, our main results remain qualitatively similar.

In columns 5 and 6, we remove potential outliers from our sample. We exclude the top 1% and top 5% of municipalities in the distribution of drug-related homicides before the introduction of NAFTA, respectively. The coefficient of interest remains stable in comparison to our baseline specification.

[15] Castillo, Mejía, and Restrepo (2020) use the following transformation: $ln(h_{it} + r)$, where $h$ is the homicide rate in municipality $i$ at time $t$ and $r$ is the homicide rate at the 90th percentile of distribution.

#### 3.4.3.2 Trade shocks and unemployment

Our results can be interpreted to reflect the consequences of an increase in drug-trafficking profits if the predicted routes were indeed used for transporting drugs to the U.S. border. However, due to the complementarity between legal and illegal trade, it is conceivable that our routes overlap with general trade routes leading to an omitted variable bias if the change in legal trade due to NAFTA differentially affected regions alongside these routes. For example, if regions alongside the predicted routes suffered by more from import competition with U.S. producers, this may have increased unemployment, leading to a decrease in the opportunity costs of using violence that may create an upward bias on our estimated coefficient.

A similar bias could arise if the predicted routes are correlated with larger local agricultural or manufacturing sectors that suffer from import competition. In line with this argument, Dell, Feigenberg, and Teshima (2019) provide evidence that increased Chinese manufacturing exports to the U.S. negatively affected the Mexican manufacturing sector and caused job losses. The authors argue that the increase in unemployment due to competition with China increased cocaine trafficking and violence in the 2000s. In our case, an increase in exports from the U.S. to Mexico due to NAFTA could negatively affect Mexican workers pushing them to participate in the drug-trafficking industry leading to a rise in drug-related violence.

To test whether job losses induced by changes in trade with the U.S. after the introduction of NAFTA affect our results, we adopt two strategies: first, we check whether Mexican municipalities that were more exposed to imports from the U.S. to Mexico suffer higher levels of drug-related homicides, and second, we test whether municipalities that suffered higher increases in unemployment rates after the introduction to NAFTA suffer higher levels of drug-related homicides.

To calculate the magnitude of the trade shock for different sectors at the local level, we follow the trade literature and use a standard shift share approach. In our setting, the (time-invariant) share reflects the fraction of the respective population in municipality $i$ that is employed in the primary or secondary sector in 1990 using data from the *Instiuto Nacional para el Federalismo y Desarrollo Municipal* (INAFED).[16] The (time-varying) aggregate shift is calculated as the sector-specific (agricultural or industrial) value of imports of goods and services from the U.S. to Mexico in U.S. dollars as a fraction of the Mexican GDP from 1991 to 1999. These data were obtained from the World Bank's WITS database which does not contain information for the year 1990, resulting in a smaller sample.

The proxies for local sector-specific trade shocks are added to our preferred specifica-

[16] INEGI originally collected these data in the decennial General Census of Population and Housing in 1990. These data reflect the spatial distribution of sector-specific activity at the municipality level.

tion in Table 3.3. For better comparability these variables are standardized with mean zero and unit standard deviation. Since sector-specific exports are unavailable for 1990, we estimate our relationship in a sample for 1991–99 in column 1 for comparison. Column 2 controls for the exposure to the trade shock in the primary sector. Following trade liberalization, agricultural imports, especially of maize constituted a substantial negative shock for the agricultural sector in Mexico. A large number of small farms were no longer able to compete with U.S. maize after tariffs were abolished. We find that this negative shock does not significantly affect violence. Column 3, adds a similarly defined control variable that accounts for the trade shock in the secondary sector. We find that increased competition with the U.S. manufacturers led to an increase in violence in regions more exposed to this shock. This finding resonates with Dell, Feigenberg, and Teshima (2019) who argue that increased competition with China decreased the opportunity cost of violence for Mexican workers in the manufacturing sector. Across these specifications, our main findings remain qualitatively unchanged.

Table 3.3: Robustness to the shock in legal trade due to NAFTA

| Dep. var.: | Drug-related homicides per 100.000 inhabitants | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Route × post NAFTA | 1.880*** | 1.871*** | 1.858*** | 2.238*** | 1.860*** |
| | (0.626) | (0.624) | (0.625) | (0.591) | (0.628) |
| 1st sector emp. 1990 × Agricultural imports share | | −0.138 | | | 0.797 |
| | | (0.821) | | | (0.889) |
| 2nd sector emp. 1990 × Industrial imports share | | | 1.090** | | 1.348** |
| | | | (0.487) | | (0.527) |
| Δ male unemp. (1990-2000) × post NAFTA | | | | −0.413 | −0.416 |
| | | | | (0.622) | (0.756) |
| Δ female unemp. (1990-2000) × post NAFTA | | | | 0.079 | 0.015 |
| | | | | (0.273) | (0.311) |
| Municipality FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Baseline controls | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean homicides pre-Nafta | 8.30 | 8.30 | 8.30 | 8.19 | 8.32 |
| Observations | 21,582 | 21,582 | 21,582 | 23,930 | 21,537 |

*Notes:* The table shows results from estimating equation 3.1. The unit of observation is a municipality. All control variables introduced in this table are standardized with zero mean and unit standard deviation. Baseline controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors clustered at the municipality level in parenthesis. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

To capture the bias arising purely from changes in unemployment after the introduction of NAFTA, we add controls for the municipality-level change in unemployment for males and females from the *Instiuto Nacional para el Federalismo y Desarrollo Municipal* (INAFED). These data were originally collected by INEGI in the decennial General Census of Population and Housing and are thus only available for 1990 and 2000. Hence, we calculate the change in the gender-specific unemployment rate between these two years. Adding these variables in column 4, we do not find a significant relationship between an increase of male or female unemployment after the introduction of NAFTA and the increase of drug-violence.

Adding all of the controls at the same time in column 5, we find that the interaction term $Route \times post\, NAFTA$ remains positive and statistically significant with a magnitude of 1.86, which is not statistically different from our preferred specification in column 1. These results suggest that after the introduction of NAFTA, the increase of drug-related homicides in municipalities with a predicted drug-trafficking route is not driven by the shock to legal trade or increasing unemployment.

#### 3.4.3.3 Other omitted variables

This section addresses concerns with respect to the various potential drivers of violence in Mexico discussed in the literature. Again, we add these determinants as control variables to our preferred specification to account for the potential bias arising from omitted variables. Results of these robustness tests are displayed in Table 3.4.

Table 3.4: Robustness to alternative mechanisms

| Dep. var.: | Drug-related homicides per 100.000 inhabitants | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Route × post NAFTA | 2.253*** | 2.248*** | 2.038*** | 2.045*** |
| | (0.588) | (0.584) | (0.568) | (0.565) |
| Maize suitability × maize price | 3.735*** | | | 3.302*** |
| | (1.053) | | | (1.072) |
| Ejidos area × maize price | | −0.161 | | −0.210 |
| | | (0.774) | | (0.769) |
| Municipality alternation | | | −0.149 | −0.201 |
| | | | (0.365) | (0.363) |
| State alternation | | | 2.945*** | 2.804*** |
| | | | (0.661) | (0.666) |
| Municipality FE | ✓ | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ | ✓ |
| Baseline controls | ✓ | ✓ | ✓ | ✓ |
| Mean homicides pre-Nafta | 8.22 | 8.22 | 8.22 | 8.22 |
| Observations | 23,980 | 23,980 | 23,980 | 23,980 |

*Notes:* The table shows results from estimating equation 3.1. The unit of observation is a municipality. *Maize suitability × maize price* and *Ejidos area × maize price* are standardized with zero mean and unit standard deviation. Baseline controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors clustered at the municipality level in parenthesis. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

In column 1, we include a control variable that aims to capture a detrimental NAFTA shock that especially affected maize farmers. This control is inspired by Dube, García-

Ponce, and Thom (2016) who show that decreasing maize prices induced farmers to shift toward the cultivation of illegal crops in municipalities more suited to growing maize. NAFTA led to a decline in the price of maize in Mexico due to increased competition with U.S. producers, arguably leading to the estimated increase in the cultivation of illegal crops. To rule out that the increase in drug-related violence on trafficking routes is driven by an increase in the cultivation of illegal crops rather than drug-trafficking, we add an interaction term between the municipal-level attainable yields for maize from the FAO-GAEZ database and the annual fluctuation in the national price of maize in the 1990s as a control variable.

The results imply that the decrease in the national maize price actually decreased violence. While we have no explanation for this finding, we find that our coefficient of interest, the interaction term between *Route* and $post\,NAFTA$, is stable in size respect to our baseline specification.

In column 2, we aim to account for the fact that the decline in maize prices may have especially affected small farms called *ejidos*, created under communal land reforms since 1930. De Janvry, Sadoulet, and De Anda (1995) argue that most of these farms used obsolete technology and were relatively unproductive so that farmers lived in poverty already before NAFTA. In 1992, Mexico passed legislation for the privatization and sale of the such land potentially further confounding our effects.[17] Hence, we include an interaction term between the cumulative number of hectares reformed through the *ejido* systems until 1990 per hectares of a municipality's area and the national price of maize as control variable.[18] While the coefficient on this variable indeed shows that municipalities with more ejidos reacted to declining maize prices with an increase in violence, this does not confound our drug-trafficking effect.

In column 3, we aim to control for the fact that the political landscape started to change for the first time during the 1990s in Mexico. Trejo and Ley (2018) argue that political alteration and the rotation of parties increased the degree of inter-cartel violence by undermining the informal networks of protection that had facilitated the cartels' operations under the one-party rule of PAN. When a different party came to power, the informal network of protection was dismantled, and cartels fought for territories, because they lost the protection of the formal institutions.

Based on the data from Trejo and Ley (2018), we create two indicator variables to measure party alternation at the state and at the municipality level. The variable *state*

[17]Murphy and Rossi (2016) document an association between land reform and homicides in Mexico occurring after the 1992 liberalization of the ejido sector. The authors show that homicides decreased in municipalities that were more exposed to the land reform after 1992.

[18]We thank Aldo Elizalde for sharing these data from the *"Padron e Historial de Nucleos Agrarios" (PHINA)*, originally collected by the *Mexican Secretaría de Desarrollo Agrario, Territorial y Urbano (SEDATU)*. This institution provides information about the number of hectares reformed through the ejido system by municipality in intervals of 10-years between 1930 and 1990.

*alternation* (*municipality alternation*) switches from zero to one after an incumbent governor (mayor) from the PRI was replaced by a candidate from a different party. Information on party alternation is only provided for the time period 1995–2006. However, since we are interested to exclude a confounding effect that coincides with NAFTA, this shortcoming seems less crucial. Furthermore, the results are in line with Trejo and Ley (2018), who show that alternation of political parties at the state level but not at the municipal level is associated with an increased in drug-related homicides. Our coefficient of interest remains positive and statistically different from zero.

In column 4, we include all control variables at the same time. The result shows that the introduction of NAFTA is associated with an increase in 2.0 drug-related homicides per 100.000 inhabitants in municipalities with a drug-trafficking route compared to municipalities without a drug-trafficking route. This result is not statistically different from our main specification.

### 3.4.4 Falsification tests

We conduct two falsification tests to check the validity of our results. These either use homicides that are expected to be unrelated to drug trafficking as outcomes or use optimal routes for legal trade as the treatment variable.

**Drug-unrelated deaths** In line with our hypothesis, we expect that the introduction of NAFTA is only associated with drug-related homicides which predominantly occur among males at the age of 15 to 39. Hence, inspecting the effect of NAFTA's introduction on homicides in other demographic groups constitutes a valid falsification of our hypothesis. We do not expect routes to predict changes in homicides of females and older people, or deaths from suicides or traffic fatalities. To conduct these falsification tests, we estimate event study models as embedded in equation 3.2 and replace the dependent variable for homicides of young females (15–39 years), homicides of older males or females (55–64 years), deaths from suicides, and traffic fatalities. Figure B.5 in the Appendix plots the point estimates of the different falsification tests over time. None of the plots show a pattern of higher post-NAFTA increases in placebo deaths in municipalities on a route. These results support our interpretation that the estimated increases in drug-related homicides in our main specification reflect competition over trafficking routes.

**Placebo routes** An alternative interpretation of our results suggests that violence increases on routes that are used for trade in general. In a second falsification test, we again exploit the fact that maize was Mexico's primary export commodity which was especially affected by the introduction of NAFTA. If municipalities alongside trade routes that connected maize producing regions with the U.S. suffered more strongly from the trade shock, these might have experienced a stronger decrease in the opportunity costs of using violence. To address this issue, we estimate the event-study model embedded in equation 3.2 substituting our predicted optimal drug-trafficking route indicator with a predicted optimal maize-trading route indicator. To create such optimal placebo routes, we use Dijkstra's algorithm (Dijkstra, 1959) to connect U.S. land ports of entry with maize producing municipalities as origins, i.e., municipalities above the 90th percentile of attainable maize yields.[19] Figure B.6 in the Appendix plots the point estimates when using the placebo routes as the dependent variable. The plot does not indicate a differential increase in drug-related homicides across municipalities with and without optimal maize-trading routes after the introduction of NAFTA. These results support our interpretation of predicted routes as actual trafficking routes.

## 3.5 Spillover and displacement effects

This section investigates how NAFTA affected the spatial distribution of drug-related violence.

**Spillover effects** DTOs typically strive to control substantial territories and plazas instead of scattered pockets. This may lead to violence in regions that are not traversed by a trafficking route simply because DTOs try to conjoin territory. To study such spillovers, we investigate the effects of NAFTA's introduction on municipalities that are immediate neighbors of regions with a trafficking route.

Table B.4 in the Appendix estimates equation 3.1 and replicates the main table 3.2 but adds an indicator for being the neighbor of a trafficking municipality after NAFTA's introduction. A neighbor is defined as a municipality $j$ that shares a side or an edge with a municipality $i$ with a predicted drug-trafficking route. The results in columns 1–3 show that the interaction term $Neighbor \times post\,NAFTA$ is always positive and statistically different from zero with a size of approximately 2.3 homicides per 100,000. At the same time, compared to our baseline specification, the coefficient for municipalities traversed by a route increases substantially from 2.3 to 3.2 or 4.1 respectively. This result suggests that violence spilled over to neighboring municipalities. When not accounting for such spillovers, we underestimate the true increase in violence concentrating on the most valuable municipalities.

To inspect the dynamics of the main effect and the spillovers, we estimate equation 3.2 including the $Neighbor_j$ indicator interacted with time dummies ($year_t$). Figure B.7 in the Appendix shows the point estimates ($\beta_t$) of this flexible difference-in-differences approach for municipalities $i$ that are traversed by routes in panel A and for neighboring

[19] Results are robust to using municipalities above the median and above the 95th percentile of attainable maize yields as origins.

municipalities $j$ in panel B. Results in panel A and B show a similar pattern of no discernible pre-trends in violence between treated and control group municipalities. Upon the introduction of NAFTA, both route and neighbor municipalities experience a lasting shift in violence that is, however, only significantly different from zero for municipalities on routes. The smaller and slightly declining effect in neighbor municipalities shows that these regions are less contested.

**Displacement effects** Figure B.3 in the Appendix shows that municipalities on routes saw an increase of violence after NAFTA whereas municipalities off routes eventually saw a decrease, compared to pre-NAFTA levels. Hence, we aim to understand whether NAFTA induced a reorganization of violence in Mexico such that violence was diverted from low trafficking regions and concentrated on regions that were strategically more valuable to traffickers.

Figure B.8 in the Appendix shows local polynomial regressions of the (conditional) change in violence before and after NAFTA on the arcsinh transformed distance in km between a municipality's centroid and the closest predicted optimal drug trafficking route. The estimated coefficient on homicides is positive for distances up to ca. 55 km and becomes negative for municipalities further away. The graph also shows explicit humps for municipalities whose centroid is directly on the line and for those in 10–55 km distance. Given their average area is 809 $km^2$, these will typically be municipalities traversed by a route. This figure illustrates that NAFTA generated substantial displacement effects at the local level. Violence was diverted from remote regions to municipalities in close proximity to trafficking routes.

## 3.6 Evidence for lasting competition between DTOs

In this section, we use data on detected drug-related homicides between 2007 and 2010 to study whether NAFTA had lasting consequences on the geography of violence in Mexico and whether this violence is indeed driven by competition between DTOs. Starting in 2007, the newly elected president Felipe Calderón dispatched the military to fight the cartels and reduce drug violence. This marked a shift towards unprecedented levels of violence accumulating to 50,000 drug-related homicides during Calderón's term. Next to the fact that the 2007–2010 data collected by the Office of the Mexican Attorney-General specifically count homicides classified as drug trade-related by the officials, they also distinguish between inter-cartel homicides, i.e., those resulting from rivalries between DTOs, and homicides resulting from aggression and confrontations between drug-trafficking organizations and the police and military forces.

To study whether NAFTA had lasting consequences for the spatial distribution of violence across Mexican municipalities, we use the following cross-sectional specification:

$$Drug\,homicides_{i,2007-10} = \alpha + \beta Route_i + \delta Drug\,homicides_{i,1990-93} + X_i'\Gamma + \epsilon_i \quad (3.3)$$

where, $Drug\,homicides_{i,2007-10}$ reflects the average number of drug-related homicides by category during 2007–2010 per 100.000 inhabitants in 2005 in municipality $i$. Categories include homicides arising from a) conflicts between cartels, b) aggression from cartels toward state forces, and c) confrontations of cartels by state forces. To increase comparability, these variables are standardized with zero mean and unit standard deviation. $Route_i$ is the indicator variable established in Section 4.3 that assumes the value one if a municipality $i$ is traversed by a predicted optimal drug-trafficking route. $Drug\,homicides_{i,1990-1993}$ reflects the average number of male homicides between the age of 15 and 39 during 1990–1993 per 100.000 inhabitants in 1990. This variable captures the pre-NAFTA level of violence. $X_i'$ represents our baseline set of control variables including temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, and population size in 1990. We use robust standard errors in all specifications.

TABLE 3.5: Distinguishing violence among DTOs or between DTOs and the state

| Dep. var.: | Drug-related homicides (2007–10) per 100.000 inhabitants in 2005 due to: | | |
|---|---|---|---|
| | inter-cartel conflict | aggression towards state | confrontations with state |
| | (1) | (2) | (3) |
| Route | 0.225*** | 0.095 | 0.088 |
| | (0.058) | (0.059) | (0.058) |
| Drug-related homicides 1990–93 | ✓ | ✓ | ✓ |
| Baseline controls | ✓ | ✓ | ✓ |
| Observations | 2,398 | 2,398 | 2,398 |

*Notes:* The table shows results from estimating equation 3.3. The unit of observation is a municipality. Dependent variables are standardized with zero means and unit standard deviations. Baseline controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Robust standard errors in parenthesis. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 3.5 presents the results from estimating equation 3.3 by category of homicides. The results show that our predicted drug-trafficking routes are positively associated with any of the violence categories. However, we find that drug-trafficking routes are only significantly related to inter-cartel conflict in column 1, whereas they are not significantly related to aggression from cartels toward state forces in column 2 or confrontations of cartels by state forces in column 3. Homicides from inter-cartel conflict are 23% of a standard deviation more prevalent in municipalities traversed by a route. This result confirms our hypothesis that DTOs use violence when competing for trafficking routes. The results further suggest that drug-related violence continues to be higher in regions that became more valuable after NAFTA's introduction.

## 3.7 Conclusion

This paper investigates the consequences of the introduction of the North American Free Trade Agreement in 1994 on drug-related violence in Mexico. We argue that due to the complementarity between trade in legal and illegal goods, NAFTA's open border policy increased profits in the Mexican drugs-trafficking sector. This triggered the so-called rapacity effect leading to an increase in violent competition over smuggling routes among DTOs that ultimately changed the geography of violence in Mexico in the 1990s.

Using a flexible diff-in-diff model, we compare drug-related homicides per 100,000 inhabitants in municipalities with and without an optimal drug-trafficking route, before and after the introduction of NAFTA. Trafficking routes are determined via the least-cost-paths between U.S. points of entry and regions suitable for opium production. Results show that after NAFTA introduction, homicides increase by an additional 2.25 deaths per 100,000 inhabitants in municipalities with an optimal drug-trafficking route, equivalent to an increase of 27% relative to the pre-NAFTA mean.

We think that our findings generalize to all types of illegal goods that are smuggled across borders and extend to human trafficking. When considering the gains from trade liberalization, policy makers need to consider the fact that also the trade costs of illegal goods decline, triggering the rapacity of criminals. However, we also think that our findings only generalize to countries with weak institutions where corruption is a major source of income for the government officials.

# Appendix

## B.1 Descriptives

Table B.1: The size of the U.S. illegal drugs market

| | Cocaine | | Heroin | | Cannabis | | Methamphetamine | |
|---|---|---|---|---|---|---|---|---|
| Year | Consum. | Expen. | Consum. | Expen. | Consum. | Expen. | Consum. | Expen. |
| 1990 | 447 | 69.9 | 14 | 22.5 | 837 | 15.0 | 16 | 5.7 |
| 1991 | 335 | 57.1 | 12 | 20.3 | 793 | 14.0 | 10 | 3.7 |
| 1992 | 346 | 49.9 | 12 | 17.2 | 761 | 14.6 | 14 | 4.8 |
| 1993 | 331 | 45.0 | 11 | 13.8 | 791 | 12.0 | 19 | 5.1 |
| 1994 | 323 | 42.8 | 11 | 13.2 | 874 | 12.2 | 34 | 7.6 |
| 1995 | 321 | 40.0 | 12 | 13.2 | 848 | 10.2 | 54 | 9.2 |
| 1996 | 301 | 39.2 | 13 | 12.8 | 874 | 9.5 | 54 | 10.1 |
| 1997 | 275 | 34.7 | 12 | 11.4 | 960 | 10.5 | 35 | 9.3 |
| 1998 | 267 | 34.9 | 14 | 11.1 | 952 | 10.8 | 27 | 8 |
| 1999 | 271 | 35.6 | 14 | 10.1 | 1028 | 10.6 | 18 | 5.8 |

*Notes:* Table shows size of the U.S. drug market based on prices and quantities. *Consum.* is annual quantity of consumption in metric tons. *Expen.* is annual expenditures in Billions of 2000 USD. Source: The White House (2003).

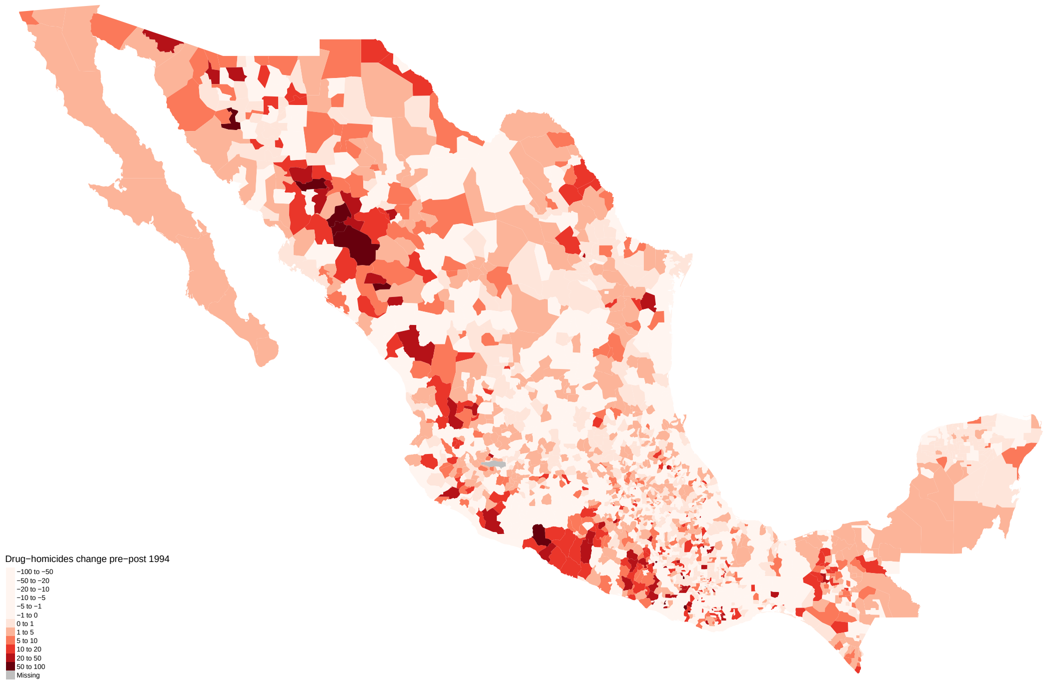



Figure B.1: Change in drug-related homicides after NAFTA's introduction

*This figure illustrates average changes in drug-related homicides, i.e., homicides of makes aged 15–39 per 100,000 inhabitants, comparing the periods 1990–93 and 1994–1999. Darker shades of red indicate higher increases after 1994.*

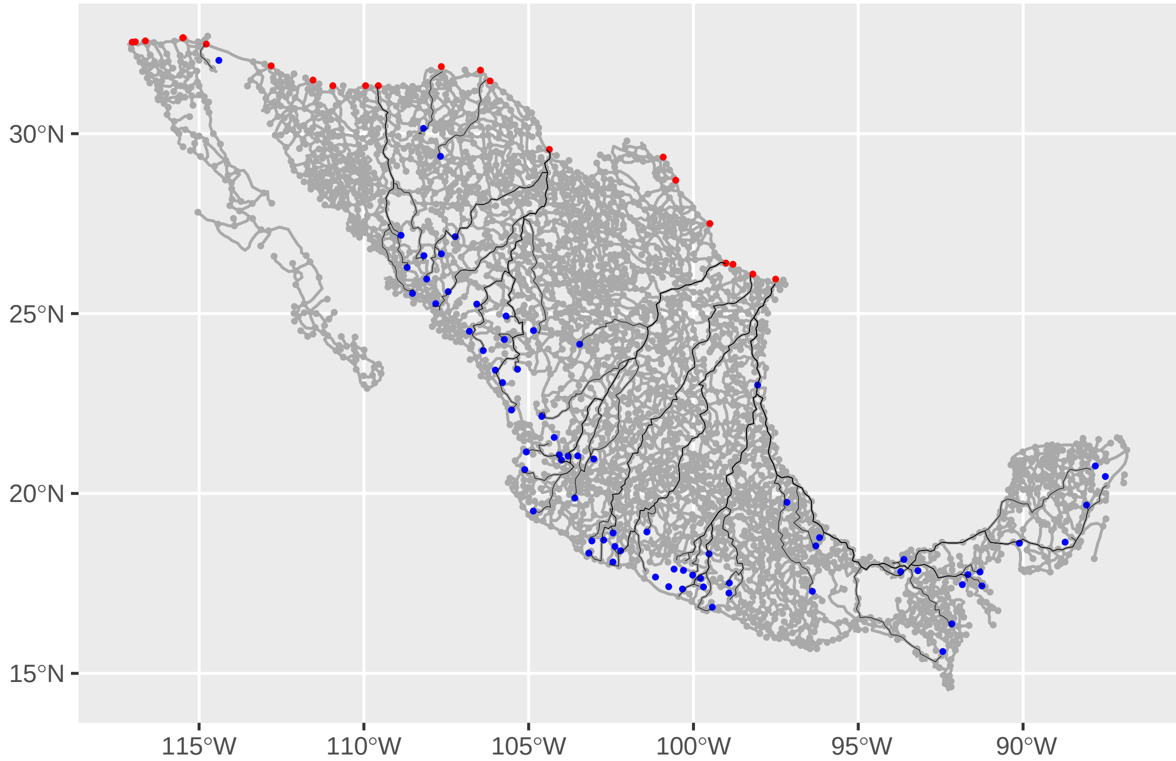



FIGURE B.2: Predicted drug-trafficking routes

*This figure shows the optimal predicted drug-trafficking routes as black lines. Grey lines indicate the full road network in Mexico as of 1993. Red dots depict the 22 land ports of entry on the Mexican-U.S. border. Blue dots depict major known drug-producing and drug-trafficking municipalities (see Section 3.3 for details).*


Figure B.3: Drug-related homicides by routes in the 1990s

*This figure depicts the evolution of drug-related homicides by municipality with (solid) and without (dashed) a predicted drug-trafficking route over time.*

Table B.2: Descriptive Statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| *Panel data* | | | | | |
| Drug-related homicides 1990-1999 | 23,980 | 8.17 | 20.07 | 0 | 607.29 |
| Drug-related homicides 1990-1993 | 9,592 | 8.22 | 21.67 | 0 | 571.43 |
| Drug-related homicides 1994-1999 | 14,388 | 8.14 | 18.92 | 0 | 607.29 |
| Female (age 15–39) homicides | 23,980 | 0.33 | 1.70 | 0 | 48.50 |
| Older male (55–64) homicides | 23,980 | 2.67 | 7.74 | 0 | 283.29 |
| Older female (55–64) homicides | 23,980 | 0.24 | 2.60 | 0 | 283.29 |
| Traffic fatalities | 23,980 | 8.75 | 17.04 | 0 | 636.94 |
| Suicides | 23,980 | 0.38 | 2.28 | 0 | 109.65 |
| State alternation | 14,388 | 0.10 | 0.30 | 0 | 1 |
| Municipality alternation | 14,388 | 0.19 | 0.39 | 0 | 1 |
| Number DTOs | 23,980 | 0.01 | 0.11 | 0 | 2 |
| *Cross-section* | | | | | |
| Inter-cartel conflict homicides | 2,398 | 5.79 | 17.38 | 0 | 368.93 |
| Aggression deaths | 2,398 | 0.15 | 1.37 | 0 | 39.96 |
| Confrontation deaths | 2,398 | 1.39 | 11.40 | 0 | 271.45 |
| Route | 2,398 | 0.25 | 0.43 | 0 | 1 |
| Length of route | 2,398 | 13.22 | 27.81 | 0 | 239.24 |
| Number of tributaries | 2,398 | 1.20 | 3.43 | 0 | 35 |
| Distance to route | 2,398 | 42,112.47 | 42,020.36 | 1.98 | 369,022.10 |
| Neighbour | 2,398 | 0.28 | 0.45 | 0 | 1 |
| Temperature | 2,398 | 19.93 | 3.94 | 10.68 | 27.93 |
| Precipitation | 2,398 | 1,025.31 | 550.71 | 77.99 | 3,925.58 |
| Soil pH | 2,398 | 64.86 | 9.74 | 1.33 | 85.74 |
| Cannabis suitability | 2,398 | 0.24 | 0.32 | 0 | 1 |
| Opium suitability | 2,398 | 0.06 | 0.15 | 0 | 1 |
| Maize suitability | 2,398 | 4.86 | 1.80 | 0 | 8 |
| Distance to U.S. border | 2,398 | 733,146.80 | 272,202.50 | 0 | 1,334,989 |
| $\Delta$ male unemp. 1990–2000 | 2,397 | −0.02 | 0.04 | −0.88 | 0.44 |
| $\Delta$ female unemp. 1990–2000 | 2,393 | −0.02 | 0.04 | −1.00 | 0.11 |
| 1st sector emp. 1990 | 2,398 | 0.53 | 0.24 | 0 | 1.00 |
| 2nd sector emp. 1990 | 2,398 | 0.18 | 0.13 | 0 | 0.85 |
| Ejidos area | 2,398 | 0.30 | 0.31 | 0 | 6.67 |
| Pop 1990 | 2,398 | 33,857.04 | 100,953.10 | 149 | 1,650,205 |
| *Time Series* | | | | | |
| Agricultural imports | 9 | 0.72 | 0.15 | 0.52 | 1.02 |
| Industiral imports | 9 | 13.43 | 4.18 | 7.60 | 17.84 |
| Maize price | 10 | 3.70 | 0.76 | 2.46 | 4.75 |

*Notes:* The table shows summary statistic for all variables included in the empirical analysis.

## B.2 Variable description

### B.2.1 Dependent variables: homicides and deaths

**Drug-related homicides.** The number of homicides of males between the age of 15 and 39 per 100,000 inhabitants in 1990 in each Mexican municipality for the period 1990–1999. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Female (age 15–39) homicides.** The number of homicides of females between the age of 15 and 39 per 100,000 inhabitants in 1990 in each Mexican municipality for the period 1990–1999. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Older male (55–64) homicides.** The number of homicides of males between the age of 55 and 64 per 100,000 inhabitants in 1990 in each Mexican municipality for the period 1990–1999. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Older female (55–64) homicides.** The number of homicides of females between the age of 55 and 64 per 100,000 inhabitants in 1990 in each Mexican municipality for the period 1990–1999. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Suicides.** The number of suicides per 100,000 inhabitants in 1990 in each Mexican municipality for the period 1990–1999. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Traffic fatalities.** The number of deaths due to traffic accidents per 100,000 inhabitants in 1990 in each Mexican municipality for the period 1990–1999. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Inter-cartel conflict homicides.** The number of homicides related to drug-trafficking organization rivalry per 100,000 inhabitants in 2005 in each Mexican municipality for the period 2007-2010. Data obtained from Empirical Studies of Conflict (ESOC). Total population in 2005 obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Aggression deaths.** The number of deaths related to DTO attacks against military and police forces per 100,000 inhabitants in 2005 in each Mexican municipality for the period

2007-2010. Data obtained from Empirical Studies of Conflict (ESOC). Total population in 2005 obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Confrontation deaths.** The number of deaths observed during a government intervention per 100,000 inhabitants in 2005 in each Mexican municipality for the period 2007-2010. Data obtained from Empirical Studies of Conflict (ESOC). Total population in 2005 obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

### B.2.2 Main explanatory variables: routes for drug-trafficking

**Route.** Indicator that takes value 1 if municipality $i$ is traversed by a predicted optimal drug-trafficking route, created using Dijkstra's algorithm (Dijkstra, 1959). Dijkstra's algorithm creates optimal transportation paths from origins to destinations following a network. The network is the main roads and highways in Mexico as of 1985, drawn from The Digital Chart of the World (DCW). Destinations consist of all 22 Mexico-U.S. land border-crossings collected from the Bureau of Transportation Statistics. Origins are municipalities of drug production or transit. We classify municipalities as origins if they are above the 95th percentile of cannabis and opium poppy eradication in hectares per area between 1990 and 1993, or if there was a positive amount of cocaine seized between 1990 and 1993. These amount to 76 origin municipalities. Data are obtained from Dube, García-Ponce, and Thom (2016).

**Length of route.** This variable measures the length predicted drug-trafficking routes in kilometers in each municipality. See '*Routes*' for prediction of routes.

**Number of tributaries.** This variable counts the number of tributaries that had flown into the routes upstream from an traversed municipality. Following the concept of a stream order to indicate the branching of a river system introduced by Shreve (1966), this variable adds the accumulated tributaries at any confluence of two routes. See '*Routes*' for prediction of routes.

**Distance to route.** This variable measures the euclidean distance from a municipality centroid to the nearest predicted drug-trafficking route in kilometers. See '*Routes*' for prediction of routes.

**Neighbor.** Indicator that takes value 1 if municipality $j$ shares a side or an edge with a municipality $i$ with a predicted drug-trafficking route. See '*Routes*' for prediction of routes.

### B.2.3 Baseline controls

**Temperature.** The average temperature in degrees Celsius in a municipality, calculated as the average temperature of all grid-cells in the municipality during the 1970—2000 period, constructed by temporally and spatially aggregating time series information on mean monthly temperature at a geospatial resolution of 30 arc seconds, obtained from WorldClim (version 2) by Fick and Hijmans (2017).

**Precipitation.** The average precipitation in hundreds of millimeters in a municipality, calculated as the average precipitation of all grid-cells in the municipality during the 1970—2000 period, constructed by temporally and spatially aggregating time series information on mean monthly precipitation at a geospatial resolution of 30 arc seconds, obtained from WorldClim (version 2) by Fick and Hijmans (2017).

**Soil pH.** The average soil pH in a municipality, calculated as the average soil pH of all grid-cells in the municipality, constructed information at a geospatial resolution of 30 arc seconds. The pH scale runs logarithmically from 0 to 14, where 0 is a highly acidic value, 14 is highly alkaline, and 7 is neutral. Data obtained from the Atlas of the Biosphere by IGBP-DIS (1998).

**Cannabis suitability.** Index which captures the suitability of a municipality to grow cannabis. To create this index, we divide Mexico's area into grid cells of 0.05 x 0.05 degrees of latitude by longitude and create a dummy variable that takes a value 1 if cell $i$ falls within the optimal intervals for growing cannabis and 0 otherwise. We define the optimal suitability for growing cannabis in terms of precipitation, temperature, and soil pH (see variable descriptions above). According to FAO EcoCrop, the optimal temperature to grow cannabis is between 15–28 degrees Celsius; annual precipitation should be between 600–1200 mm; and soil pH between 6–7. Using this information, we calculate the share of 0.05 x 0.05 grid cells within each municipality that potentially could grow cannabis.

**Opium suitability.** Index which captures the suitability of a municipality to grow opium poppies. To create this index, we divide Mexico's area into grid cells of 0.05 x 0.05 degrees of latitude by longitude and create a dummy variable that takes a value 1 if cell $i$ falls within the optimal intervals for growing opium poppies and 0 otherwise. We define the optimal suitability for growing opium poppies in terms of precipitation, temperature, and soil pH (see variable descriptions above). According to FAO EcoCrop, the optimal temperature to grow opium poppies is between 15 and 24 degrees Celsius; annual precipitation should be between 800–1200 mm; and soil pH between 6.5–7.5. Using this information, we calculate the share of 0.05 x 0.05 grid cells within each municipality that potentially could grow opium poppies.

**Maize suitability.** The average agroclimatic attainable yield for maize at the municipality level at a geospatial resolution of 30 arc seconds. We calculate the average of low-input-level, intermediate-input-level, and high-input-level rain-fed for maize at the municipality level. Data obtained from from FAO-GAEZ V3.

**Distance to U.S. border.** The euclidean distance in kilometers from a municipality centroid to the nearest Mexico-U.S. land port of Entry.

**Total population 1990.** Total population in 1990 at the municipality level obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

### B.2.4 Robustness controls

**1st sector emp. 1990.** The population working in the primary sector in 1990 at the municipality level, divided by the total active population in 1990 at the municipality level. Data obtained from Instituto Nacional para el Federalismo y el Desarrollo Nacional.

**2nd sector emp. 1990.** The population working in the secondary sector in 1990 at the municipality level, divided by the total active population in 1990 at the municipality level. Data obtained from Instituto Nacional para el Federalismo y el Desarrollo Nacional.

**Agricultural imports.** The total value of agricultural goods and services imported from the United States to Mexico as a percentage of the total Mexican GDP from 1991 to 1999. Data obtained from World Integrated Trade Solution (WITS). The WITS provides information about the value in USD of imports of different industries, we classify as agricultural sector: animal, vegetable.

**Industrial imports.** The total value of industrial goods and services imported from the United States to Mexico as a percentage of the total Mexican GDP from 1991 to 1999. Data obtained from World Integrated Trade Solution (WITS). The WITS provides information about the value in USD of imports of different industries, we classify as industrial sector: chemicals, food products, footwear, fuels, hides and skins, mach and elec, metals, minerals, miscellaneous, plastic or rubber, stone and glass, textiles and clothing, transportation, wood.

**Δ male unemp. 1990–2000.** The percentage change in male unemployment rates between 1990 and 2000 at the municipality level. Data obtained from Instituto Nacional para el Federalismo y el Desarrollo Nacional.

**Δ female unemp. 1990–2000.** The percentage change in male unemployment rates between 1990 and 2000 at the municipality level. Data obtained from Instituto Nacional para el Federalismo y el Desarrollo Nacional.

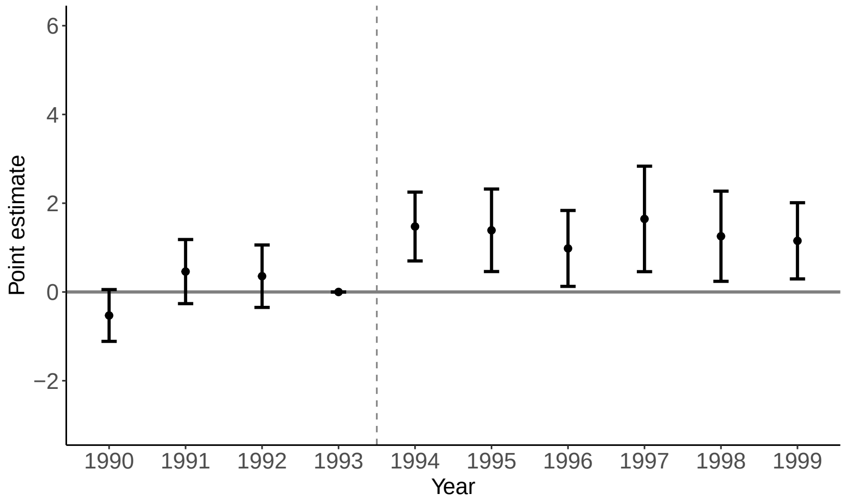
**Maize price.** National maize price in 2010 thousand Mexican pesos, obtained from Dube, García-Ponce, and Thom (2016) originally collected by *Servicio de Información Agroalimentaria y Pesquera (SIAP).*

**Ejidos area.** The cumulative area of the land redistributed under the *ejido* system divided by total area of a municipality in 1990. Data obtained from Elizalde (2020b).

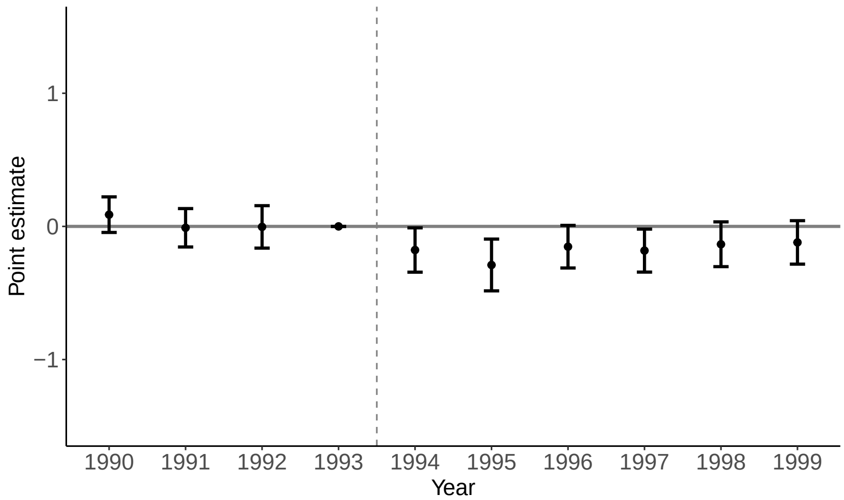
**Municipality Alternation.** Indicator variable assuming the value 1 if the major of the municipality is not affiliated with the PRI and 0 if the PRI candidate maintained power at the municipality level from 1995 to 1999. Data obtained from Trejo and Ley (2018).

**State Alternation.** Indicator variable assuming the value 1 for all municipalities in a state if the state governor is not affiliated with the PRI and 0 if the PRI candidate mantained power from 1995 to 1999. Data obtained from Trejo and Ley (2018).

## B.3 Robustness checks



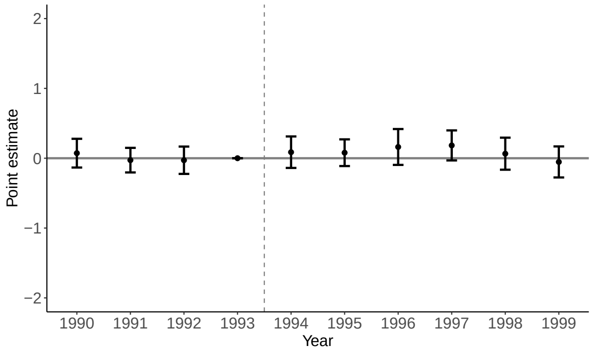

(A) Length of the route



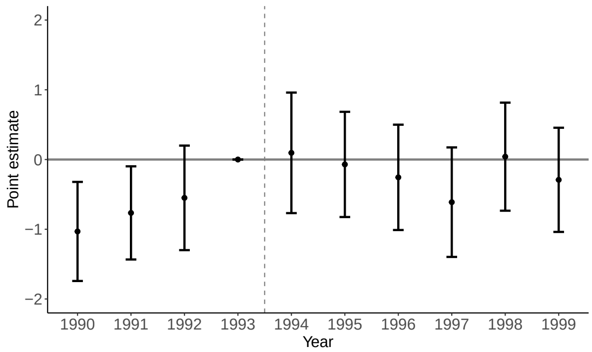

(B) Number of tributaries

FIGURE B.4: Event study estimates using alternative measures of drug-trafficking route exposure

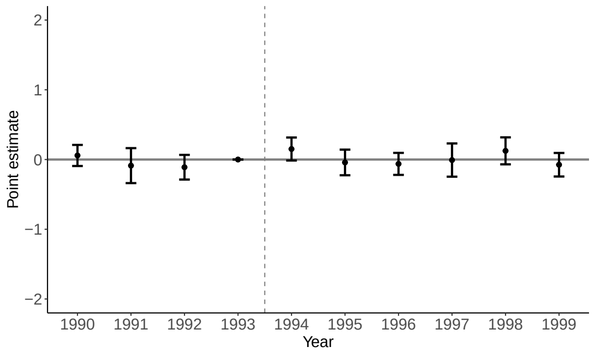
*These figures show event-study estimates based on equation 3.2 using alternative measures of drug-trafficking route exposure. Panel A plots the estimated point coefficient from a variable that measures the length of trafficking routes in a municipality, interacted with time dummies. Panel B plots the estimated point coefficient from a variable that counts the number tributaries that flow into the route upstream from the traversed municipality, interacted with time dummies. This estimation is conditional on a route indicator, interacted with time dummies. The unit of observation is a municipality. Controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors are clustered at the municipality level. 95% confidence band shown.*



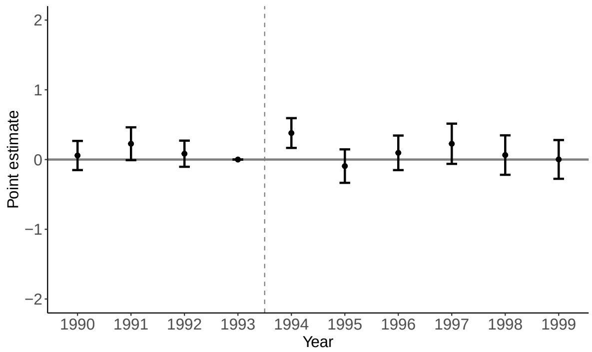

(A) Female (age 15–39) homicides



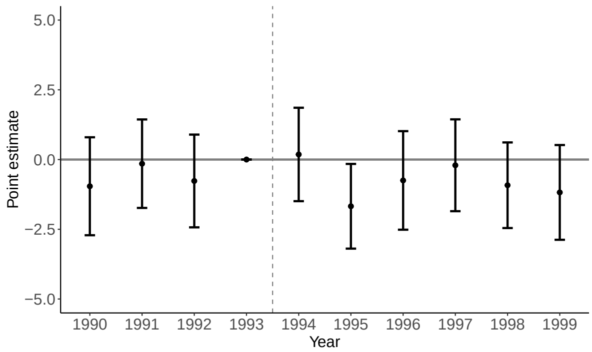

(B) Older male (55–64) homicides




(C) Older female (55–64) homicides




(D) Suicides




(E) Traffic fatalities

Figure B.5: Falsification test: drug-unrelated deaths

*These figures show event-study estimates based on equation 3.2 using (arguably) drug-unrelated death rates as outcomes for falsification purposes. Panel A uses homicides of females aged 15–39; Panel B uses homicides of males aged 55-64; Panel B uses homicides of females aged 55-64; Panel D uses all suicides, Panel E uses traffic fatalities. The unit of observation is a municipality. Controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors are clustered at the municipality level. 95% confidence band shown.*

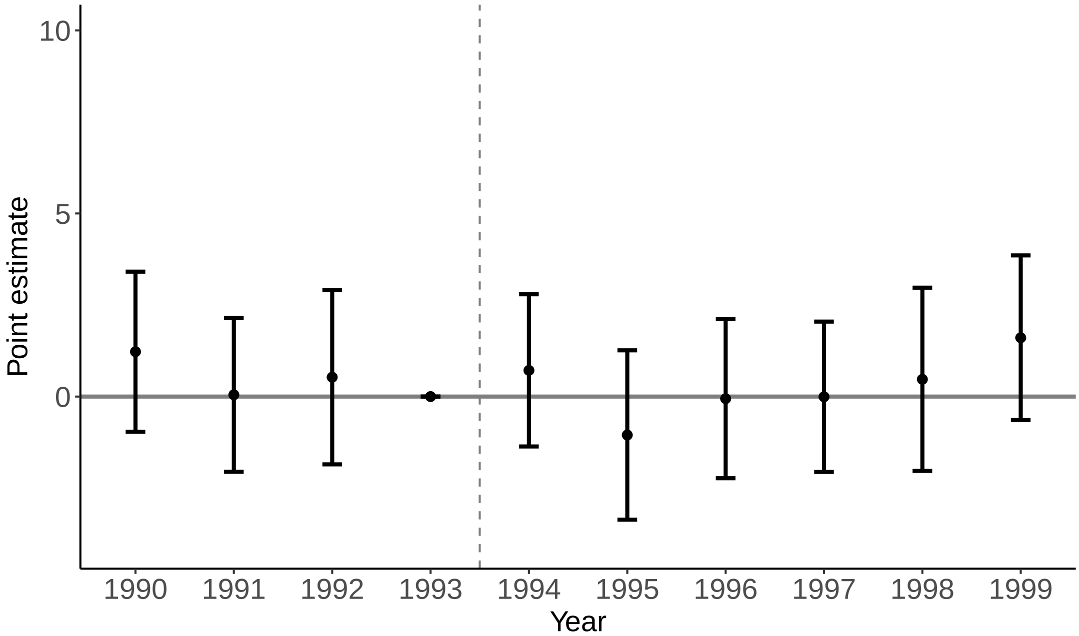
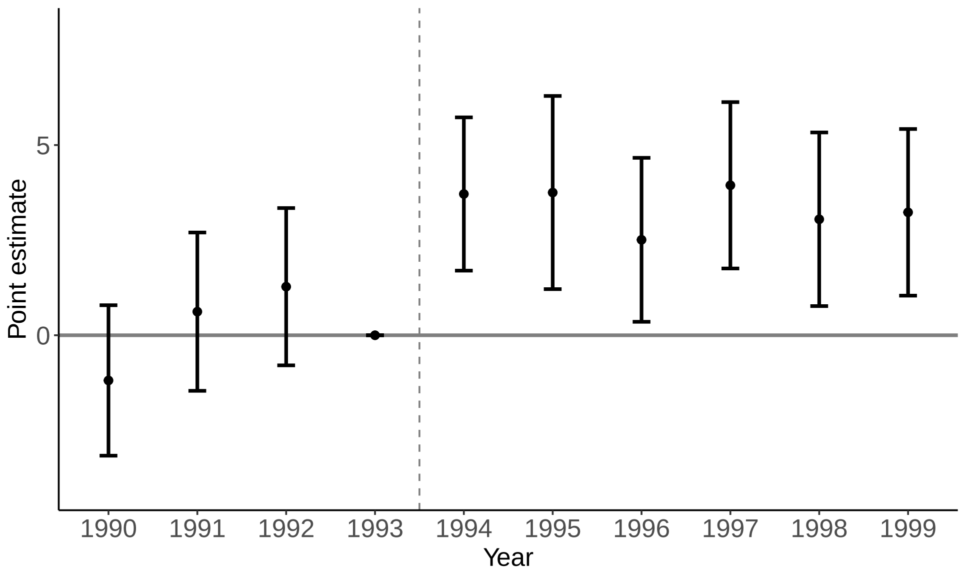
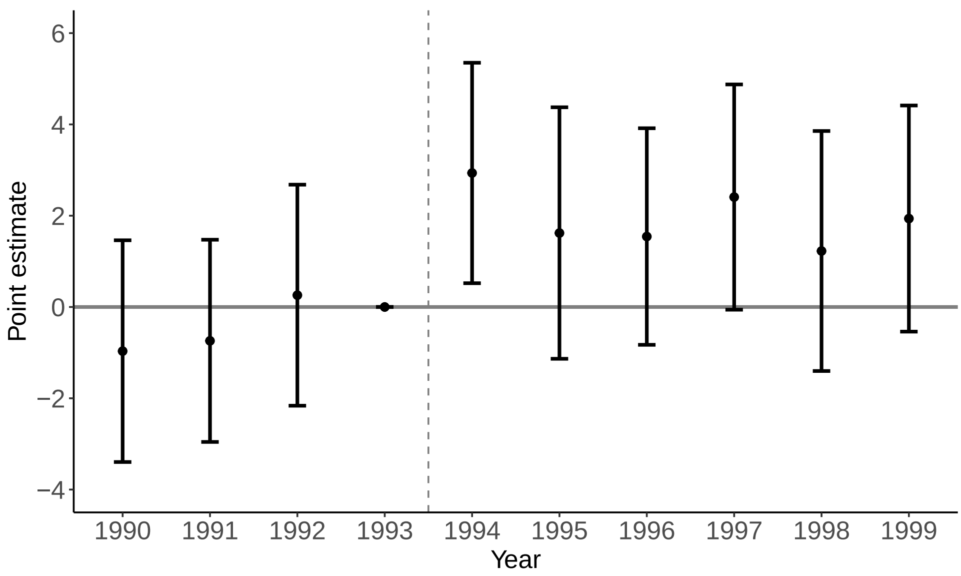



FIGURE B.6: Falsification test: optimal maize trading routes

*This figure shows event-study estimates based on equation 3.2 using predicted optimal maize-trading routes as placebo treatment indicators. Optimal maize-trading routes are predicted least cost paths using municipalities among the 90th percentile of maize production as origins instead of (known) drug-producing or drug-trafficking regions. The unit of observation is a municipality. Controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors are clustered at the municipality level. 95% confidence band shown.*




(A) Route municipalities




(B) Neighbor municipalities

Figure B.7: Event study of spillover effects

*These figures show event-study estimates based on equation 3.2 adding an indicator for being a neighbor of a trafficking municipality. Panel A shows the dynamic effect for municipalities that are traversed by a route; Panel B shows the dynamic effect for neighbor municipalities from the same regression. The unit of observation is a municipality. Controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors are clustered at the municipality level. 95% confidence band shown.*

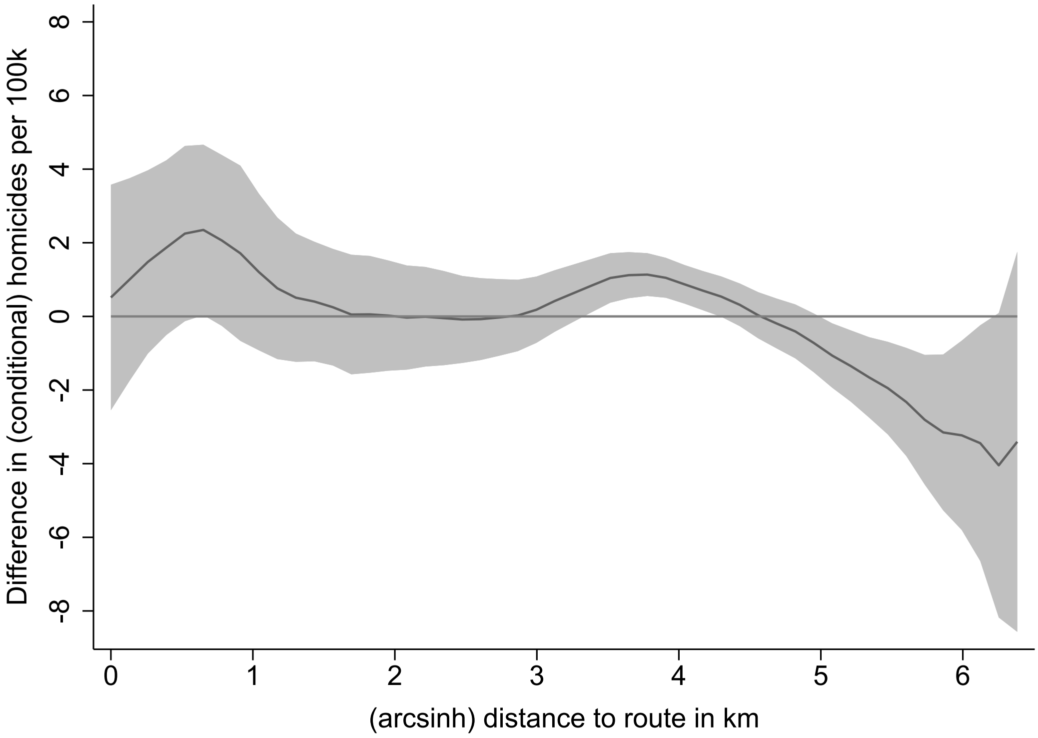



Figure B.8: The displacement of violence

*This figure shows results from local polynomial regressions of the residual change in drug-related homicides of males (15–39) between the periods 1990–93 and 1994–99 on the inverse hyperbolic since (arcsinh) transformed distance in km between a municipality's centroid and the closest predicted optimal drug-trafficking route. The residual change in homicides is calculated from the difference in residual homicides drawn from regressions of homicides on the baseline control variables (temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990), separately for the periods 1990–93 and 1994–99. 95% confidence band shown.*

Table B.3: Robustness to alternative transformations and sample restrictions

| Dep. var.: | Drug-related homicides per 100.000 inhabitants | | | | | |
|---|---|---|---|---|---|---|
| | Ln+1 | Arcsinh | Ln+q.90 | Origins | Drop top 1% | Drop top 5% |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Route x post NAFTA | 0.106*** | 0.119*** | 2.253*** | 1.883*** | 2.271*** | 2.500*** |
| | (0.038) | (0.045) | (0.588) | (0.594) | (0.591) | (0.602) |
| Origins x post NAFTA | | | | 3.231* | | |
| | | | | (1.821) | | |
| Municipality FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Baseline controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| *N* | 23,980 | 23,980 | 23,980 | 23,980 | 23,740 | 22,820 |

*Notes:* The table shows results from estimating equation 3.1. The unit of observation is a municipality. Column 1 transforms the dependents variable using the natural logarithm, adding 1; column 2 uses the inverse hyperbolic sine transformation; column 3 uses the natural logarithm, adding the value at the 90th percentile of the homicides distribution. Column 4 adds an indicator for municipalities that are identified as origins for the routes. Columns 5 and 6 drop municipalities in the top 1% and top 5% of the homicides distribution. Baseline controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Robust standard errors in parenthesis. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table B.4: Spillover effects for neighbor municipalities

| Dep. var.: | Drug-homicides per 100.000 inhabitants | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Route × post NAFTA | 3.189*** | | 4.076*** |
| | (0.675) | | (0.842) |
| Neighbor × post NAFTA | 2.308*** | 2.252*** | 2.289*** |
| | (0.555) | (0.539) | (0.555) |
| Length of routes × post NAFTA | | 1.576*** | |
| | | (0.395) | |
| Number of tributaries × post NAFTA | | | −0.189*** |
| | | | (0.055) |
| Municipality FE | ✓ | ✓ | ✓ |
| Time FE | ✓ | ✓ | ✓ |
| Baseline controls | ✓ | ✓ | ✓ |
| Mean homicides pre-Nafta | 8.22 | 8.22 | 8.22 |
| Observations | 23,980 | 23,980 | 23,980 |

*Notes:* The table shows results from estimating equation 3.1. The unit of observation is a municipality. *Neighbor* is defined as a municipality that shares a side or an edge with a municipality with a predicted drug-trafficking route. *Length of route* is standardized with zero mean and unit standard deviation. The variable *Number of tributaries* counts the number of tributaries that had flown into the route upstream from a traversed municipality. Baseline controls include temperature, precipitation, soil ph, maize suitability, optimal conditions for cannabis and opium production, distance to U.S. border, population size in 1990. Standard errors clustered at the municipality level in parenthesis. *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level.

# Chapter 4

# Violence and Ethnic Identity: Evidence from Perú

This chapter is based on Hidalgo (2022).[1]

## 4.1 Introduction

People's identity defines who they are and their behavior according to Akerlof and Kranton (2010, p. 13, 28). One of the most salient identities for human beings is their ethnic identity, where the economic literature has found a clear relationship between ethnic diversity and conflict.[2] Most of these studies focus on understanding how ethnic diversity affects violent conflicts. Yet, there is a lack of understanding in the economic literature on how violence affects ethnic identification.

In this paper, I intend to advance the understating of the consequences of violence on ethnic identity by using the Shining Path conflict as a natural experiment. At the beginning of the 1980s, Peru's shining path guerrilla started its revolutionary activity by settling in rural areas and pressuring the population to embrace class identity. In this conflict, more than 70.000 people were killed, and many more suffered torture, rape, and displacement. The Indigenous population was the most affected ethnic group representing 75% of the victims of this conflict while only representing 30% of the total population.

This setting provides two advantages for studying the consequences of violence on ethnic identity. First, the Shining Path conflict did not target Indigenous people per se, but the Indigenous people were their main victims. The Shining path ideology emphasized that any other identity but the class identity was alienation from capitalism and enforced the abandonment of those other identities (Jima-González and Paradela-López, 2020). In this context, all people, including the Indigenous people, had the decision to abandon their

[1]I thank Gustav Agneman, Miriam Artiles, Aldo Elizalde, Francisco Eslava, Eric Melander, Alberto Hidalgo, Erik Hornung, Juan Pablo Posada, Susanne Prantl, Carola Stapper for helpful comments.

[2]see: Alesina, Baqir, and Easterly (1999), Arbatlı et al. (2020), Esteban, Mayoral, and Ray (2012), Montalvo and Reynal-Querol (2005), and Rohner, Thoenig, and Zilibotti (2013a,b)

identity to embrace the class identity or suffer potentially violent consequences.[3] Second, the Shining Path conflict lasted from 1982 to 1992 and affected approximately one-third of Peruvian districts. Therefore, this conflict provides a large source of variation at the time and spatial level, which allows me to exploit differences in exposure to violence between cohorts of individuals over a ten year period across one-third of Peruvian districts.

To study the consequences of violence on ethnic identity, I combine individual-level data on ethnic identity and violent event-level data from 1958 until 1992. I collect individual-level data on ethnic identitiy from the *Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza* (ENAHO). I measure ethnic identity in two ways. First, based on self-identification, an individual is classified as Indigenous if individual $i$ identifies itself as Quechua, Aymara or another Indigenous group, and I classify it as non-Indigenous if individual $i$ identifies as Mestizo. Second, based on mother tongue, an Individual is classified as Indigenous if their mother tongue is Quechua, Aymara or another Indigenous language, and they are classified as non-Indigenous if Spanish (Castellano) is their mother tongue.

I collect data on violent events from the Peruvian Truth and Reconciliation Commission (TRC) database. This dataset is the most comprehensive dataset on the Shining Path conflict. It provides event-level data on the location, the time, the type of violence, and the ethnic identity of the victim of the violent event for more than 36,000 unique violent events from 1980 until 2000. I combine the TRC dataset with the ENAHO dataset to create an exposure to the violence variable at the individual level. The idea of the identification strategy is to compare ethnic identification differences of individuals that were exposed and not exposed to violence during the years identity is formed (0-19) with respect to ethnic identification differences of individuals that were exposed and not exposed to violence when identity was already formed (>19). Last, since I am interested in assessing the impact of conflict on Indigenous identity, I only consider violent events where an Indigenous person is a victim.

The main concern of the identification strategy is that violent events did not occur randomly, and the Shining Path could have targeted some specific districts based on the identity characteristic of the population. To address this issue, I use a border sample design that allows me to select a set of districts that are equal in socio-economic characteristics but suffered different levels of violence.

I implement a difference-in-differences strategy that exploits quasi-random variation levels of exposure to violence across individuals using the bordering sample districts. This allows me to compare ethnic identification differences of individuals exposed and not exposed to violence during the ethnic identity formation years (0-19) compared to individuals

[3] *"First they [rural population] will be given political reeducation, and if that fails, they will receive what we get now: dictatorship, prison ... death. The genocide which we have accepted, they will have to accept"* (Wheat, 1990, p. 48).

exposed and not exposed to violence when ethnic identity was already formed (>19). To further understand the consequences of violence on identity, I use an cohort-level analysis and compare the probability of individuals identifying as Indigenous conditionally on the year-cohort ∈ (0-4, 5-9, 10-14, 15-19) that they were exposed to violence compared to the cohort (20-24) when identity was supposed to be formed at the time of violence.

The estimates show that individuals exposed to violence are 2.3 percentage points less likely to identify as Indigenous and 3.4 percentage points less likely to speak an Indigenous language as their mother tongue than comparable individuals that were not exposed to violence. These results are meaningful in size since a 2.3 (3.4) percentage points decrease in the probability of identifying as Indigenous (speak an indigenous language as mother tongue) represents a 5.75% (13.6%) decrease with respect to the mean (40%). Furthermore, the results of the cohort-level analysis show that individuals exposed to violence in their childhood were the most affected. The estimates suggest that individuals who experienced violence between 0 and 4 years old are 7.8 percentage points less likely to speak an indigenous language as their mother tongue and 4.7 percentage points less likely to identify Indigenous than individuals who did not experienced violence in similar cohorts. Similarly, individuals who experienced violence between 5 and 9 years old are 5.3 percentage points less likely to speak an indigenous language as their mother tongue and 3.2 percentage points less likely to identify Indigenous than individuals who did not experienced violence in similar cohorts. I do not find statistical differences in the probability of speaking an Indigenous language as a mother tongue or identifying as Indigenous between the 10-14 and 15-19 cohorts and the 20-24 cohort.

The validity of the identification strategy relies on the assumption that in the absence of violence, indigenous identification would have followed common trends between cohorts of individuals exposed and not exposed to violence. To provide evidence for this assumption, I compare Indigenous identification differences between individuals that were exposed to violence and not exposed to violence when their ethnic identity was already formed (cohorts 25-29, and 30-34) with respect to our baseline cohort (20-24). I find no statistical differences in ethnic identification between these cohorts. Further, I compare Indigenous identification of individuals born in districts that experienced violence after violence occurred and the baseline cohort (20-24). The results show no difference between these two groups in their likelihood of identifying as Indigenous or speaking an Indigenous language as their mother tongue.

I conduct additional robustness checks to validate the main results. First, I test whether exposure to non-Indigenous violence (when the victim is not an Indigenous person) impacts ethnic identification. The estimates show that exposure to non-Indigenous violence has no consequences on the probability of an individual identifying as Indigenous. Second, I check whether migration is confounding the main results. When controlling for migration, the results remain similar, and when using migration as an outcome, I do not find a significant effect of exposure to violence on migration. Third, I check whether exposure to violence had a consequence on class and national identitiy. I find no effect on class-identity and a negative association between exposure to violence and a national identity feeling. Last, I test whether exposure to violence affected the different Indigenous groups differently. The results show that violence mostly affected Indigenous identification of the Quechua population and Indigenous groups that were not Aymara.

To understand the mechanism driving the main results, I build on the recent advances in understanding the dynamics of violent conflicts, trust, and ethnic identity. Rohner, Thoenig, and Zilibotti (2013a,b) argue that exposure to violence in a conflict erodes generalized trust in a society and enhances interpersonal trust and ethnic identification. The main difference between Rohner, Thoenig, and Zilibotti (2013a,b), and the Peruvian case is that in the case of Rohner, Thoenig, and Zilibotti (2013a,b) violence was perpetuated between different ethnic groups, and in the Shining Path conflict in Peru, it was perpetuated locally by people living in the same village. Therefore, I expect the conflict to erode interpersonal trust and Indigenous identification in Indigenous homogeneous regions.

To test the consequences of exposure to violence on trust, I collect data on interpersonal trust from the LAPOP surveys, generalized trust from the Latinobarometro surveys, and institutional trust from the ENAHO surveys. The results show that individuals exposed to violence have lower interpersonal trust, and (weakly) higher generalized trust than comparable individuals that were not exposed to violence. Further, the results show that trust in different state institutions is not different between individuals that were exposed to violence and individuals that were not exposed to violence. I hypothesize that these results are opposite to Rohner, Thoenig, and Zilibotti (2013a,b) because individuals were exposed to violence by their own Indigenous group.

To test whether own-Indigenous group violence is the mechanism driving the main results, I construct a measure of Indigenous diversity at the district level using ethnic identity data of individuals born before 1950 following Alesina, Baqir, and Easterly (1999). This index measures the probability that two randomly drawn people from the same district belong to different Indigenous groups. The rationale for using these measures is that in more Indigenous homogeneous districts, the likelihood of suffering violence by your own ethnic group is higher, leading to a lower Indigenous identification. To test this hypothesis, I interact the Indigenous diversity index with the exposure to violence variable. The results show that in Indigenous homogeneous districts, Individuals exposed to violence are less likely to speak an indigenous language as a mother tongue and to identify as Indigenous than comparable individuals not exposed to violence. Furthermore, a 1sd increase in ethnic diversity index increases the probability of speaking an indigenous language by 5.8 percentage points and increases the probability of identifying as Indigenous by 1.1

percentage point.

These results complement the advances from Rohner, Thoenig, and Zilibotti (2013a,b) in our understanding of the consequences of violence on ethnic identity. The result suggests that violence has heterogeneous effects on ethnic identity depending on whether violence was perpetrated by a different ethnic group or by your own ethnic group. If a different ethnic group committed violence, we could expect an increase in ethnic identity. However, if your ethnic group committed violence, we expect a decrease in ethnic identity

This paper contributes to three branches of the economic literature. First, it contributes to the growing literature on economics and identity.[4] This literature provides theoretical and empirical evidence about how identity, or a person's sense of self, affects economic outcomes (Akerlof and Kranton, 2000). Chen and Li (2009) provide evidence that when individuals share identity, they are more prone to within-group charity and lower envy. Furthermore, individuals are more likely to choose social-welfare-maximizing actions when sharing a group's identity. Xin Li (2010) shows that ethnic and national identities shape the tax morale of a country. The author finds that ethnically fractionalized countries have poorer tax morale than ethnically homogenous countries. In a lab experiment, Hargreaves Heap and Zizzo (2009) measure the social value of groups. The authors find that trust falls due to negative discrimination against outsiders when including outsiders in a group.

However, the economic determinants of ethnic identity have remained largely unexplored. Clots-Figueras and Masella (2013) provide evidence of education policies on shaping ethnic identity. Using the change of education policy by introducing Catalan along side with Spanish in Catolonia, the authors show that individuals that were exposed to Catalan have stronger Catalan feelings. Battu and Zenou (2010) provide evidence that the social environment and attachment to culture of origins are determinants of ethnic identity. Additionally, the study finds that ethnic identity has an impact on labour market success. Aspachs-Bracons et al. (2008) provide evidence that mandatory language policies have an effect on ethnic identity, whereas policies in which parents can choose the language have no impact on ethnic identity. I contribute to this literature by providing empirical evidence that exposure to violence has a negative association with Indigenous identity.

This paper contributes to the established literature linking ethnicity and conflict.[5] This literature shows, theoretically and empirically, how different measures of ethnic diversity correlate with civil conflict. For example, seminal work by Alesina, Baqir, and

[4]see: Akerlof and Kranton, 2000, 2002; Aspachs-Bracons et al., 2008; Bénabou and Tirole, 2011; Chen and Li, 2009; Clots-Figueras and Masella, 2013; Constant and Zimmermann, 2008; Hargreaves Heap and Zizzo, 2009; Kranton, 2016; Schüller, 2015; Xin Li, 2010

[5]see: Collier and Hoeffler (2004b), Collier, Hoeffler, and Rohner (2009); Montalvo and Reynal-Querol (2005); Esteban, Mayoral, and Ray (2012), Esteban and Ray (2011), Rohner (2011), Caselli and Coleman (2013)

Easterly (1999) argues that the underprovision of public goods is related to ethnic heterogeneity, and ethnic conflict is an important determinant of public finances. Montalvo and Reynal-Querol (2005) provide empirical evidence that ethnic polarization is an important determinant of civil wars, and La Porta et al. (1999) provide evidence that ethnic heterogeneity is correlated with low efficiency in governments and corruption. In a series of articles Esteban, Mayoral, and Ray (2012) and Esteban and Ray (2011) study the link between ethnicity and conflict. First, the authors develop a theoretical model to understand the link between ethnic polarization and ethnic fractionalization with conflict, and second, they provide empirical evidence for their theoretical predictions. They show that ethnic polarization, fractionalization, and the Gini-Greenberg index are significantly correlated with conflict.

However, our understanding of the effect of violent conflict on ethnic identity is still limited. Rohner, Thoenig, and Zilibotti (2013b) develop a theory of persistent ethnic conflicts. The authors argue that conflict lead to a collapse of trust, fuelling more ethnic conflict. Rohner, Thoenig, and Zilibotti (2013a) provide empirical evidence for Rohner, Thoenig, and Zilibotti (2013b) theoretical predictions. In Rohner, Thoenig, and Zilibotti (2013a), the authors study the impact of social conflict on trust and ethnic identity. They find that more intense conflict is associated with lower generalized trust and higher ethnic identification. This paper complements this literature by providing empirical evidence regarding the heterogeneous effects of exposure to violence on ethnic identity depending on whether violence was perpetrated by a different ethnic group or by your own ethnic group.

Last, I contribute to the growing literature studying the causes and consequences of the Shining Path guerrilla. Albertus (2020) studies the determinants for the Shining Path uprising. The author provides evidence that greater land reform dampened Shining Path conflict events. García-Ponce (2017) investigates the association between Shining Path insurgency and women's engagement in local politics. The author provides evidence that gender quotas in Peru were more successful in municipalities exposed to the uprising. Galdo (2013) studies the consequences of Shining path conflict on labor market outcomes. The study finds that exposure to violence during the first 36 months of life is most detrimental to future labor-market outcomes. The author finds that a 1 standard deviation increase in the level of violence exposure during the first 36 months of life is associated with 5% lower monthly earnings, a 3.5% lower probability of working in the formal sector and a 6% lower probability of working in large firms. Leon (2012) studies the consequences of the Shining Path conflict on human capital accumulation. The author finds that individuals exposed to conflict accumulate less years of education, and that these results are stronger in the short-term than in the long-run. I contribute to this literature by proving evidence that the Shining path conflict not only affected economic outcomes but changed individuals' identities.

## 4.2 Background

### 4.2.1 The shining path conflict

On May 17th, 1982, in the locality of Chuschi in the department of Ayacucho, a group of 5 hooded men broke into premises where the electoral boxes and ballots were kept and burnt them. This event was claimed by the Shining Path and marked the beginning of the civil conflict in Peru. The confrontation between the Shining Path (Sendero Luminoso or Sendero) guerrilla and the Peruvian state was one of the most brutal civil conflicts endured throughout Latin America during the 20th century. During this conflict, 70,000 people were estimated to be killed, and many more suffered torture, rape, and displacement. The conflict mostly affected the poor rural population, especially the indigenous population (CVR, 2003).

The Shining Path organization results from an ideological conflict within the Peruvian Communist Party (PCP). Abimael Guzman, the leader of the Shining Path organization, advocated for a violent revolution to achieve socialism in contrast to the majority of leaders of the PCP. Guzman saw the Maoist *"Cultural Revolution"* as the most remarkable political event humanity has ever seen (CVR, 2003, p. 15). The division between the Shining Path and PCP occurred in 1970, and Abimael Guzman lost the ideological fight. At the time of the rupture, the Shining Path had 12 members in Ayacucho and 50 in the rest of the country (CVR, 2003, p. 17).

During the 1970s, Guzman's organization was primarily based in the department of Ayacucho, where he was a professor at the University of Humanga. In this decade, the organization members were university professors, university students and rural teachers, who held clandestine meetings to re-construct the PCP with Guzman's ideas. Following other communist parties, they constructed the Shining Path from top to bottom and from ideology to politics. Furthermore, this group of professors and students would be at the top of the hierarchy that was supposed to bring the peasantry to the revolutionary success (CVR, 2003, p. 17 -19).

Based on their study and learnings from Marxism, the main Shining Path leaders started the diffusion of their ideology. They started the diffusion of their ideology first to university students who would eventually graduate as teachers and would be able to diffuse the Shining Path ideology through secondary schools. In this way, the Shining Path created a small but ideologically cohesive groups dependent on the party. The idea was that once the armed conflict started, the peasantry would join the Shining Path guerrilla under the party's direction.

May 17th, 1982, marked the beginning of the armed conflict. In this year, the Shining Path was present in the rural areas of Ayacucho, and its discourse of social justice and equality was accepted primarily by poor farmers. From 1983, the Shining Path started a more coercive campaign to conquer the peasantry. This increased the death toll of people that were against the revolution, community leaders and "richer" peasants that the Shining Path identified as enemies of the people. (CVR, 2003, p. 43-46).

The Shining Path rhetoric was caught on the young population, mainly individuals between 12 and 30 years old. This rhetoric was appealing because of the promise of a "new world order" that would bring justice and social equality. These new ideas were an earthquake to the Andean cultural traditions, especially for the Indigenous people, because in their traditional structures, power was held by the older population, and all the population respected this. This "new order" changed the hierarchical structures by giving the power to children and young adults, while the rest of the population lived in fear of losing their lives if they did not agree with this (CVR, 2003, p. 47).

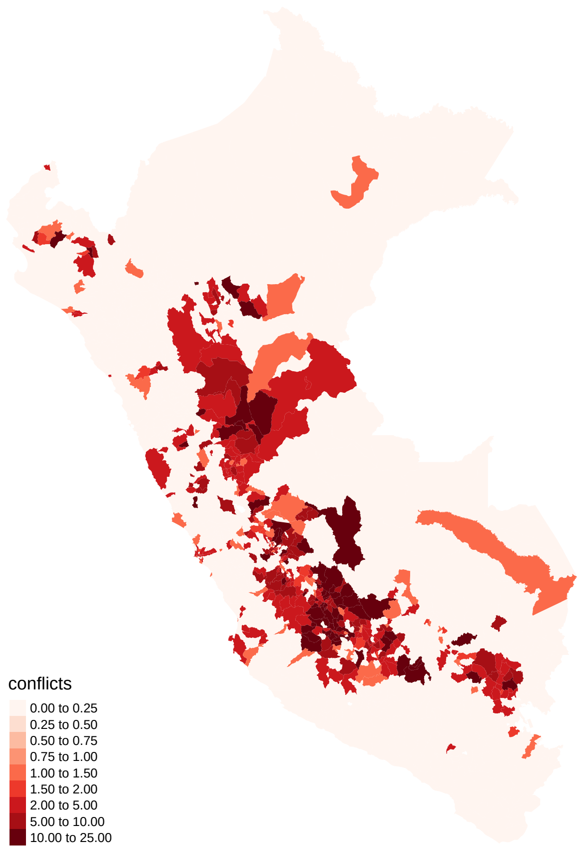



FIGURE 4.1: Violent events: 1980-2000

*Notes:* This figure shows the average number of violent events between 1980 to 2000 at the district level in Peru. Violent events includes murders, disappearances, illegal detentions, kidnappings, forced recruitment, torture, rape, and woundings. The data was collected from Peru 's Truth and Reconciliation Commission.

According to the CVR (2003), the recruitment of guerrilla fighters was through coercion, violence, and deception. Each community or family had a quota of children that had to be given to the fight, and families that opposed to give their children were killed to later enroll their kids in the guerrilla. The disruption of life of the peasantry and the constant fear of being kidnapped or killed caused discomfort in the rural population, who organized in peasant patrols to fight the Shining Path, which led to an increase of violence after 1983.

Even though the Shining Path conflict was one of the bloodiest in Latin America, the Shining Path considered themselves an "armed group without arms". Their leaders argue that their best arm was their ideology and that the organization did not depend on arms. The Shining Path leaders expected each militant to acquire arms by their own means. Hence, most were acquired by assaulting defenceless policemen or by stealing dynamite from mines.

The conflict extended until 1992 when state forces captured Abimael Guzman. By 1992, the conflict had extended to one-third of all Peruvian municipalities, covering 75% of all departments. Figure 4.1 shows the spatial distribution of conflict events at the district level in Peru.[6] The figure shows that violence was spread through Peru but was mostly prevalent in the Andean rural region. After the capture of Guzman, violence declined and by 1995 had stopped in most of the regions.

### 4.2.2 The shining path and Indigenous people

To understand the role of indigenous people and ethnic identity during the Shining Path insurgency, it is important to understand the Shining Path ideology developed by its leader Abimael Guzman. This ideology was called "Pensamiento Gonzalo", and was an application of communist ideas from Peruvian socialists, like Mairátegui and Maoist ideas into the Peruvian context (Jima-González and Paradela-López, 2020, 2021).

The foundation documents of the Shining Path state that the Shining Path is a "Marxist-Leninist-Maoist" political organization (Shining Path, 1988). Mostly aligned with Mao's ideology, which is relevant to the Peruvian context. According to Mao's ideology, the revolutionary subject is the peasant in contrast to Marxist-Leninist theory, who sees the proletariat as the revolutionary subject. Furthermore, Mao's theory identifies rural areas as the starting point of the revolution and not industrial cities as Marx suggested. Last, due to the anti-revisionism paranoia, high levels of violence and repression were legitimized, specially to those that rejected joining the movement (Jima-González and Paradela-López, 2020).

The Shining Path leaders saw the indigenous people as the "soldiers of the revolution". According to the Truth and Reconciliation Commission (CVR, 2003) the Shining Path

[6] conflict events include: killings, sexual assaults, kidnapping and torture.

leadership exhibited a sense of superiority toward the Indigenous people. For example, in an article by the Peruvian Communist Party [Shining Path] in 1976 they stated: *"We need to build the Party [and revolution] in the countryside, we need to mobilise, organise, and arm the peasants [...] under the direction of the [Communist] Party."* Furthermore, they added: *"The war belongs to the peasants under the leadership of the proletariat [The Communist Party]. The peasantry, because it constitutes the base of the national democratic revolution, is also the base of the people's war. This is a peasant war, or it is nothing."* (PCP, 1976). The Shining Path saw the Indigenous, who compose the majority of the rural population, as soldiers that would allow them to reach their political objectives.

The Shining Path understood that the peasantry and not the Indigenous were the revolutionary subject. Hence, Guzman's ideology, *Pensamiento Gonzalo* included the re-education of the Indigenous people because they were not aware of their own alienation (Jima-González and Paradela-López, 2020). Therefore, the Shining Path emphasised the class identity (peasantry) over any other feeling or identity, such as ethnic or gender identity, because these types of identities were perceived as alienation from capitalism. Along these lines, the Shining Path enforced the abandonment of ethnic [Indigenous] identity to embrace the peasant identity (Jima-González and Paradela-López, 2020).

The Shining path never included any Indigenous values in their official documents. Further, Indigenous people were never part of Shining Path's leadership, which was headed by Guzman and a group of academics. According to Peru's Truth and Reconciliation Commission (CVR, 2003, p 129), the Shining Path never took into consideration the cultural differences and identities of the peasantry but considered the peasantry as a homogenous subject.

The use of violence was central in *Pensamiento Gonzalo*, first as a tool for convincing people to join the revolution and second as a way to repress people that did not join them (Jima-González and Paradela-López, 2020). For Guzman, maintaining the "puritiy of the revolution" justifies violence and repression. In a speech to teachers in the Department of Huamanga in 1974, Guzman expressed that: *"war is necessary for the realization of the revolution"* (Gúzman, 1974).

In this context, the Indigenous population had the choice to join the movement and abandon their Indigenous identity and practices to embrace the peasant-class identity or suffer the potentially violent consequences. Anecdotal evidence suggests that this idea was impregnated in the mid-level leaders in the Shining Path. For example, Wheat (1990) pointed out that a leader of the Shining Path organization expressed: *"First they [rural population] will be given political reeducation, and if that fails, they will receive what we get now: dictatorship, prison ... death. The genocide which we have accepted, they will have to accept"* (Wheat, 1990 p. 48).

Following these ideas, The Shining Path started an execution campain against individuals and organizations that refused to collaborate with them (Kent, 1993). Since their main areas of action were the rural areas of Peru, the Indigenous people was the most affected population. Jima-González and Paradela-López (2020) provides anecdotal evidence of part of the Indigenous communities affected by the Shining Path guerrilla, for example the assassination in 1982 of Alejandro Huamán, leader of the Uchuraccay, the killing of 55 Ashaninka peasants in 1993, and the killing of 80 peasants in Lucanamarca in 1983 because they were presumably collaborating with the Peruvian government. In words of Jima-González and Paradela-López (2020), these killings were: *"Justified by Pensamiento Gonzalo's idea that the "blood-quota" was necessary for the triumph of the revolution and warned of the need to prepare for the inevitable bloodbath"* (Jima-González and Paradela-López, 2020, p. 204).

#### 4.2.3 Identity Formation

There is an extensive literature discussing the consequences of [ethnic] identity (Akerlof and Kranton, 2010). However, a less common question in the economic literature is how identity is formed (Casey and Dustmann, 2010).

The concept of ethnicity and identity has been discussed in depth in the psychological literature. Tajfel and Turner (2004, p. 17) defines [social] *"identity"* as those aspects of an individual self-image that are derived from social categories (e.g. ethnicity, gender, class...) to which she perceives belonging herself. Holcomb-McCoy (2005) define *"ethnicity"* as a set of behaviours, languages, and worldview that are associated with cultural heritage. Hence, ethnic identity is a membership to a specific ethnic group. Furthermore, the psychology literature argues that (ethnic) identity is formed in three different stages (Erikson, 1968; Marcia et al., 1980; Phinney, 1989). First, is a phase where individuals do not explore or do not have a conscious of identity. Next comes a phase of exploration, where individuals actively question and choose among different alternatives. Last, individuals commit to an identity (Phinney, 1989, p. 38).

The (psychologyc) empirical literature suggests that ethnic identity is formed during childhood (0-9) and adolescence (10-19). Brand, Ruiz, and Padilla (1974) shows that ethnic awareness begins at the age of 4. Further, once attitudes towards identity are formed, they tend to increase with age. Davey and Mullin (1980) measure the ethnic identification and ethnic preferences of children aged 7-11. The authors find that children perceive the advantages of being white, and only few children are ethnocentric in the decision to make friends. Marks et al. (2007) study ethnic identification among children of different ages, from 6 to 12 years old. The authors find that older cohorts show greater ethnic identity exploration than younger cohorts. The majority of studies focus on the development of ethnic development during adolescence.[7] Seminal work by Phinney (1989) show that about half of the adolescent have not explored their ethnic identity, one quarter is exploring their identity and one quarter have achieved their identity. Yip (2014) finds that adolescents that achieved ethnic identity have a high level of identity salience across different situations.

Following the literature on the identity formation process, I hypothesize that if exposure to violence affects ethnic identity, it would affect individuals exposed to violence between the ages of 0 and 19, when identity is formed. I expect that exposure to violence do not affect ethnic identity formation of individuals exposed to violence after the age of 20.

The direction of the consequences of violence on ethnic identity is unclear. On the one hand, exposure to violence could decrease generalized trust, leading individuals to shelter in their ethnic group and reinforcing ethnic identity. This is relevant for inter-ethnic group violence (Rohner, Thoenig, and Zilibotti, 2013a,b). On the other hand, I hypothesize that exposure to intra-ethnic group violence could hamper individuals from self-identifying as their ethnic group following an opposite dynamic as suggested by Rohner, Thoenig, and Zilibotti (2013a,b).

## 4.3 Empirical framework

### 4.3.1 Data

**Dependent Variable.** The dependent variable is an individual's ethnic identity, more specifically whether an individual is Indigenous or not. I measure ethnic identity in two ways. First, I measure Indigenous identity based on individuals' self-identification whether an individual self-identify as Indigenous or not, and second based on their mother tongue.

I collect data at the individual level from the *Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza* (ENAHO) originally collected by *Instituto nacional de Estadística e Informática* (INEI). This survey collects individual-level data on the living conditions of people across all 24 Peruvian departments (first administrative unit). I use the survey waves from 2013 to 2017 because these waves include questions regarding individuals' indigenous identification and mother tongue. The surveys are organised in 29 different modules where the INEI collects information regarding the livelihood of Peruvians. For the analysis, I use the following modules: Module 1: house characteristics, (*Características de la Vivienda y del Hogar*); Module 2: charateristics of individuals (*Características de los Miembros del Hogar*); Module 3: Education (*Educación*); Module 4: Health (*Salud*); Module 5: Income and employment (*Empleo e Ingresos*); ands Module 85: governance, democracy, and transparency (*Gobernabilidad, Democracia y Transparencia*).

[7] for a review, see: Phinney (1990)

I collect the dependent variable from the education, and employment and income modules. In the employment & income module, we can find the question: *"According to your ancestors and your customs, do you consider yourself: Quechua, Aymara, Other Indigenous community, Black, White, Mestizo, or unknown?"* I classify an individual as Indigenous if the individual answered Quechua, Aymara and Other Indigenous community. I classify an individual as non-Indigenous if the Individual answer mestizo. The main reason I only use mestizo as a base is historical. Mestizo is a person with both Spanish and Indigenous ancestry, hence, if an Individual changes their ethnic identity, it would move from Indigenous to mestizo or mestizo to Indigenous rather than to white or black.

From the education module, I use the question: *"What is the language or mother tongue you learned in your childhood?"* The potential answers were: Quechua, Aymara, another Indigenous language, Spanish (Castellano), English, Portuguese, another foreign language, sign language or other. In this case, I classify an individual as Indigenous if the individual answered Quechua, Aymara, or another Indigenous language. If the individual answered Castellano, I would classify them as non-Indigenous. Similarly, as the previous case, Castellano (Spanish) is the most common language in Peru, hence, if an individual did not learn an Indigenous language, Castellano would be the most likely language to learn.

Figure C.1 in the appendix shows the share of individuals each cohort from 1950 to 1992 by ethnicity. In subfigure C.1a, I use mother tongue as a measure of ethnicity and in subfigure C.1b, I use the self-identification measure. The figure shows an increase of Non-Indigenous identification in younger cohorts. This could be a result of the violent consequences of the Shining Path conflict.

**Independent Variable.** I collect data on Peru's conflict from the Truth and Reconciliation Commission (TRC). This dataset provides data on violent events at the event level. It provides information about the location, time, type of violence and victim of the event from the beginning of the conflict in 1980 until 2000.

The TRC collected 36,019 unique reports of violent events from testimonies of human rights abuses through interviews and visits to communities affected by the conflict. The conflict events came from 16,917 testimonies and included murders, disappearances, illegal detentions, kidnappings, forced recruitment, torture, rape, and woundings.

To construct the independent variable, I combine the conflict event dataset from TRC and the individual-level dataset from ENAHO. Using the location and time of the event, I construct a panel dataset that counts the number of conflict events in each of the 1873 Peruvian districts from 1980 until 1992. Further, the ENAHO dataset provides information on individuals' district and year of birth (Module 2). I combine these two datasets and create the variable *exposure*, which takes the value of one if an individual was born in a district that suffered a violent event and takes the value of zero otherwise.

Furthermore, the dataset allows me to differentiate whether individuals were exposed to violence in their district of birth during their ethnic identity formation years (0-19) or when ethnic identity was already formed (>19). Last, since I am interested in assessing the impact of violence on Indigenous identity, I consider only violent events with an indigenous person as the victim. According to the TRC, approximately 75% of the victims of the conflict spoke Quechua or another Indigenous language as their first language (Laplante and Theidon, 2007).[8] Table C.1 in the appendix presents the summary statistics of all the variables used in the empirical analysis.

### 4.3.2 Sampling design

The main concern of the empirical strategy is that the Shining Path did not randomly select districts to perpetuate their violent attacks. Figure 4.1 shows the spatial distribution of conflict events in Peruvian districts from 1980 to 2000. The figure shows that most Shining Path attacks were concentrated in the Andean region of Peru and the Shining path was mostly present in rural areas. Even though Indigenous people were not the main target of the Shining Path, the Indigenous population were the main victims, and they represented approximately 75% of the total death of the conflict (Laplante and Theidon, 2007).

To investigate the effect of the shining path conflict on ethnic identity, the ideal experiment would be to randomly allocate shining path's terrorist attacks in Peruvian districts and compare whether individuals who were born before and after the attacks in districts with and without attacks self-identify ethnically different. However, this field experiment is not feasible. Hence, to provide evidence for the consequences of conflict on ethnic identity, I select a set of bordering municipalities that were, arguably, similar before the conflict started but suffered different violence intensity. A similar strategy was used by Lichter, Löffler, and Siegloch (2021), where the authors use a border discontinuity design to provide evidence for the long-run effects of government surveillance on civic capital and economic performance, using East Germany district borders.

The objective of this section is to find a set of municipalities that were equal in socio-economic characteristics before the Shining Path conflict started but suffered different levels of violence due to a random characteristic. The sampling design is based on Albertus (2020) study. The author studies the effect of land reform on conflict in Peru. Using a regression discontinuity design, the author finds that districts in core areas of land reform witnessed less conflict than the adjacent districts in peripheral areas where less land reform occurred.

[8] Note that the variable could suffer from measurement error, if an individual was born in a district that suffered a conflict event but the individual moved to a different district (false positive) or if an individual was born in a district that did not suffer a conflict but moved to a district that expose the individual to a conflict event (false negative).

Figure 4.2 shows the standardized mean differences coefficients of different variables between core and periphery districts. The mean comparison indicates no differences between core and periphery districts in the vote for Marxist political parties (coef 3), and there are no differences in the share of indigenous people (coef 4) before the conflict started. Further, the results show that core municipalities suffer a higher number of conflict attacks and a higher number of deaths in the conflict, coef (1) and (2), respectively. These results suggest that when using the bordering discontinuity sample, conflict events and deaths were not confounded by the voting to Marxist or the share of indigenous people. Last, I do not find statistical differences in geographical and socio-economic characteristics, such as population, the literacy rate in 1961 and 1972, the number of social movements prior to the conflict, the number of hectares of private land, the share of cultivated land in each district, the number of state employees and land reform. I find that core districts had a lower elevation and slope than periphery districts. The latter can be controlled with district-fixed effects.[9] Given these results, I use the bordering sample as the main sample in the anaylsis.

[9] In appendix C.4 see a discussion about the differences between the sample design in this paper and the RDD in Albertus (2020).

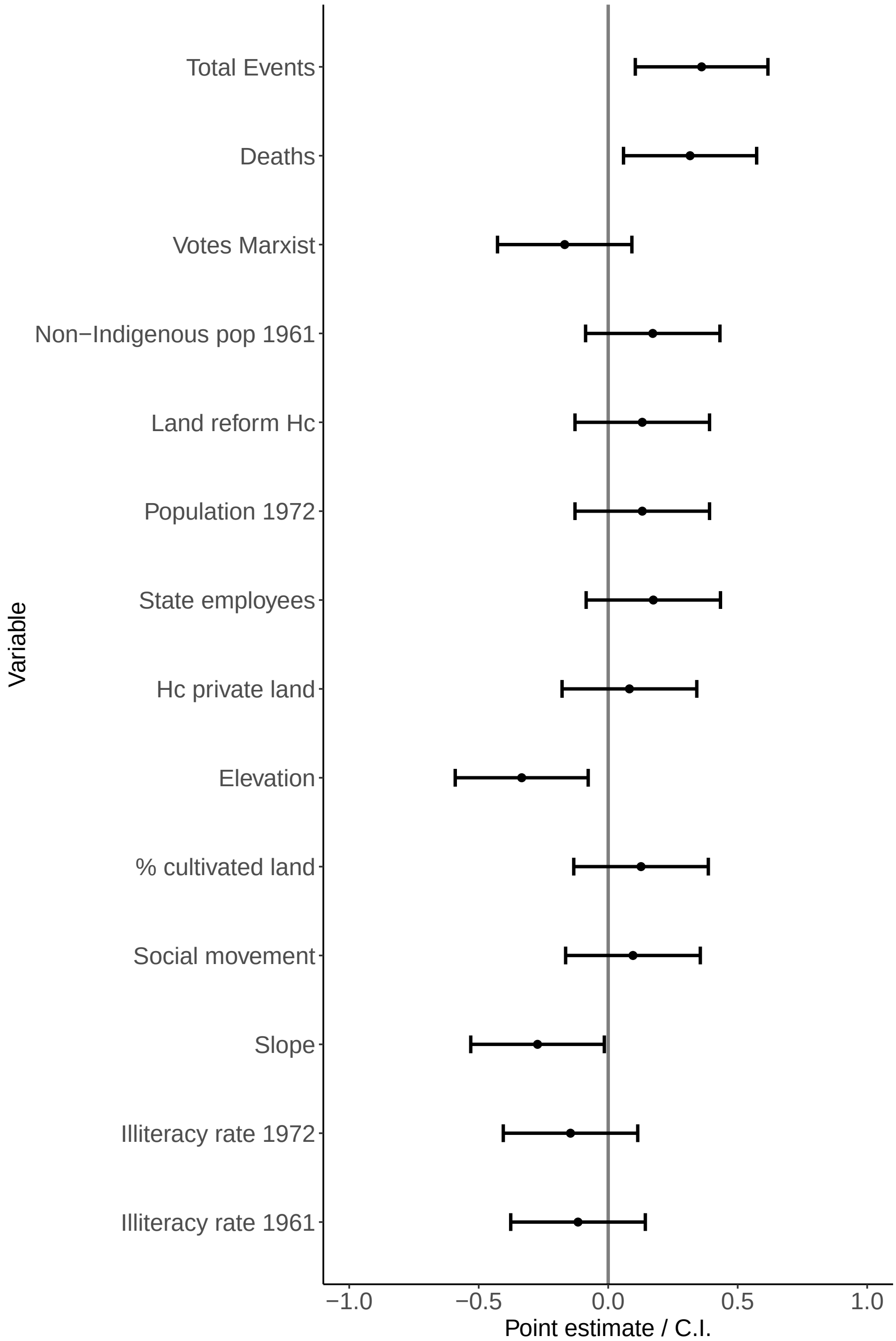



FIGURE 4.2: Mean comparison

*Notes:* The figures plots the standardized mean differences of different variables between core and periphery districts with 95% confidence intervals. The base category is core districts. The unit of observation is the districts. The variables are total number of violent events, total number of death, share of votes to Marxist parties , share of non-Indigenous population in 1961, illiteracy rate in 1961, , illiteracy rate in 1972, peasant movements, average slope, share of cultivated land, Hectares of private land, average elevation, number of state employees, population in 1972 (thousands), Hectares of land reform. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

### 4.3.3 Empirical strategy

This paper's objective is to study the effect of violence on ethnic identity. To capture the impact of the Shining Path violence on ethnic identification, I assume that the district and year of birth determine the exposure to violence of individuals. Furthermore, I use the border sample to select a set of municipalities which are similar in different socio-economic characteristics, except for the level of violence intensity, to provide quasi-random variation in the exposure to violence of individuals. Last, I compare ethnic identification differences between individuals who were exposed and not exposed to violence when they were between 0 and 19 years old with respect to the ethnic identification differences of individuals who were exposed and not exposed to violence when they were older than 19 years old. I implement this empirical strategy in a difference-in-differences regression framework applying the following specification:

$$Y_{ikdtp} = \beta Exposure_{ip} \times T_{ipk} + \gamma_k + \gamma_t + \gamma_d + \gamma_p + \delta c_i + \epsilon_{iktp} \quad (4.1)$$

Where $Y_{ikdtp}$ is a dummy that captures the ethnic identity $\in$(Indigenous or non-Indigenous) of individual $i$, born in year $k$, living in district $d$, interviewed in year $t$, and born in district $p$. I use two measures of ethnic identity, $Identify\ Indigenous$, which takes the value of 1 if the individual identifies as Indigenous and 0 if the individual identity as mestizo. $Mother\ tongue$, which takes the value of 1 if an individual speaks an indigenous language as his mother tongue and 0 if the individual speaks Castellano (Spanish) as his mother tongue.

$Exposure_i$ takes the value of 1 for individuals that were exposed to violence in their district of birth and 0 otherwise. $T_{idk}$ takes the value of 1 for individuals exposed to violence between the age of 0 and 19 and 0 otherwise. $\hat{\beta}$ is our coefficient of interest, which captures the difference in ethnic identification between individuals who were exposed and not exposed to violence when they were between 0 and 19 years old with respect to the ethnic identification differences of individuals who were exposed and not exposed to violence when they were older than 19 years old. $\gamma_k$ are year-of-birth fixed effects. $\gamma_t$ are year-at-interview fixed effects, $\gamma_d$ are district-at-interview fixed effects, and $\gamma_p$ are district-of-birth fixed effects. $\delta c_i$ are individual-level controls, such as age, age$^2$, and sex.

This econometric setting allows me to mitigate potential endogeneity concerns. Year-of-birth fixed effects allow me to compare individuals within the same year of birth and control for a possible decrease in indigenous identification over time. District fixed effects control for time-invariant factors such as geography. Year-at-interview fixed effects allow us to control for shocks that occurred during the year of the interview, for example, nationwide economic recession. District-of-birth FE allows us to compare individuals that were born within the same district. Furthermore, our border sample allow us to select municipalities that were similar in socio-economic characteristics, like the share of indigenous population, and political views, but suffered different levels of violence due to an exogenous border design. Our main identifying assumption is that, in the absence of violence, cross-cohort ethnic self-identification would have been similar between cohorts with different levels of violence exposure.

I use a more flexible difference-in-differences strategy to understand the consequences of violence on ethnic identity. In this specification, I compare how exposure to violence at different age cohorts impacts ethnic identity. Furthermore, this specification allows me to compare individuals' ethnic identification born after violence was perpetrated in their district of birth compare to the baseline cohort (20-24). To do so, I estimate the following regression equation:

$$Y_{ikdtp} = \sum_{c} \beta_c Exposure_{ip} \times D_{pc} + \gamma_k + \gamma_t + \gamma_d + \gamma_p + \delta c_i + \epsilon_{ikdtp} \tag{4.2}$$

Here, $D_c$ is an indicator whether individual $i$ experienced violence when belonging to cohort $c \in$ (post, 0-4, 5-9, 10-14, 15-19, 25-29, 30-34 ). Our baseline cohort is those individuals that were between 20 and 24 when violence occurred in municipality $p$. The rest of the variables remain the same as in our previous specification. The coefficients of interest, $\hat{\beta}_c$, capture cohort-specific treatment effects. More specifically, they capture the ethnic identification differences between individuals within each cohort $c$ that were exposed and not exposed to violence with respect to the baseline.

## 4.4 Results

Table 4.1 shows the results of our specification embedded in equation 4.1. The dependent variable in columns (1)-(2) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (3)-(4), the dependent variable is whether an individual identifies Indigenous or not. In columns (1) and (3), I use the full sample of individuals. In columns (2) and (4), I restrict the sample to the districts in the border sample. In all specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Additionally, I control for individual characteristics, such as age, age square, and sex. Last, I cluster the standard errors at the district-year level.

The results show a negative association between exposure to violence and individuals speaking an indigenous language as their mother tongue, as well as identifying as Indigenous. When using the full sample, individuals that were exposed to violence are 1.4 percentage points (pp. henceforth) less likely to speak an indigenous language than similar individuals that were exposed to violence during the identity formation years. Further, there is no statistical significant association between exposure to violence and self-identifying as Indigenous.

When using the border sample, the negative association between exposure to violence and Indigenous identity become larger in absolute terms. The results show that individuals exposed to violence are 3.4 pp. less likely to speak an indigenous language than individuals that did not experienced violence during their ethnic identity formation years. Similarly, individuals exposed to violence are 2.3 pp. less likely to identify Indigenous compared to individuals that were not exposed to violence. The effect is non-negligible in size, approximately 25% of individuals in the border sample speak an indigenous language as their mother tongue and 40% identify Indigenous. Hence, a 3.4 pp. (2.3 pp.) decrease in the likelihood of speaking an indigenous language (identifying as Indigenous) represents a 13.6% (5.75%) with respect to the mean (40%).

Table 4.1: Total

| | Indigenous language as mother tongue | | Identify Indigenous | |
|---|---|---|---|---|
| | all districts | border sample | all districts | border sample |
| | (1) | (2) | (3) | (4) |
| Exposure violence x T | $-0.014^{**}$ | $-0.034^{**}$ | $-0.005$ | $-0.023^{**}$ |
| | (0.006) | (0.016) | (0.005) | (0.009) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ |
| District-born FE | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 0.23 | 0.25 | 0.36 | 0.40 |
| SE Cluster | district-year | district-year | district-year | district-year |
| $N$ | 208,768 | 33,326 | 168,301 | 27,842 |
| Adjusted $R^2$ | 0.660 | 0.628 | 0.545 | 0.480 |

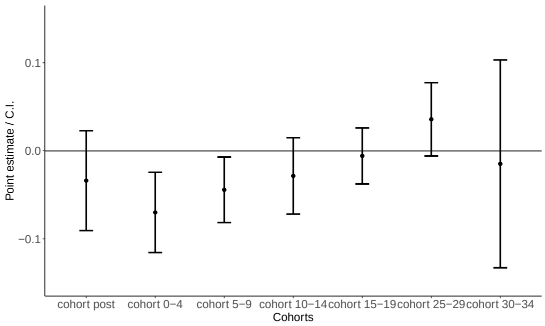
*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1)-(2) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (3)-(4), the dependent variable is whether an individual identifies Indigenous or not. In columns (1) and (3), I use the full sample of individuals. In columns (2) and (4), I restrict the sample to the districts in the border sample. In all specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

These results are robust to different specifications. Tables C.3 and C.2 in the appendix are replicas of table 4.1, but in table C.3 I remove the cohort that never was exposed to violence, and in table C.2 I cluster the standard errors at the district of birth.

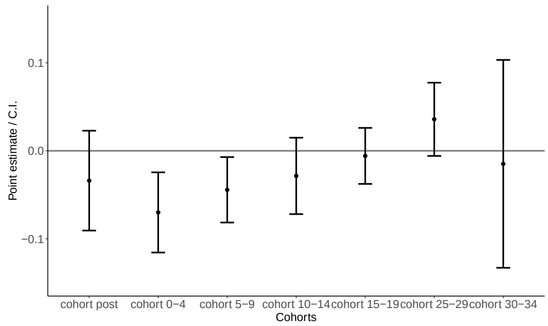
I perform cohort analysis to understand which cohorts were more affected by violence. Figure 4.3 plots the $\beta_c$ coefficients of equation 4.2. This figure shows differences in ethnic identification between individuals that were exposed to violence at different ages with respect to differences in ethnic identification of our baseline cohort (20-24 years old). Furthermore, this specification allows us to provide evidence regarding the assumption that Indigenous identification would be similar between cohorts in the absence of violence. To do so, I compare the ethnic identity of individuals exposed to violence, after the identity formation years, between the ages of 25-29 and 30-35 with our baseline (cohort 20-24). Further, I compare the ethnic identity of individuals born after a violent event occurred in a district (*post* cohort) with respect to the baseline cohort.[10] Table C.6 in the appendix shows the full set of results of this analysis.

The results show that ethnic identity is mostly affected when individuals are exposed to violence during childhood (0-9 years old). The estimates suggest that individuals exposed to violence between 0 and 4 years old are 7 pp. less likely to speak an Indigenous language as their mother tongue and 4 pp. less likely to identify Indigenous than individuals that were not exposed to violence. Furthermore, individuals that were exposed to violence between the age of 5 and 9 are 4.4 pp. less likely to speak an Indigenous language as their mother tongue, but the likelihood of identifying as Indigenous is not different from zero. Last, The results show no statistical differences in the probability of speaking an Indigenous language as a mother tongue or identifying as Indigenous among individuals that were exposed to violence between the age 10 and 19.

To provide evidence for the assumption that ethnic identification is not different between cohorts in the absence of violence. I compare ethnic identification differences between cohorts exposed to violence when their identity was already formed (cohorts 25-29, and 30-34) and the baseline cohort. The results show no statistical differences in Indigenous identification between cohorts 25-29, and 30-34, with respect to our baseline. Last, I find no differences in Indigenous identification between individuals born after violence was perpetrated (*post* cohort) in the district they were born and our baseline cohort.




(A) Mother tongue




(B) Indigenous identification

FIGURE 4.3: Cohort-level analysis

*Notes:* The figure plots $\beta_{\tau}$ coefficients estimated from equation 4.2 with 95% confidence intervals. The omitted cohort is 20-24. The explanatory variable is the interaction term between exposure to violence and exposure to violence after 19 years old. In subfigure 4.3a the dependent variable is is whether an individual speaks an Indigenous language as their mother tongue or not. In subfigure 4.3b is the dependent variable is whether an individual identifies Indigenous or not. I restrict the sample to the districts in the border sample. In both specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level.

[10]Post cohort example: an individual born in the year 1987 in district $i$ when the last violent event reported in district $i$ was in 1984.

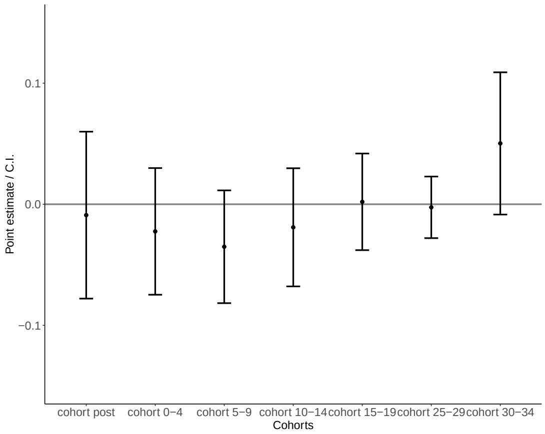
These results suggest that the Shining Path violence affected ethnic identity in Peru. However, the conflict did not affect all individuals equally. Individuals who were exposed to violence during childhood were less likely to speak an indigenous language as their mother tongue and less likely to identify as indigenous.

The attachment theory from the psychological literature could help understand why violence affected the development of an Indigenous identity during infancy and early childhood but not adolescence (Ainsworth, 1989; Bowlby, 1982). This theory argues that individuals during childhood and later in life have attachment figures for their emotional and social development, which includes the identity formation process (Parkes, Stevenson-Hinde, and Marris, 2006). During infancy and childhood, the main attachment figure is the mother. However, during adolescence, identity formation is not only influenced by parents but also by others in society (e.g. teachers in school) (Cama and Sehgal, 2021). The Shining Path ideology forced individuals to embrace a class identity abandoning all other identities, including ethnic identity. This could have led parents to stop transmitting their languages and culture to children fearing potential violent consequences. When violence occurs during adolescence, children could have already learned an Indigenous language during their childhood, mitigating the negative consequences of violence on Indigenous identity.

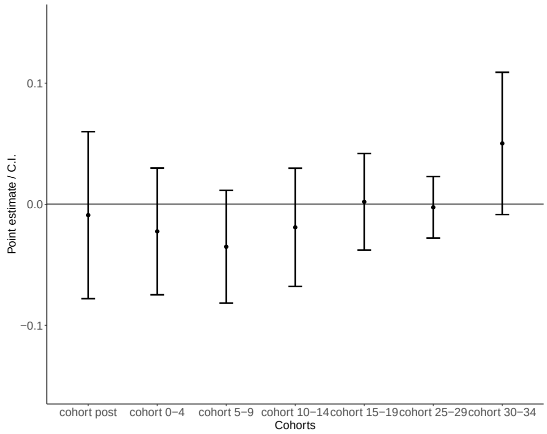
### 4.4.1 Robustness checks

#### 4.4.1.1 Non-Indigenous violence.

This paper shows that individuals exposed to violence are less likely to identify Indigenous and to speak an Indigenous language as their mother tongue. In the main analysis, I use only violent events where the victim was an Indigenous person (approximately 77.3% of the violent events).




(A) Mother tongue




(B) Indigenous identification

FIGURE 4.4: Non-Indigenous violence

*Notes:* The figure plots $\beta_\tau$ coefficients estimated from equation 4.2 with 95% confidence intervals. The omitted cohort is 20-24. The explanatory variable is the interaction term between exposure to non-Indigenous violence and exposure to non-Indigenous violence after 19 years old. In subfigure 4.4a the dependent variable is is whether an individual speaks an Indigenous language as their mother tongue or not. In subfigure 4.4b is the dependent variable is whether an individual identifies Indigenous or not. I restrict the sample to the districts in the border sample. In both specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level.

In this section, I use violent events where the victim was not an Indigenous person and test whether exposure to a violent event of a non-Indigenous person affects Indigenous identity. To do so, I use the cohort analysis and compare Indigenous identification differences between Individuals that were exposed and not exposed to non-Indigenous violence with respect to the baseline difference (20-24 cohort).

Figure 4.4 shows the coefficient of interest ($\beta_c$) of this event study analysis. Subfigure 4.4a shows that exposure to non-Indigenous violence has no effect on the probability of an individual speaking an Indigenous language as a mother tongue. Subfigure 4.4b shows that exposure to non-indigenous violence has no consequences on the likelihood of an individual identifying as Indigenous. These results suggest that violence mainly affects Indigenous identification when the victim is an indigenous person.

#### 4.4.1.2 Migration

The economic literature has shown a relationship between conflict and migration.[11] In the full sample, the data shows that individuals that were exposed to violence are 8.4 pp. more likely to migrate than individuals that were not exposed to violence (see column in 7 in table C.4 in the appendix). Hence, migration could be a potential confounder. The datset set allow us to control for the potential biases of migration.

[11]See: Guatemala (Morrison, 1993); Mexico (Basu and Pearlman, 2017); Colombia (Silva and Massey, 2015); Africa (De Vries and Guild, 2019)

In the main specification, individuals are exposed to violence if there was a violent event in their district of birth when they were between 0 and 19 years old. I do not differentiate between individuals that migrated and individuals that stayed in their district after a violent event. To further understand whether individuals that migrated and did not migrate have different ethnic identification after a violent event, I create a dummy variable, *migration*, which takes value 1 if the individual is living in a different district, at the time of the interview, than the one he was born and 0 if he lives in the same district as he was born. Furthermore, I use the main identification strategy embedded in equation 4.1 and include the *migration* in different ways.

Table C.4 in the appendix shows the results when including the *migration* dummy. In columns (1)-(3), the dependent variable is whether an individual speaks an Indigenous language as mother tongue. In columns (4)-(6), the dependent variable is whether an individual identifies Indigenous or not. Columns (1) and (4) are our baseline results similar to columns (2) and (4) in table 4.1. In columns (2) and (5), I include the migration dummy as a control variable. In columns (3) and (6), I interact the migration dummy with our measure of exposure to violence. In all regressions, I include district fixed effects, year-interview fixed effects, and year of birth fixed effects. Further, I include the individual controls, sex, age, and age square. Last, the standard errors are clustered at the district-year level.

The results show that when controlling for migration, the results are not statistically different from our baseline. The results show that individuals that were exposed to violence are 3.4 pp. less likely to speak and Indigenous and 2.3 pp. less likely to identify as indigenous than comparable individuals not exposed to violence. The interaction term results show that violence mostly affected individuals that were exposed to violence and did not migrate. The baseline comparison group is no-exposed to violence and no-migration, hence, each coefficient is compared to this group. The results show that individuals that were exposed to violence and did not migrate are 5.1 pp. less likely to speak an Indigenous language as their mother tongue and 3.6 pp. less likely to identify as Indigenous than Individuals that were not exposed to violence and did not migrate. Furthermore, the results show no statistical differences in ethnic identification between individuals that were not exposed to violence and migrated and individuals that were not exposed to violence and did not migrate. Last, I do not find a difference in the likelihood of speaking an indigenous language as a mother tongue or indigenous self-identification between individuals exposed to conflict and migrated and individuals not exposed to violence and not migrated.

To further understand the consequences of the Shining Path violence on migration, I use the *migration* dummy as an outcome using equation 4.1. Column (8) in table C.4 shows that the probability of migrating is not different for individuals exposed to violence before and after age 20. These results show that the migration was not a confounder driving the the main results.

#### 4.4.1.3 Other identities

This paper studies the consequences of violence on ethnic identity. As explained in section 4.2, the Shining path started its revolutionary activity and pressured the population to embrace a class identity. In this section, I study the consequences of violence on other types of identities, namely class identity and national identity.

To test the consequences of violence on class identity, I collect data on individuals' membership in different class-related organizations from the ENAHO surveys. To the best of my knowledge, there is no survey data on self-identification on class identity in Peru. Therefore, I use membership in workers and professionals associations, membership in political parties, and membership in a peasant community as a proxy for class identity. I expect that if class identity is salient in an individual, the likelihood of being part of one of these organizations is higher compared to an individual that does not emphasize class identity.

I use the specification in equation 4.1 and replace the dependent variable with a dummy variable indicating the probability of an individual belonging to a worker association, a professional association, a political party organization, and a peasant community. Table 4.2 show the results of this analysis. The estimates suggest that exposure to violence has no consequence on the probability of an individual participating in a workers' or professionals' association nor in the likelihood of being part of a political party organization, which suggest that the exposure to the Shining path violence had no consequences on class identity. Further, I find that exposure to violence decreases an individual's probability of being part of a peasant community. Note that the Shining Path ideology pressured the individuals to embrace a peasant identity. These results suggest that the exposure to their violence had the opposite effects. The likelihood of belonging to a peasant association is lower when an individual is exposed to the Shining Path's violent activity.

To test for the consequences of violence on National identity, I collect data on national identity from the Latinbarometro survey.[12] From this database, I use the question: *"How proud are you to be Peruvian?"* The respondents could answer: 1) very proud, 2) proud, 3) little proud, and 4) not proud. From this question, I create a dummy variable which takes the value of one if the respondent answered very proud or proud and the value of 0 if the respondent answered little proud or not proud. Further, I create a ordered variable, which takes values of 1, 2, 3 and 4 if the respondent answered not proud, little proud, proud, and very proud, respectively.

I combine the Latinbarometro survey and the TRC database and use the specification in equation 4.1 to understand the consequences of exposure to violence on National iden-

[12]In section 4.5.1, I discuss the potential shortcomings of using the Latinbarometro survey.

tity. Table 4.3 shows that individuals exposed to violence are less likely to be proud of their Peruvian identity than individuals that were no exposed to violence (columns 1 and 3). The results remain similar when controlling for indigenous identity (Columns 2 and 4). These results suggest that the Shining Path violence had a consequence on national identity, which could be attributed to the lack of state intervention, especially in the early years of the conflict.

Table 4.2: Class-identity

| | Workers | Professionals | Political Party | Peasant community |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Exposure violence x T | 0.008 | 0.005 | 0.002 | $-0.023^{**}$ |
| | (0.008) | (0.009) | (0.005) | (0.011) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ |
| District-born FE | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Mean DV | 0.037 | 0.37 | 0.011 | 0.22 |
| *N* | 13,868 | 13,868 | 13,868 | 13,868 |
| Adjusted $R^2$ | 0.084 | 0.062 | 0.002 | 0.641 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1),(2), (3), (4), are whether an individual belongs or not to a workers association, a professionals associations, a political party, and a peasant community, respectively. In all specifications, I include, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

#### 4.4.1.4 Ethnic group

Peru is a multiethnic society. In the main specification, I only compare Indigenous and mestizos. However, in Peru, there are more than 50 different Indigenous groups. In our dataset, we have Quechua, which represents 22% of the individuals in our sample, Aymara (0.6%) and other Indigenous populations (2.9%). In this section, I check whether exposure to violence has heterogeneous effects depending on the affected Indigenous group. To do so, I use the main specification but splitting the sample into Quechua, Aymara, and other Indigenous groups. Note that our baseline sample is always Castellano for the mother tongue language variable and mestizo for the self-identification variable.

Table C.5 presents heterogeneous effects of the association between exposure to violence and different Indigenous identity. The results show that individuals exposed to violence were less likely to speak Quechua or Other Indigenous languages as a mother

TABLE 4.3: National identity

| | National identity dummy | | National identity | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Exposure violence x T | $-0.057^{***}$ | $-0.052^{**}$ | $-0.099^{***}$ | $-0.098^{**}$ |
| | (0.021) | (0.020) | (0.036) | (0.038) |
| No-Indigenous | | $0.084^{***}$ | | $0.177^{***}$ |
| | | (0.027) | | (0.044) |
| District FE | ✓ | ✓ | ✓ | ✓ |
| Survey FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Mean DV | 0.83 | 0.83 | 3.38 | 3.38 |
| SE Cluster | district-year | district-year | district-year | district-year |
| *N* | 2,675 | 2,656 | 2,675 | 2,656 |
| Adjusted $R^2$ | 0.010 | 0.010 | 0.024 | 0.024 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1)-(2) is the ordered measure of national identity. In columns (3)-(4), the dependent variable is the dummy variable measure of national identity. In columns (2) and (4), I include a dummy variable indicating whether the individual is Indigenous or not. In all specifications, I include, year-birth fixed effects, year-survey fixed effects, district fixed effects. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

tongue, and less likely to identify as Quechua and Other indigenous groups than comparable individuals that were not exposed to violence in the identity formation years compared to differences between individuals exposed and not exposed to violence after the identitiy formation years. The results regarding the Aymara population are not statistically different from zero, due to the lack of statistical power. Hence, it is not possible to draw a conclusion on the consequence of violence with respect to this Indigenous group.

The majority of the Indigenous population in Peru is Quechua, representing around 22% of the total population, whereas the Aymara population represents less than 1%. I hypothesize that the negative association between exposure to violence for the Quechua population but not the Aymara population is own-Indigenous group violence. Since the Quechua population is the majority, the likelihood of experiencing own-group violence is higher than for the Aymara population, leading to a lower Indigenous identification. This mechanism is explained in the next section.

## 4.5 Dynamics of Conflict and Mechanism

The results show that those individuals exposed to the shining path violence in their identity formation years have a lower probability to identify indigenous and a lower probability of speaking an Indigenous language as a mother tongue. These results are particularly interesting in light of recent advances in understanding the dynamics of conflict and ethnic identity.

Rohner, Thoenig, and Zilibotti (2013b) propose a theoretical model to understand the dynamics of violent conflict, trust and ethnic identity. The authors suggest that civil wars are persistent over time, especially in multi-ethnic societies. The theory suggests that conflict and violence erodes trust in a society, leading individuals to identify more with their own ethnic group and increasing the probability of future conflicts, plunging the society into a vicious cycle, which the authors describe as a *"war trap"*. To provide evidence for their theoretical predictions, Rohner, Thoenig, and Zilibotti (2013a) use the civil conflict in Uganda between 2002 and 2005. The findings show that in regions where fighting was more intense, individuals decreased generalize trust and increased ethnic identity and interpersonal trust.

In Peru, we do not see a persistent *"war trap"* or persistent violent events towards identity as suggested by Rohner, Thoenig, and Zilibotti (2013b). Figure 4.5 shows that the Peruvian conflict was short and not persistent. It started to spike in 1983 until 1985. Between 1986 and 1987 there was a period of low conflict intensity, and it rose again from 1988 until Abimael Guzman was captured in 1992. After 1995, the conflict was negligible and not persistent. The main difference between the Peruvian civil conflict and the theoretical predictions by Rohner, Thoenig, and Zilibotti (2013b) are the perpetrators of the violent attacks. In the Peruvian case, the perpetrators were individuals from the same village as the victims, whereas in Rohner, Thoenig, and Zilibotti (2013b), they were from different ethnic groups. Therefore, I hypothesize that the mechanism driving the negative relationship between exposure to violence and Indigenous identity is own-ethnic group violence.

To understand the dynamics of violent conflict on trust and ethnic identification, I follow Rohner, Thoenig, and Zilibotti (2013a,b) and study the consequences of the Shining Path violence on generalized and interpersonal trust. Furthermore, I study the consequences of own group violence on ethnic identification.

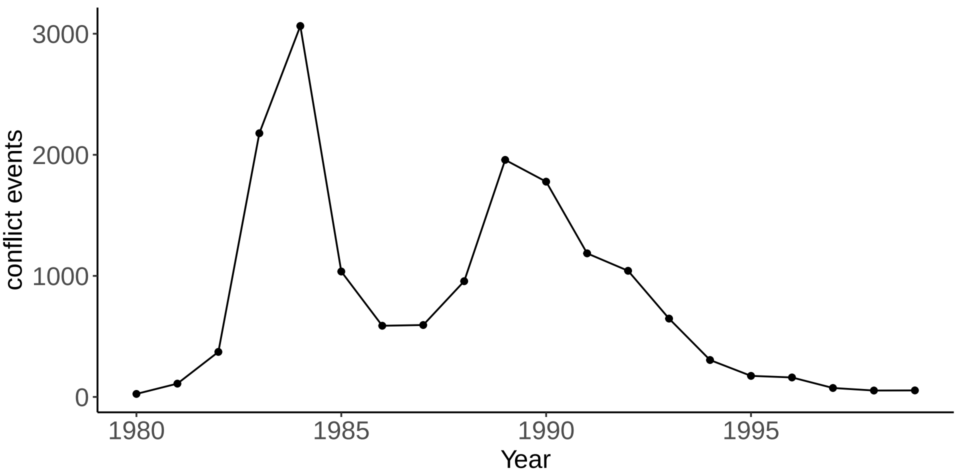



FIGURE 4.5: Violent Conflict events by year (1980-2000)

*Notes:* This figure shows the average number of conflict violent events between 1980 to 2000 in Peru. Violent events includes murders, disappearances, illegal detentions, kidnappings, forced recruitment, torture, rape, and woundings. The data was collected from Peru 's Truth and Reconciliation Commission.

### 4.5.1 Trust

**Interpersonal Trust.** One of the primary mechanisms explaining the dynamics of war is that conflict increases interpersonal trust. Hence, individuals tend to trust their neighbours and relatives more than individuals out of their community. To understand the consequences of the exposure to violence on interpersonal trust, I collect data on interpersonal trust from Latin American Public Opinion Project surveys (LAPOP).[13] I use the LAPOP waves of 2006, 2012, 2014 and 2016. From these surveys I use the variable *IT1*, which asks the question: *"...And speaking of the people from around here, would you say that people in this community are very trustworthy, somewhat trustworthy, not very trustworthy or untrustworthy...?"*. This variable takes value 1 if the individual answered very trustworthy, 2 for somewhat trustworthy, 3 for not very trustworthy, and 4 for untrustworthy. Note that for a more straightforward interpretation of the coefficients, I reverse the original value, hence, the variable takes value 1 for untrustworthy, 2 for not very trustworthy, 3 for somewhat trustworthy, and 4 for very trustworthy.

The use of LAPOP survey comes with different shortcomings, however, to the best of my knowledge, there is no other survey data on interpersonal trust in Peru. The first shortcoming comes from the fact that the data is sampled at the province level instead of the district level. The province level is the second administrative layer in Peru, and the Peruvian territory is divided in 196 provinces. Therefore, when constructing the exposure to violence using this data, individuals were in provinces that were affected by the conflict are classified as exposed to violence independently whether their district recorded a violent event or not. The consequence of this measurement error would push the coefficient of

[13]Source: The AmericasBarometer by the LAPOP Lab, here

interest towards zero, downward biasing the estimation, since individuals that were not affected by the treatment (violent event exposure) are now "treated".

The second shortcoming is the lack of information about the individuals' district or province of birth. In this case, I make the assumption that an individual province of residence is their province of birth. The main results show that individuals exposed to violence and did not migrate are driving our results. The consequence of being unable to differentiate between migrants and not migrants is that the coefficient of interest could be potentially biased towards zero because those individuals who migrated from provinces exposed to violence are now "treated", and I cannot control for it. These shortcomings of using the LAPOP surveys dataset push our coefficient to zero, hence, in the absence of this classical measurement error, I would expect a higher absolute number of the coefficient with lower standard errors.

To test whether exposure to the shining path conflict affected interpersonal trust. I use the specification in equation 4.1. In this case, the dependent variable is my measure of interpersonal trust, and the unit of treatment is the province-year due to the lack of data at the district level. Table C.8 in the appendix show the results on interpersonal trust. Column (1) shows that violence affected individuals' interpersonal trust. The result shows that an individual that was exposed to violence had 6.5% (0.167 units) lower interpersonal trust with respect to the mean than individuals that were not exposed to violence.

In Columns (2), I study whether the effect of trust was different between Indigenous and non-Indigenous individuals. To do so, I create a dummy (*No-Indigenous*) which takes value 1 for non-Indigenous and zero for Indigenous individuals. Note that the inclusion of this variable could lead to potential bias due to the inclusion of a "bad control", hence, the results should be interpreted with caution as a mean comparison and not as a causal effect. In column (2), I include in the *No-Indigenous* dummy as a regressor. The result shows that conditionally on the ethnic group, individuals exposed to the violence had a lower interpersonal trust than those not exposed to violence. Additionally, Non-Indigenous individuals had a 4.4% (0.11 units) higher interpersonal trust with respect to the mean (2.46) than Indigenous people.

These results show that individuals exposed to violence had lower interpersonal trust than individuals that were not exposed to violence. In contrast to Rohner, Thoenig, and Zilibotti (2013b) I find that violent conflict decreased ethnic identity and interpersonal trust specially to Indigenous people. This could be the reason for the not persistent wars in Peru after the Shining Path conflict.

**Generalized Trust.** Rohner, Thoenig, and Zilibotti (2013a,b) argue that violent conflict erodes generalised trust, leading individuals to identify more with their own ethnic group. To understand whether the Shining Path conflict affected generalized trust, I collect data from the Latinobarometro surveys. I use the wave from 2002 to 2015 and select the question: *"Generally speaking, would you say that you can trust most people, or that you can never be too careful when dealing with others?"*. The respondents could answer: 1) You can trust most people or 2) One can never be too careful when dealing with others. Using this information, I create the variable generalize trust, which takes value 1 if the respondent answered: *"You can trust most people"*, and 0 if the respondent answered: *"One can never be too careful when dealing with others"*.

To study how exposure to violence affects an individual's probability of trusting most people (generalized trust). I use a similar identification strategy to the one studying interpersonal trust. Here, I combine the Latinbarometro survey with the TRC conflict database at the province level and use the specification in equation 4.1. Note that similarly to the LAPOP surveys, the Latinbarometer surveys are representative at the province level and do not account for migration. Hence, the identification strategy will suffer the same shortcomings.

Table C.9 in the appendix is a replica of table C.8 but the dependent variable is the probability of an individual answering: *"You can trust most people"*. In column (1), I only include the exposure to conflict variable. The results show that individuals exposed to violence are 1.4 pp. more likely to trust someone than individuals that were not exposed to violence. In column (2), I include the *No-Indigenous* dummy as a control. The results are similar and suggest that individuals exposed to violence are 1.4 pp. more likely to trust someone than individuals that were not exposed to violence. Note that these results are not statistically different from zero.

To further understand how exposure to violence affected trust. I collect data on individual institutional trust from the ENAHO survey module 85 from the waves 2013 to 2017. I collect information about individuals' trust on different institutions. The survey asks the question:*Currently, do you have trust in the following institutions?* The individual could rank from 1 for no trust, 2 for little trust, 3 for enough trust, 4 for a lot of trust. I select trust in the following institutions: district government, national government, the judiciary power, the national congress, and the catholic church. Furthermore, I select a variable that measures whether an individual knows what a democracy or not is.

To test whether the Shining path conflict had an association with institutional trust in Peru, I use the empirical specification embedded in equation 4.1. I use as a dependent variable our trust measures to different institutions, namely trust to: the district government, the national government, the judiciary power, the congress, and the catholic church. Furthermore, I compare whether individuals exposed to the conflict have a higher probability of knowing what democracy is than individuals not exposed to conflict.

Table C.10 in the appendix show the results of the consequences of exposure to conflict on institutional trust. In column (1), the dependent variable is confidence to the district government. In columns (2), (3), (4), (5), (6), the dependent variables are confidence to the national government, the judiciary power, whether an individual knows what democracy and confidence in the catholic church, respectively. The rest of the table resembles the main identification strategy. The independent variable is exposure to conflict. The results show that individuals exposed to conflict do not have different trust in institutions than comparable individuals who were not exposed to conflict. These results could be explained due to the lack of government intervention during the first years of the conflict.

### 4.5.2 Own-Indigenous group violence

The main difference between the Shining Path violence and the setting proposed by Rohner, Thoenig, and Zilibotti (2013a,b) is who committed the violent events. In Rohner, Thoenig, and Zilibotti (2013a,b) violence was perpetrated by a different ethnic group than your own, which decreased generalized trust and increased interpersonal and ethnic identity. In the Peruvian context, as explained in section 4.2, violence was perpetrated by younger generations of the same district that joined the Shining path. Therefore, if own-Indigenous group violence drives the results, I expect the negative association between exposure to violence and negative Indigenous identity is more salient in Indigenous homogeneous districts.

To test this mechanism, I construct a measure of Indigenous diversity at the district level following Alesina, Baqir, and Easterly (1999). This index measures the probability that two randomly drawn people from the same district belong to different Indigenous groups. A higher value of the index represents a higher probability of encountering another individual of a different Indigenous group.

The idea of using the Indigenous diversity index is that in more homogeneous districts, the likelihood of suffering own-Indigenous-group violence is higher, leading to a lower Indigenous identification. To create Indigenous diversity measures at the district level, I use the ENAHO survey and only individuals born before 1950. I use these people because they were at least 32 years at the time of the first attack of the Shining Path, hence, their ethnic identity was already formed. Further, I use the self-identification measure to construct the index. Last, I create a dummy variable from the Indigenous diversity index, which takes the value of 1 if the Indigenous diversity index is larger than 0, hence more than one Indigenous group the district, and takes the value of 0 if there is only one Indigenous group in the district.

To test the own-Indigenous group violence mechanism, I use the regression specification embedded in equation 4.1 and interact the main explanatory variable, exposure to violence, with the measure of ethnic diversity, either as a continuous variable or a dummy. Note that in this specification, the baseline group are individuals not exposed to violence in fully homogeneous districts, hence, only one Indigenous group was present in the district.

Table 4.4 shows the results using the Indigenous diversity index. In columns (1) and (2), I include the Indigenous diversity index dummy, and the dependent variables are whether an Individuals speak an Indigenous language as a mother tongue (column 1) or whether the individual identifies as Indigenous (column 2). In columns (3) and (4), I include the Indigenous diversity index continuous, and the dependent variables are whether an Individuals speak an Indigenous language as a mother tongue (column 3) or whether the individual identifies as Indigenous (column 4). In all specifications I include the full set of fixed effects, district fixed effects, year-interview fixed effects, district-birth fixed effects, and year-birth fixed effects. Further, I include individual controls, sex, age, and age square.

TABLE 4.4: Own-Indigenous group violence

| | Mother tongue | Identify indigenous | Mother tongue | Identify indigenous |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Exposure Violence x T | −0.069*** | −0.045** | −0.058*** | −0.029*** |
| | (0.022) | (0.017) | (0.019) | (0.011) |
| Exposure Violence x T x Indigenous diversity *dummy* | 0.061*** | 0.034 | | |
| | (0.023) | (0.024) | | |
| Exposure Violence x T x Indigenous diversity index | | | 0.213*** | 0.042 |
| | | | (0.057) | (0.053) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ |
| District-birth FE | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| SE Cluster | district-year | district-year | district-year | district-year |
| *N* | 29,911 | 25,566 | 29,911 | 25,566 |
| Adjusted $R^2$ | 0.622 | 0.472 | 0.622 | 0.472 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. In columns (1) and (2), the explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19 interacted with the Indigenous diversity dummy. In columns (3) and (4), the explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19 interacted with the Indigenous diversity index. In columns (1) and (3), the dependent variable is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (2) and (4), the dependent variable is whether an individual identifies Indigenous or not. In all specifications, I restrict the sample to the districts in the border sample. I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

The results show that in Indigenous homogeneous districts, individuals exposed to violence during the ethnic identity formation years are between 6.9 pp. and 5.8 pp. less likely to speak an indigenous language as their mother tongue and between 4.5 pp. and 2.9 pp. less likely to identify as Indigenous than individuals not exposed to violence during their ethnic identity formation years, respect to the difference in ethnic identification between individuals that were and were not exposed to violence after ethnic identity was formed.

Further, the results show that individuals exposed to violence during their youth in districts with more than one Indigenous group are 6.1 pp. more likely to speak an Indigenous language and 3.4 pp. more likely to identify as indigenous than individuals not exposed to violence in their identity formation years respect to the difference in ethnic identification between individuals that were and were not exposed to violence after ethnic identity was formed. This translates into a 1sd (0.17 units) increase in the Indigenous diversity index is associated with an increase of 5.8 pp. in the likelihood that an individual speaks an Indigenous language as mother tongue and 1.1 pp. increase in the probability that an individual identifies as Indigenous.

These results show an increasing number of Indigenous groups in a municipality, hence, a higher probability that the perpetrator of violence was an Indigenous person from a different Indigenous group increases Indigenous identity. On the other hand, in Indigenous homogeneous communities, the likelihood that the perpetrator is from the same Indigenous group decreases the probability that an individual identifies as Indigenous—suggesting that own-Indigenous group violence could be the mechanism driving the main results.

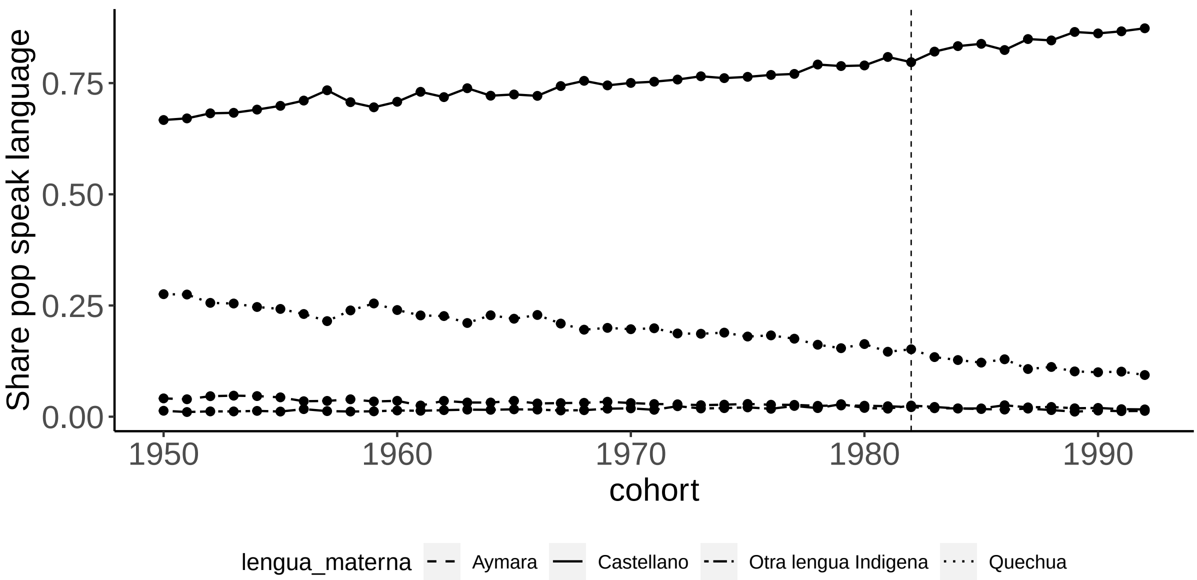
## 4.6 Conclusion

This paper studies the consequences of violence on ethnic identity using the Shining Path conflict as a natural experiment. The results show that individuals that were exposed to violence are less likely to speak an Indigenous language as their mother tongue and identify Indigenous. The mechanism that could explain these findings are own-ethnic group violence. Results show that in Indigenous homogeneous districts, individuals that were exposed to violence are less likely to have an Indigenous identity compared to individuals that were not exposed to violence. Furthermore, the likelihood of having an Indigenous identity increased with Indigenous heterogeneity.

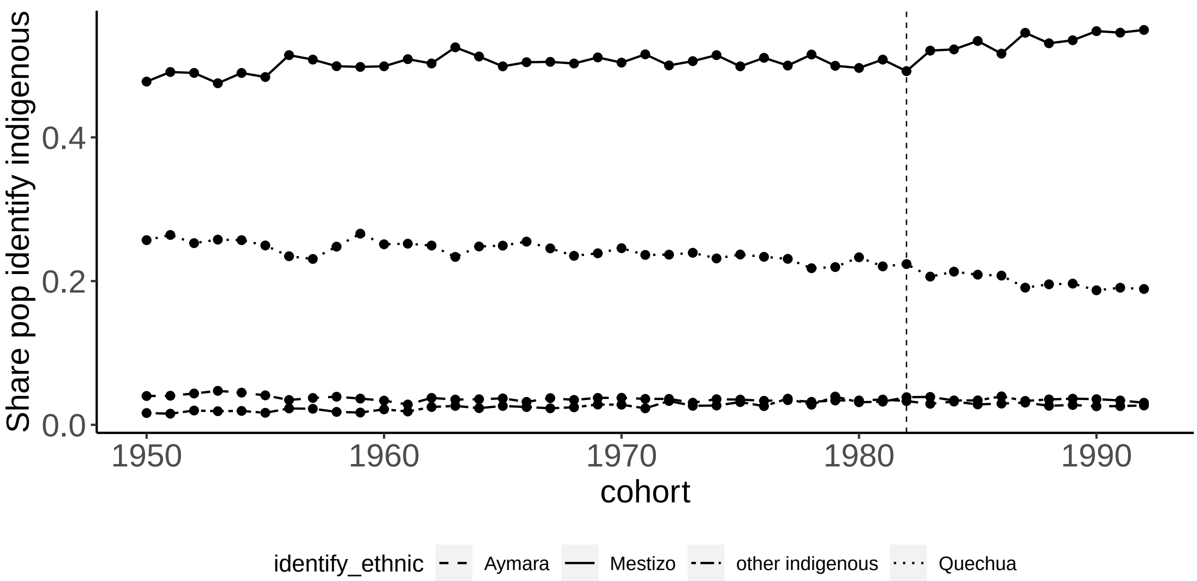
These result complement the advances from Rohner, Thoenig, and Zilibotti (2013a,b) in our understanding of the relationship between violence and ethnic identity, the result suggests that violence has heterogeneous effects on ethnic identity depending on whether violence was perpetrated by a different ethnic group or your own ethnic group. If a different ethnic group committed violence, we could expect an increase in ethnic identity. However, if your ethnic group committed violence, we expect a decrease in ethnic identity.

# Appendix

## C.1 Figures




(A) Indigenous language

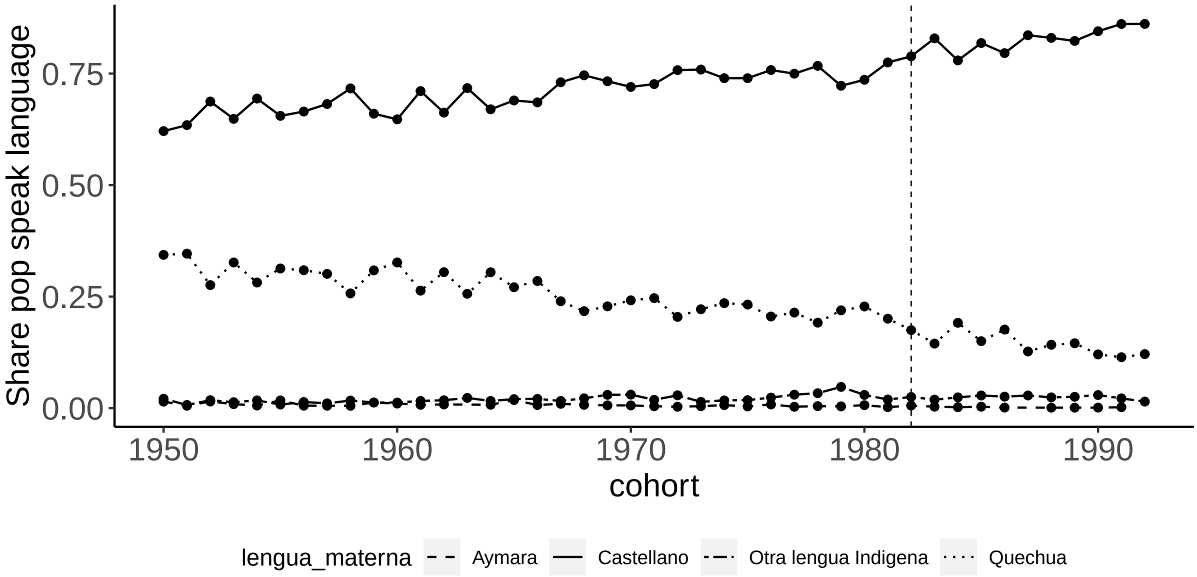



(B) Identify Indigenous

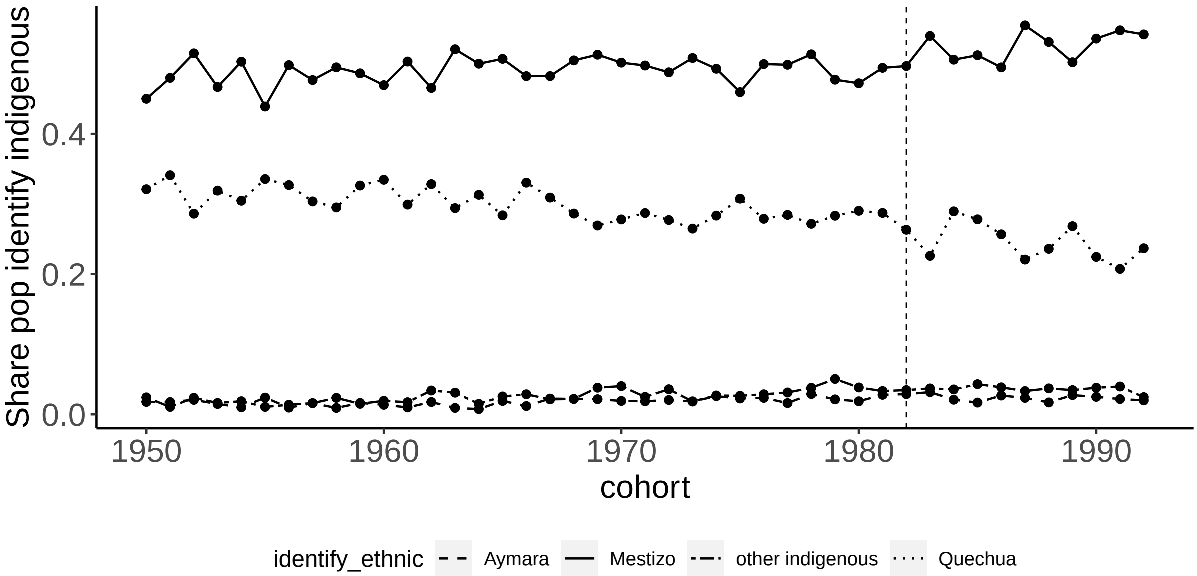
Figure C.1: Ethnic identification: full sample



*Notes:* This figure plots the share of individuals in each cohort by ethnicity. Subfigure C.1a measures ethnicity as which language individuals speak as their mother tongue. Subfigure C.1b measures ethnicity as which ethnic group the individual self-identify as. The data was collected from the ENAHO surveys Peru from 2013 to 2017.



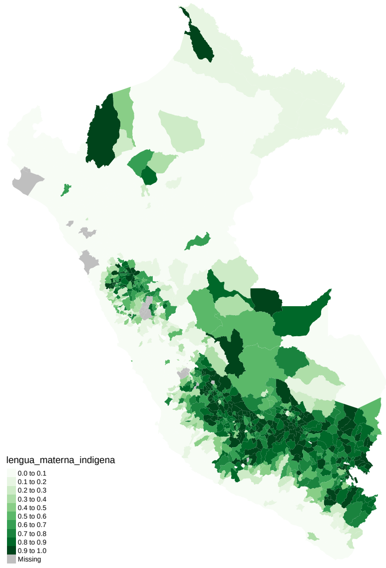

(A) Indigenous language



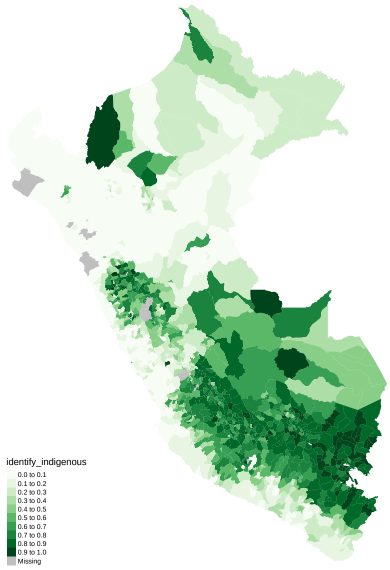

(B) Identify Indigenous

Figure C.2: Ethnic identification: Border sample

*Notes:* This figure plots the share of individuals in each cohort by ethnicity. I restrict the sample to the bordering sample. Subfigure C.2a measures ethnicity as which language individuals speak as their mother tongue. Subfigure C.2b measures ethnicity as which ethnic group the individual self-identify as. The data was collected from the ENAHO surveys Peru from 2013 to 2017.



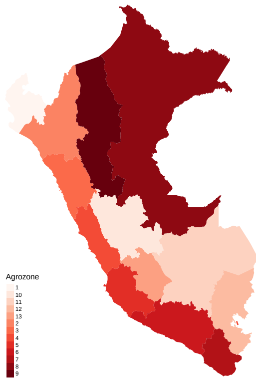

(A) Indigenous language



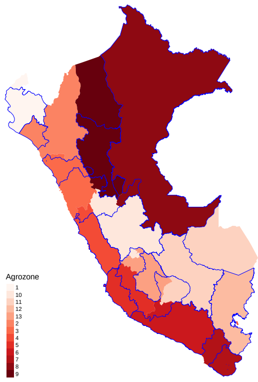

(B) Identify Indigenous

Figure C.3: Indigenous Identity

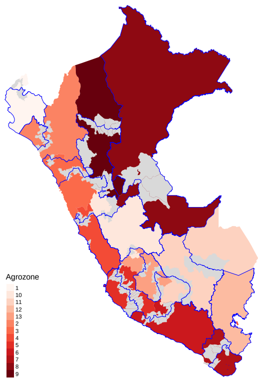
*Notes:* This figure shows the spatial distribution of the Indigenous people in Peru. Subfigure C.3b plots the share of individuals that speak an Indigenous language as their mother tongue in each Peruvian district. Subfigure C.3a plots the share of individuals that identify indigenous in each Peruvian district. The data was collected from the ENAHO surveys Peru from 2013 to 2017.




(A) Agrozones




(B) Agrozones - core and periphery




(C) Agrozones - core and periphery - sample

FIGURE C.4: Agrozones

*Notes:* This figure shows a graphical description of the bordering discontinuity design. Panel C.4a shows Peruvian territory divided by the 13 agrozones, each color represents one agrozone. Panel C.4b adds to the agrozones the borders between "core" and "periphery" districts. The blue line depicts the bordering design. panel C.4c shows in grey all the bordering districts selected for the analysis.

## C.2 Tables

TABLE C.1: Summary Statistics

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| *Full Sample* | | | | | |
| Identify Indigenous | 204,247 | 0.36 | 0.48 | 0 | 1 |
| Indigenous mother tongue | 253,332 | 0.22 | 0.41 | 0 | 1 |
| Exposure violence < 20 | 254,301 | 0.17 | 0.38 | 0 | 1 |
| Exposure violence | 254,301 | 0.25 | 0.44 | 0 | 1 |
| Age | 254,301 | 38.95 | 10.11 | 20 | 64 |
| Year born | 254,301 | 1,976 | 10.00 | 1,958 | 1,992 |
| Sex (Female = 1) | 254,301 | 0.52 | 0.50 | 0 | 1 |
| Migration | 254,301 | 0.70 | 0.46 | 0 | 1 |
| *Border Sample* | | | | | |
| Identify Indigenous | 27,842 | 0.40 | 0.49 | 0 | 1 |
| Indigenous mother tongue | 33,326 | 0.24 | 0.43 | 0 | 1 |
| Exposure violence < 20 | 33,449 | 0.27 | 0.44 | 0 | 1 |
| Exposure | 33,449 | 0.36 | 0.48 | 0 | 1 |
| Exposure non-Indigenous violence < 20 | 33,449 | 0.25 | 0.43 | 0 | 1 |
| Exposure non-Indigenous violence | 33,449 | 0.35 | 0.48 | 0 | 1 |
| Age | 33,449 | 38.98 | 10.14 | 20 | 59 |
| Year Born | 33,449 | 1,976 | 10.01 | 1,958 | 1,992 |
| Sex (Female = 1) | 33,449 | 0.52 | 0.50 | 0 | 1 |
| Migration | 33,449 | 0.65 | 0.48 | 0 | 1 |
| Indigenous diversity index | 30,018 | 0.10 | 0.17 | 0.00 | 0.49 |
| Indigenous diversity index dummy | 30,018 | 0.42 | 0.49 | 0 | 1 |
| *Border Sample variables* | | | | | |
| Violent events | 229 | 21.56 | 83.19 | 0 | 661 |
| Deaths | 229 | 40.00 | 168.46 | 0 | 1,981 |
| Vote Marxist | 229 | 0.12 | 0.12 | 0.00 | 0.88 |
| Non-Indigenous population 1961 | 229 | 0.59 | 0.40 | 0.003 | 1.00 |
| Illiteracy rate 1961 | 229 | 0.53 | 0.22 | 0.01 | 1.00 |
| Illiteracy rate 1972 | 229 | 0.38 | 0.15 | 0.08 | 0.68 |
| Slope | 229 | 6.70 | 3.58 | 0.07 | 18.23 |
| Social movement | 229 | 0.17 | 0.77 | 0 | 6 |
| Share cultivated land | 229 | 5.25 | 5.58 | 0.00 | 37.19 |
| Hectares private land | 229 | 23.16 | 177.98 | 0.00 | 2,680.50 |
| Elevation | 229 | 2.55 | 1.28 | 0.04 | 4.60 |
| State employees 1961 | 229 | 2.74 | 1.45 | 0.00 | 8.37 |
| Population 1972 (thousands) | 229 | 8.44 | 22.16 | 0.22 | 230.81 |
| *log* Land reform hectares | 229 | 4.23 | 10.30 | 0.00 | 81.45 |

*Notes:* The table shows summary statistic for all variables included in the empirical analysis.

Table C.2: Total: District SE

| | Indigenous language as mother tongue | | Identify Indigenous | |
|---|---|---|---|---|
| | all districts | border sample | all districts | border sample |
| | (1) | (2) | (3) | (4) |
| Exposure violence x T | $-0.014^{**}$ | $-0.034^{*}$ | $-0.005$ | $-0.023^{**}$ |
| | (0.007) | (0.018) | (0.005) | (0.011) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ |
| District-born FE | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 0.23 | 0.25 | 0.36 | 0.40 |
| SE Cluster | district | district | district | district |
| $N$ | 208,768 | 33,326 | 168,301 | 27,842 |
| Adjusted $R^2$ | 0.660 | 0.628 | 0.545 | 0.480 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1)-(2) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (3)-(4), the dependent variable is whether an individual identifies Indigenous or not. In columns (1) and (3), I use the full sample of individuals. In columns (2) and (4), I restrict the sample to the districts in the border sample. In all specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.4: Migration

| | Indigenous Language as mother tongue | | | identify Indigenous | | | Migration | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Exposure violence x T | $-0.034^{**}$ | $-0.034^{**}$ | $-0.050$ | $-0.023^{**}$ | $-0.023^{**}$ | $-0.035^{**}$ | | 0.006 |
| | (0.016) | (0.016) | (0.037) | (0.009) | (0.009) | (0.014) | | (0.004) |
| Migration | | 0.007 | 0.001 | | 0.008 | 0.003 | | |
| | | (0.004) | (0.008) | | (0.008) | (0.011) | | |
| Migration x Exposure | | | 0.004 | | | 0.008 | | |
| | | | (0.017) | | | (0.020) | | |
| Exposure violence x T x Migration | | | 0.025 | | | 0.018 | | |
| | | | (0.039) | | | (0.018) | | |
| Exposure | | | | | | | $0.084^{***}$ | |
| | | | | | | | (0.002) | |
| District-interview FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| District-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| SE Cluster | district-year | district-year | district-year | district-year | district-year | district-year | | district-year |
| Sample | Border | Border | Border | Border | Border | Border | Full | Border |
| $N$ | 33,326 | 33,326 | 33,326 | 27,842 | 27,842 | 27,842 | 254,301 | 33,449 |
| Adjusted $R^2$ | 0.628 | 0.628 | 0.628 | 0.480 | 0.480 | 0.480 | 0.006 | 0.829 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. In Column (7), explanatory variables is whether an individual was exposed to violence or not and I use the full sample in this column. The dependent variable in columns (1)-(3) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (4)-(6), the dependent variable is whether an individual identifies Indigenous or not. In columns (7) and (8), the dependent variable is whether an individual migrated or not. In columns (2) and (5), we include a dummy variable indicating whether an individual migrated or not as a control variable. In columns (3) and (6), we interact the main explanatory variable with the *migration* dummy. In all specifications except for column (7), I restrict the sample to the districts in the border sample. I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

TABLE C.3: Total: No post cohort

| | Indigenous language as mother tongue | | Identify Indigenous | |
|---|---|---|---|---|
| | all districts | border sample | all districts | border sample |
| | (1) | (2) | (3) | (4) |
| Exposure violence x T | −0.051*** | −0.046 | −0.009 | −0.022* |
| | (0.010) | (0.030) | (0.006) | (0.013) |
| | (0.006) | (0.016) | (0.005) | (0.009) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ |
| District-born FE | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 0.23 | 0.25 | 0.36 | 0.40 |
| SE Cluster | district-year | district-year | district-year | district-year |
| $N$ | 204,434 | 32,417 | 164,565 | 27,048 |
| Adjusted $R^2$ | 0.666 | 0.628 | 0.548 | 0.477 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1)-(2) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (3)-(4), the dependent variable is whether an individual identifies Indigenous or not. In columns (1) and (3), I use the full sample of individuals except for the post cohort. In columns (2) and (4), I restrict the sample to the districts in the border sample. In all specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

TABLE C.5: Ethnic groups

| | Indigenous language as mother tongue | | | Identify Indigenous | | |
|---|---|---|---|---|---|---|
| | Quechua | Aymara | Other Indigenous | Quechua | Aymara | Other Indigenous |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Exposure violence x T | −0.031** | 0.002 | −0.007* | −0.011 | −0.007 | −0.012** |
| | (0.015) | (0.002) | (0.004) | (0.007) | (0.007) | (0.005) |
| Intensive margin controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District-interview FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 0.22 | 0.006 | 0.029 | 0.22 | 0.006 | 0.0294 |
| SE Cluster | district-year | district-year | district-year | district-year | district-year | district-year |
| $N$ | 32,395 | 25,471 | 26,074 | 26,130 | 17,509 | 17,869 |
| Adjusted $R^2$ | 0.660 | 0.135 | 0.658 | 0.559 | 0.255 | 0.500 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1)-(3) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (4)-(6), the dependent variable is whether an individual identifies Indigenous or not. In columns (1) and (4), I only include the Quechua population. In columns (2) and (5), I only include the Aymara population. In columns (3) and (6), I only include Other Indigenous population. In all specifications, I restrict the sample to the districts in the border sample. I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.6: Total

| | Indigenous language as mother tongue | | Identify Indigenous | |
|---|---|---|---|---|
| | all districts | border sample | all districts | border sample |
| | (1) | (2) | (3) | (4) |
| Exposure cohort post | −0.130*** | −0.034 | −0.017 | −0.001 |
| | (0.017) | (0.029) | (0.031) | (0.031) |
| Exposure cohort 0-4 | −0.088*** | −0.070*** | −0.020 | −0.040** |
| | (0.010) | (0.023) | (0.018) | (0.018) |
| Exposure cohort 5-9 | −0.049*** | −0.044** | −0.007 | −0.025 |
| | (0.008) | (0.019) | (0.020) | (0.020) |
| Exposure cohort 10-14 | −0.045*** | −0.029 | −0.014 | −0.017 |
| | (0.006) | (0.022) | (0.020) | (0.020) |
| Exposure cohort 15-19 | −0.014*** | −0.006 | 0.009 | 0.008 |
| | (0.005) | (0.016) | (0.013) | (0.013) |
| Exposure cohort 25-29 | 0.007 | 0.036* | 0.004 | 0.023 |
| | (0.011) | (0.021) | (0.024) | (0.024) |
| Exposure cohort 30-34 | 0.003 | −0.015 | 0.012 | 0.020 |
| | (0.017) | (0.060) | (0.035) | (0.035) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ |
| District-birth FE | ✓ | ✓ | ✓ | ✓ |
| Year-interview FE | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 0.23 | 0.25 | 0.36 | 0.40 |
| *N* | 208,768 | 33,326 | 168,301 | 27,842 |
| Adjusted $R^2$ | 0.661 | 0.629 | 0.545 | 0.480 |

*Notes:* The table shows results from estimating equation 4.2. The unit of observation is the individual. The omitted cohort is 20-24. The dependent variable in columns (1)-(2) is whether an individual speaks an Indigenous language as their mother tongue or not. In columns (3)-(4), the dependent variable is whether an individual identifies Indigenous or not. In columns (1) and (3), I use the full sample of individuals. In columns (2) and (4), I restrict the sample to the districts in the border sample. In all specifications, I include the full set of fixed effects, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.7: Total

| | totalevents | deaths | marxistleft | nonindigenous_pc_1961 | illiterate_pc_1961 | illiterate_pc_1972 | slope |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| inside_agrozone_core | 30.004*** | 53.254** | 0.021 | −0.067 | −0.026 | −0.021 | −0.977** |
| | (10.868) | (22.093) | (0.016) | (0.053) | (0.029) | (0.019) | (0.471) |
| *N* | 229 | 229 | 229 | 229 | 229 | 229 | 229 |
| $R^2$ | 0.032 | 0.025 | 0.007 | 0.007 | 0.003 | 0.005 | 0.019 |

| | mov_comunidad | cult_land | prop_ha_ths | elev_1k | lemployees1961 | population_1972ths | LRpercap_calweighted_log |
|---|---|---|---|---|---|---|---|
| | (8) | (9) | (10) | (11) | (12) | (13) | (14) |
| inside_agrozone_core | 0.151 | 0.533 | 22.560 | −0.428** | 0.119 | 3.861 | 1.355 |
| | (0.101) | (0.740) | (23.592) | (0.168) | (0.193) | (2.932) | (1.365) |
| *N* | 229 | 229 | 229 | 229 | 229 | 229 | 229 |
| $R^2$ | 0.010 | 0.002 | 0.004 | 0.028 | 0.002 | 0.008 | 0.004 |

*Notes:* The table shows the standardized mean differences of different variables between core and periphery districts. The base category is core districts. The unit of observation is the districts. The variables are total number of violent events, total number of death, share of votes to Marxist parties , share of non-Indigenous population in 1961, illiteracy rate in 1961, , illiteracy rate in 1972, peasant movements, average slope, share of cultivated land, Hectares of private land, average elevation, number of state employees, population in 1972 (thousands), Hectares of land reform. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.8: Trust interpersonal

| | Interpersonal trust | |
|---|---|---|
| | (1) | (2) |
| Exposure violence x T | −0.167* | −0.147 |
| | (0.101) | (0.095) |
| No-Indigenous | | 0.109** |
| | | (0.052) |
| District FE | ✓ | ✓ |
| Survey FE | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ |
| Controls | ✓ | ✓ |
| Mean DV | 2.46 | 2.46 |
| SE Cluster | province-year | province-year |
| *N* | 5,158 | 5,039 |
| Adjusted $R^2$ | 0.032 | 0.033 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable is the measure of interpersonal trust. In column (2), I include a dummy variable indicating whether the individual is Indigenous or not. In all specifications, I include, year-birth fixed effects, year-survey fixed effects, district fixed effects. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.9: Trust: Generalized

| | Generalized trust | |
|---|---|---|
| | (1) | (2) |
| Exposure violence x T | 0.014 | 0.013 |
| | (0.018) | (0.017) |
| No-Indigenous | | −0.010 |
| | | (0.030) |
| District FE | ✓ | ✓ |
| Survey FE | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ |
| Controls | ✓ | ✓ |
| Mean DV | 0.16 | 0.16 |
| SE Cluster | province-year | province-year |
| *N* | 5,905 | 5,731 |
| Adjusted $R^2$ | 0.018 | 0.011 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable is the measure of generalized trust. In column (2), I include a dummy variable indicating whether the individual is Indigenous or not. In all specifications, I include, year-birth fixed effects, year-survey fixed effects, district fixed effects. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.10: Trust: Institutions

| | district | national | judiciary | congress | know democracy | catholic church |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Exposure violence x T | 0.009 | −0.003 | 0.035 | −0.013 | −0.005 | 0.012 |
| | (0.020) | (0.015) | (0.022) | (0.020) | (0.007) | (0.025) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District-birthFE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Agrozone FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 1.88 | 1.74 | 1.78 | 1.60 | 0.53 | 2.49 |
| SE Cluster | province-year | province-year | province-year | province-year | province-year | province-year |
| *N* | 28,676 | 27,471 | 27,860 | 27,603 | 29,804 | 29,168 |
| Adjusted $R^2$ | 0.056 | 0.048 | 0.048 | 0.058 | 0.182 | 0.084 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. I restrict the sample to the bordering sample. The dependent variable in columns (1),(2), (3), (4), (5), (6) are measures of trust in the district government, the national government, the judiciary power, the congress, and the catholic church. In all specifications, I include, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth fixed effects. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

Table C.11: Trust: Institutions (Full sample)

| | district | national | judiciary | congress | know democracy | catholic church |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Exposure violence | 0.011* | −0.012 | 0.011* | −0.002 | −0.007* | 0.004 |
| | (0.006) | (0.011) | (0.006) | (0.014) | (0.004) | (0.012) |
| District-interview FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| District-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Survey FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Year-birth FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Mean DP | 1.87 | 1.74 | 1.77 | 1.60 | 0.53 | 2.52 |
| SE Cluster | province-year | province-year | province-year | province-year | province-year | province-year |
| $N$ | 181,013 | 173,637 | 176,110 | 174,345 | 187,703 | 183,955 |
| Adjusted $R^2$ | 0.056 | 0.062 | 0.052 | 0.059 | 0.157 | 0.084 |

*Notes:* The table shows results from estimating equation 4.1. The unit of observation is the individual. The explanatory variable is the interaction term between exposure to violence and exposure to violence up to age of 19. The dependent variable in columns (1),(2), (3), (4), (5), (6) are measures of trust in the district government, the national government, the judiciary power, the congress, and the catholic church. In all specifications, I include, year-birth fixed effects, year-interview fixed effects, district of interview fixed effects, and district of birth fixed effects. Baseline controls include age, age square, and sex. Standard errors clustered at the district-year level in parenthesis. Statistical significance at the 99%, 95%, and 90% confidence level denoted: ***, **, and *.

## C.3 Description of Variables

### C.3.1 Main variables

**Identify Indigenous.** Dummy variable which takes the value of 1 if the individual answered to the question: *"According to your ancestor and your customs, do you consider yourself: Quechua, Aymara, Other Indigenous community, Black, White, Mestizo, or unknown?"*, Quechua, Aymara, Other Indigenous community, and takes the value of 0 if the individual answered Mestizo. I restrict the sample to individuals born between 1958 and 1992. Data obtained from Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza (ENAHO).

**Indigenous mother tongue.** Dummy variable which takes the value of 1 if the individual answered to the question: "*What is the language or mother tongue you learned in your childhood"*, Quechua, Aymara, Other Indigenous language, and takes the value of 0 if the individual answered Castellano (Spanish). I restrict the sample to individuals born between 1958 and 1992. Data obtained from Data obtained from Encuesta Nacional de Hogares sobre Condiciones de Vida y Pobreza (ENAHO).

**Exposure to violence.** Dummy variable which takes the value 1 if individual $i$ was born in a district that was exposed to a violent attack of the Shining Path between 1980 and 1992 were the victim was an Indigenous person. I restrict the sample to individuals born between 1958 and 1992. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI), and Truth and Reconciliation Commission (TRC).

**Exposure to non-Indigenous violence.** Dummy variable which takes the value 1 if individual $i$ was born in a district that was exposed to a violent attack of the Shining Path between 1980 and 1992 were the victim was not an Indigenous person. I restrict the sample to individuals born between 1958 and 1992. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI), and Truth and Reconciliation Commission (TRC).

**Exposure to violence $< 20$.** Dummy variable which takes the value 1 if individual $i$ was born in a district that was exposed to a violent attack of the Shining Path between 1980 and 1992 between age 0 and 19 were the victim was an Indigenous person, and takes value 0 otherwise. I restrict the sample to individuals born between 1958 and 1992. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI). and Truth and Reconciliation Commission (TRC).

**Exposure to non-Indigenous violence $< 20$.** Dummy variable which takes the value 1 if individual $i$ was born in a district that was exposed to a violent attack of the Shining Path between 1980 and 1992 between age 0 and 19 were the victim was not an Indigenous person, and takes value 0 otherwise. I restrict the sample to individuals born between 1958 and 1992. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI). and Truth and Reconciliation Commission (TRC).

**Age.** Age of individual $i$. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Year born.** Year born of individual $i$. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Sex.** Dummy variable which takes value 1 for female, and 0 for male. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

**Indigenous diversity index.** Measures the probability that two randomly drawn people from the same district belong to different Indigenous groups. A higher value of the index represents a higher probability of encountering another individual of a different Indigenous group. I restrict the sample to individuals born before 1950. To create the index, I follow Alesina, Baqir, and Easterly (1999), and use the following equation C.3.1, where $Indigenous_i d$ represent the share of population self-identify from indigenous group

$i \in$ (Quechua, Aymara, Other Indigenous) in district $d$. Data obtained from Mexican Instituto Nacional de Estadística y Geografía (INEGI).

$$Indigenous\ diversity\ index = 1 - \sum_i (Indigenous_i d)^2 \tag{C.3.1}$$

**Indigenous diversity index dummy.** Dummy variable which takes the value of 1 if the Indigenous diversity index is larger than 0, and takes the value of 0 otherwise.

### C.3.2 Sampling design variables

**Total events.** The number of attacks during the Peruvian civil conflict, which include murders, disappearances, illegal detentions, kidnappings, forced recruitment, torture, rape, and woundings, at the district level from 1980 to 2000. Data obtained from Albertus (2020).

**Deaths.** The deaths during the Peruvian civil conflict at the district level from 1980 to 2000. Data obtained from Albertus (2020).

**Vote Marxist.** Marxist vote share in the 1980 election. Data obtained from Albertus (2020).

**Non-Indigenous pop in 1961.** The share of non-Indigenous population in 1961 at the district level. Data obtained from Albertus (2020).

**Non-Indigenous pop in 1961.** The share of non-Indigenous population in 1961 at the district level. Data obtained from Albertus (2020).

**Non-Indigenous pop in 1961.** Measure as the share of individuals that speak spanish as their mother tongue in 1961 at the district level. Data obtained from Albertus (2020).

**Land reform Hc.** Land reform per capita at the district level. Data obtained from Albertus (2020).

**Population in 1972.** The number of inhabitants in thousands in 1972 at the district level. Data obtained from Albertus (2020).

**State Employees.** The number of state personnel in 1961 at the district level. Data obtained from Albertus (2020).

**Hc private land.** The number hectares of private land in thousands at the district level. Data obtained from Albertus (2020).

**Elevation.** The average elevation in meters at the district level. Data obtained from Albertus (2020).

**Share cultivable land.** The share of cultivable land at the district level. Data obtained from Albertus (2020).

**Social movement.** The number of previous autonomous communal uprisings at the district level. Data obtained from Albertus (2020).

**Slope.** The average slope in degree at the district level. Data obtained from Albertus (2020).

**Illiteracy rate 1972.** Share of population without reading and writing ability in 1972 at the district level. Data obtained from Albertus (2020).

**Illiteracy rate 1961.** Share of population without reading and writing ability in 1972 at the district level. Data obtained from Albertus (2020).

## C.4 Discussion Albertus (2020)

The sample border design in this paper is substantially different than Albertus (2020) study. First, I do a bordering discontinuity design à la Lichter, Löffler, and Siegloch (2021) and select the sample of municipalities at the land reform zone borders, whereas Albertus (2020) use the exogenous border and perform a regression discontinuity design. Second, I include all the "agrarian reform zones" in Peru, and Albertus (2020) includes "agrarian reform zones" 1 to 7 and 11. Third, I perform a difference-in-differences analysis, which allows me to compare cohorts that were exposed to violence against cohorts that were not exposed to violence in districts with and without conflict events, and Albertus (2020) focuses on the effect of land reform on conflict.

The literature has studied the relationship between land reform and conflict. The evidence regarding the association between land reform and conflict is mixed. Some studies find a positive relationship[14] and others a negative relationship.[15] Remarkably, authors had found opposites results on the association between land reform and violence for same land reform and conflict. For example, in the Colombian land reform, Guardado (2018)

[14]Positve relationship: Albertus, Brambor, and Ceneviva (2018), Albertus and Kaplan (2013), Finkel, Gehlbach, and Olsen (2015), and Mason (1998)

[15]Negative relationship: Albertus (2020), Guardado (2018), Huntington (2006), Kalyvas (2006), and Wood (2003)

finds a negative relationship through the opportunity cost mechanism. The opportunity cost is the most commonly proposed mechanism for explaining the negative relationship between conflict and land reform. Landless peasants have no opportunity cost for entering a guerrilla movement, increasing the likelihood of conflict.[16] On the other hand, Albertus and Kaplan (2013) finds a positive relationship between land reform and conflict through the "low-intensity" land reform mechanism. A "low-intensity" land reform could incentivize recipients of land reform to fight for more land redistribution when land reforms do not provide significant improvements for those who received land, leading to an increase in the probability of conflict. Last, another mechanism explaining the positive relationship between land reform and conflict is the lack of government investment in infrastructures such as irrigation or market access that landowners provided before the reform, leading to a decrease in production and an increase in the likelihood of conflict (Kapstein, 2017).

The main difference between the bordering sample of this paper is that I include all agrozones and Albertus (2020), omitted Zones 8, and 9 because no major land reform applies; Zone 10 because they split into two zones 10, and 13 in 1974; and Zone 12 because it was not split by core and peripheral areas. The inclusion of Zones 8, and 9 could explain the positive relationship through the "low intensity" mechanism. If, in these Zones, land reform was low, it could incentivize peasants to join the Shining Path Guerrilla due to the Shining Path rhetoric of social justice. Furthermore, the inclusion of Zones 10 and 13 could explain the positive relationship between land reform and conflict through the lack of state capacity mechanism. Zones 10 and 13 are located in the departments of Junin and Pasco, which are bordering departments to the department of Ayacucho, the crane of the Shining Path guerrilla. As discussed in the background section, the Shining path propagated fast through rural areas because of the low state presence. Therefore, I hypothesize that the lack of state presence in those departments could have caused a lack of government infrastructure leading to an increasing conflict in areas with higher land reform.

[16] This mechanism explains Albertus (2020) findings

# Bibliography


Acemoglu, Daron, Camilo García-Jimeno, and James A Robinson (2015). State capacity and economic development: A network approach. *American Economic Review* 105(8), pp. 2364–2409.

Adams, Richard EW and Murdo J Macleod (2000). History of the native peoples of the Americas. Cambridge University Press.

Ainsworth, Mary S (1989). Attachments beyond infancy. *American psychologist* 44(4), p. 709.

Akerlof, George A and Rachel E Kranton (2000). Economics and identity. *The quarterly journal of economics* 115(3), pp. 715–753.

— (2002). Identity and schooling: Some lessons for the economics of education. *Journal of economic literature* 40(4), pp. 1167–1201.

— (2010). Identity economics. Princeton University Press.

Albertus, Michael (2020). Land reform and civil conflict: Theory and evidence from Peru. *American Journal of Political Science* 64(2), pp. 256–274.

Albertus, Michael, Thomas Brambor, and Ricardo Ceneviva (2018). Land inequality and rural unrest: theory and evidence from Brazil. *Journal of Conflict Resolution* 62(3), pp. 557–596.

Albertus, Michael and Oliver Kaplan (2013). Land reform as a counterinsurgency policy: Evidence from Colombia. *Journal of Conflict Resolution* 57(2), pp. 198–231.

Alesina, Alberto, Reza Baqir, and William Easterly (1999). Public goods and ethnic divisions. *The Quarterly journal of economics* 114(4), pp. 1243–1284.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg (2003). Fractionalization. *Journal of Economic growth* 8(2), pp. 155–194.

Alesina, Alberto and Eliana La Ferrara (2005). Ethnic diversity and economic performance. *Journal of economic literature* 43(3), pp. 762–800.

Andreas, Peter (1996). U.S.–Mexico: Open Markets, Closed Border. *Foreign Policy* 103, pp. 51–69.

— (2012). Border games: Policing the US-Mexico divide. Cornell University Press.

Angeles, Luis and Aldo Elizalde (2017). Pre-colonial institutions and socioeconomic development: The case of Latin America. *Journal of Development Economics* 124, pp. 22–40.

Angrist, Joshua D. and Adriana D. Kugler (2008). Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia. *Review of Economics and Statistics* 90(2), pp. 191–215.

Arbatlı, Cemal Eren, Quamrul H Ashraf, Oded Galor, and Marc Klemp (2020). Diversity and conflict. *Econometrica* 88(2), pp. 727–797.

Arrowsmith, Aaron (1810). A New Map of Mexico and Adjacent Provinces Compiled from Original Documents de la Colección Cartográfica SCOP. The National Archives of Mexico.

Arteaga, Fernando (2018). The Historical Legacy of (Pre?) Colonial Indigenous Settlements in Mexico.

Asher, Sam and Paul Novosad (2020). Rural Roads and Local Economic Development. *American Economic Review* 110(3), pp. 797–823.

Aspachs-Bracons, Oriol, Irma Clots-Figueras, Joan Costa-Font, and Paolo Masella (2008). Compulsory language educational policies and identity formation. *Journal of the European Economic Association* 6(2-3), pp. 434–444.

Bandyopadhyay, Sanghamitra and Elliott Green (2016). Precolonial political centralization and contemporary development in Uganda. *Economic Development and Cultural Change* 64(3), pp. 471–508.

Basu, Sukanya and Sarah Pearlman (2017). Violence and migration: evidence from Mexico's drug war. *IZA Journal of Development and Migration* 7(1), pp. 1–29.

Battu, Harminder and Yves Zenou (2010). *Oppositional identities and employment for ethnic minorities: Evidence from England.*

Baum-Snow, Nathaniel, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics* 99(3), pp. 435–448.

Bazzi, Samuel and Christopher Blattman (2014). Economic Shocks and Conflict: Evidence from Commodity Prices. *American Economic Journal: Macroeconomics* 6(4), pp. 1–38.

Becker, Gary S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76. Ed. by Nigel G. Fielding, Alan Clarke, and Robert Witt, pp. 169–217.

Bénabou, Roland and Jean Tirole (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics* 126(2), pp. 805–855.

Berman, Nicolas and Mathieu Couttenier (2015). External Shocks, Internal Shots: The Geography of Civil Conflicts. *Review of Economics and Statistics* 97(4), pp. 758–776.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig (2017). This Mine Is Mine! How Minerals Fuel Conflicts in Africa. *American Economic Review* 107(6), pp. 1564–1610.

Bess, Michael K (2014). Routes of conflict: building roads and shaping the nation in Mexico, 1941–1952. *The Journal of Transport History* 35(1), pp. 78–96.

— (2017). Routes of Compromise: Building Roads and Shaping the Nation in Mexico, 1917-1952. U of Nebraska Press.

Bess, Michael Kirkland (2016). Revolutionary Paths: Motor Roads, Economic Development, and National Sovereignty in 1920s and 1930s Mexico. *Mexican Studies/Estudios Mexicanos* 32(1), pp. 56–82.

Blankespoor, Brian, Theophile Bougna, Rafael Garduño Rivera, and Harris Selod (2017). Roads and the geography of economic activities in Mexico. *techreport*. The World Bank.

Bowlby, John (1982). Attachment and loss: retrospect and prospect. *American journal of Orthopsychiatry* 52(4), p. 664.

Brand, Elaine S, Rene A Ruiz, and Amado M Padilla (1974). Ethnic identification and preference: A review. *Psychological Bulletin* 81(11), p. 860.

Bruhn, Miriam and Francisco A Gallego (2012). Good, bad, and ugly colonial activities: do they matter for economic development? *Review of Economics and Statistics* 94(2), pp. 433–461.

Burfisher, Mary E, Sherman Robinson, and Karen Thierfelder (2001). The Impact of NAFTA on the United States. *Journal of Economic Perspectives* 15(1), pp. 125–144.

Calderón, Gabriela, Gustavo Robles, Alberto Díaz-Cayeros, and Beatriz Magaloni (2015). The Beheading of Criminal Organizations and the Dynamics of Violence in Mexico. *Journal of Conflict Resolution* 59(8), pp. 1455–1485.

Cama, Shireen F and Priya Sehgal (2021). Racial and ethnic considerations across child and adolescent development. *Academic Psychiatry* 45(1), pp. 106–109.

Caselli, Francesco and Wilbur John Coleman (2013). On the theory of ethnic conflict. *Journal of the European Economic Association* 11(suppl_1), pp. 161–192.

Casey, Teresa and Christian Dustmann (2010). *Immigrants' identity, economic outcomes and the transmission of identity across generations.*

Castillo, Juan Camilo and Dorothy Kronick (2020). The Logic of Violence in Drug War. *American Political Science Review* 114(3), pp. 874–887.

Castillo, Juan Camilo, Daniel Mejía, and Pascual Restrepo (2020). Scarcity without leviathan: The violent effects of cocaine supply shortages in the mexican drug war. *Review of Economics and Statistics* 102(2), pp. 269–286.

Chen, Yan and Sherry Xin Li (2009). Group identity and social preferences. *American Economic Review* 99(1), pp. 431–57.

Chias, Luis, Héctor Reséndiz, and Juan C García (2010). El sistema carretero como articulador de las ciudades. *Los grandes problemas de México II. Desarrollo urbano y regional.*

Chiovelli, Giorgio (2014). Pre-colonial centralization, colonial activities and development in Latin America. *Unpublished manuscript, University of Bologna.*

Clots-Figueras, Irma and Paolo Masella (2013). Education, language and identity. *The Economic Journal* 123(570), F332–F357.

Coatsworth, John H (1979). Indispensable railroads in a backward economy: The case of Mexico. *Journal of Economic History*, pp. 939–960.

Cockburn, Alexander and Jeffrey St. Clair (1998). Whiteout: The Cia, Drugs, and the Press. London, New York: Verso.

Collier, Paul and Anke Hoeffler (1998). On Economic Causes of Civil War. *Oxford Economic Papers* 50(4), pp. 563–573.

— (2004a). Greed and Grievance in Civil War. *Oxford Economic Papers* 56(4), pp. 563–595.

— (2004b). Greed and grievance in civil war. *Oxford economic papers* 56(4), pp. 563–595.

Collier, Paul, Anke Hoeffler, and Dominic Rohner (2009). Beyond greed and grievance: feasibility and civil war. *oxford Economic papers* 61(1), pp. 1–27.

Constant, Amelie F and Klaus F Zimmermann (2008). Measuring ethnic identity and its impact on economic behavior. *Journal of the European Economic Association* 6(2-3), pp. 424–433.

CVR, Comisión de la Verdad y Reconciliación (2003). *Informe final de la Comisión de la Verdad y Reconciliación*. URL: https://www.cverdad.org.pe/ifinal/index.php (visited on 05/12/2022).

Dal Bó, Ernesto and Pedro Dal Bó (2011). Workers, Warriors, and Criminals: Social Conflict in General Equilibrium. *Journal of the European Economic Association* 9(4), pp. 646–677.

Dalgaard, Carl-Johan, Nicolai Kaarsen, Ola Olsson, and Pablo Selaya (2018). Roman roads to prosperity: Persistence and non-persistence of public goods provision.

Daniele, Gianmarco, Marco Le Moglie, and Federico Masera (2020). Pains, Guns and Moves: The Effect of the US Opioid Epidemic on Mexican Migration. *Working Paper 2020-141*. Bocconi University.

Davey, AG and PN Mullin (1980). Ethnic identification and preference of British primary school children. *Child Psychology & Psychiatry & Allied Disciplines*.

De Janvry, Alain, Elisabeth Sadoulet, and Gustavo Gordillo De Anda (1995). NAFTA and Mexico's maize producers. *World Development* 23(8), pp. 1349–1362.

De Vries, Leonie Ansems and Elspeth Guild (2019). Seeking refuge in Europe: spaces of transit and the violence of migration management. *Journal of ethnic and migration studies* 45(12), pp. 2156–2166.

DEA (1995). The Cocaine Threat to the United States: Drug Intelligence Report. *techreport*. U.S. Department of Justice: Drug Enforcement Administration.

— (2016). 2016 National Drug Threat Assessment. *techreport*. U.S. Department of Justice Drug Enforcement Administration.

Deiana, Claudio (2016). The Bitter Side of Trade Shocks: Local Labour Market Conditions and Crime in the Us. *Mimeo*.

Dell, Melissa (2015). Trafficking Networks and the Mexican Drug War. *American Economic Review* 105(6), pp. 1738–79.

Dell, Melissa, Benjamin Feigenberg, and Kensuke Teshima (2019). The Violent Consequences of Trade-induced Worker Displacement in Mexico. *American Economic Review: Insights* 1(1), pp. 43–58.

Dell, Melissa, Nathan Lane, and Pablo Querubin (2018). The historical state, local collective action, and economic development in Vietnam. *Econometrica* 86(6), pp. 2083–2121.

Dermota, Ken (1999). Snow Business: Drugs and the Spirit of Capitalism. *World Policy Journal* 16(4), pp. 15–24. URL: http://www.jstor.org/stable/40209659 (visited on 07/29/2022).

Díaz-Cayeros, Alberto, Beatriz Magaloni, and Alexander Ruiz-Euler (2014). Traditional governance, citizen engagement, and local public goods: evidence from Mexico. *World Development* 53, pp. 80–93.

Dijkstra, Edsger Wybe (1959). Two Problems in Connectionwith Graphs. *Numerische Mathematik* 1, pp. 269–271.

Dippel, Christian (2014). Forced coexistence and economic development: evidence from Native American Reservations. *Econometrica* 82(6), pp. 2131–2165.

Dix-Carneiro, Rafael, Rodrigo R Soares, and Gabriel Ulyssea (2018). Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization. *American Economic Journal: Applied Economics* 10(4), pp. 158–95.

Donaldson, Dave (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review* 108(4-5), pp. 899–934.

Dube, Arindrajit, Oeindrila Dube, and Omar García-Ponce (2013). Cross-border Spillover: Us Gun Laws and Violence in Mexico. *American Political Science Review* 107(3), pp. 397–417.

Dube, Oeindrila, Omar García-Ponce, and Kevin Thom (2016). From Maize to Haze: Agricultural Shocks and the Growth of the Mexican Drug Sector. *Journal of the European Economic Association* 14(5), pp. 1181–1224.

Dube, Oeindrila and Juan F. Vargas (2013). Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *Review of Economic Studies* 80(4), pp. 1384–1421.

Duflo, Esther and Rohini Pande (2007). Dams. *The Quarterly Journal of Economics* 122(2), pp. 601–646.

Elizalde, Aldo (2020a). On the economic effects of Indigenous institutions: Evidence from Mexico. *Journal of Development Economics* 147, p. 102530. ISSN: 0304-3878.

— (2020b). On the Economic Effects of Indigenous Institutions: Evidence from Mexico. *Journal of Development Economics* 147, p. 102530.

Elizalde, Aldo, Eduardo Hidalgo, and Nayeli Salgado (2021). Public good or public bad? Indigenous institutions and the demand for public goods. *Mimeo, University of Cologne.*

Engerman, Stanley L and Kenneth L Sokoloff (1997). Factor endowments, institutions, and differential paths of growth among new world economies. *How Latin America Fell Behind*, pp. 260–304.

Erikson, Erik H (1968). Identity: Youth and crisis. 7. WW Norton & company.

Esteban, Joan, Laura Mayoral, and Debraj Ray (2012). Ethnicity and conflict: An empirical study. *American Economic Review* 102(4), pp. 1310–42.

Esteban, Joan and Debraj Ray (2011). Linking conflict to inequality and polarization. *American Economic Review* 101(4), pp. 1345–74.

Fick, Stephen E. and Robert J. Hijmans (2017). Worldclim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *International Journal of Climatology* 37(12), pp. 4302–4315.

Finkel, Evgeny, Scott Gehlbach, and Tricia D Olsen (2015). Does reform prevent rebellion? Evidence from Russia's emancipation of the serfs. *Comparative Political Studies* 48(8), pp. 984–1019.

Fisman, Raymond and Shang-Jin Wei (2009). The Smuggling of Art, and the Art of Smuggling: Uncovering the Illicit Trade in Cultural Property and Antiques. *American Economic Journal: Applied Economics* 1(3), pp. 82–96.

Foote, Wiliam F. (1997). Mexico's troubled toll roads.

Galdo, Jose (2013). The long-run labor-market consequences of civil war: Evidence from the Shining Path in Peru. *Economic Development and Cultural Change* 61(4), pp. 789–823.

García-Ponce, Omar (2017). Women's Political Participation After Civil War: Evidence from Peru. *Manuscript. University of California, Davis. At here. pdf, accessed July* 16, p. 2019.

Garfias, Francisco (2018). Elite competition and state capacity development: Theory and evidence from post-revolutionary Mexico. *American Political Science Review* 112(2), pp. 339–357.

Gennaioli, Nicola and Ilia Rainer (2006). Precolonial centralization and institutional quality in Africa. *Institutions and norms in economic development*, pp. 21–46.

— (2007). The modern impact of precolonial centralization in Africa. *Journal of Economic Growth* 12(3), pp. 185–234.

Giuliano, Paola and Nathan Nunn (Dec. 2020). Understanding Cultural Persistence and Change. *The Review of Economic Studies.*

Goldstein, Paul J. (1985). The Drugs/violence Nexus: A Tripartite Conceptual Framework. *Journal of Drug Issues* 15(4), pp. 493–506.

Guardado, Jenny (2018). Land tenure, price shocks, and insurgency: Evidence from Peru and Colombia. *World Development* 111, pp. 256–269.

Gúzman, Abigael (1974). *LA PROBLEMATICA NACIONAL.* URL: https://cedema.org/digital_items/631 (visited on 05/12/2022).

Hargreaves Heap, Shaun P and Daniel John Zizzo (2009). The value of groups. *American Economic Review* 99(1), pp. 295–323.

Harvey, Neil (1998). The Chiapas rebellion: The struggle for land and democracy. Duke University Press.

Heldring, Leander (June 2020). The Origins of Violence in Rwanda. *The Review of Economic Studies* 88(2), pp. 730–763.

Herbst, Jeffrey (2014). States and power in Africa: Comparative lessons in authority and control. Vol. 149. Princeton University Press.

Hernán, Sergio Ortiz (1994). Caminos y transportes en México: una aproximación socioeconómica: fines de la colonia y principios de la vida independiente. Secretaria de Comunicaciones y Transportes.

Hidalgo, Eduardo (2022). Violence and Ethnic Identity: Evidence from Perú. *Mimeo, University of Cologne.*

Hidalgo, Eduardo, Erik Hornung, and Pablo Selaya (2022). NAFTA and drug-related violence in Mexico. *Mimeo, University of Cologne.*

Hodges, Donald and Ross Gandy (1983). Mexico 1910-1976: Reform Or Revolution? *London: Zed Press.*

Holcomb-McCoy, Cheryl (2005). Ethnic identity development in early adolescence: Implications and recommendations for middle school counselors. *Professional School Counseling* 9(2).

Hornung, Erik (2015). Railroads and growth in Prussia. *Journal of the European Economic Association* 13(4), pp. 699–736.

Huntington, Samuel P (2006). Political order in changing societies. Yale university press.

IGBP-DIS (1998). *SoilData(V.0) A program for creating global soil-property databases, IGBP Global Soils Data Task, France.* URL: http://atlas.sage.wisc.edu/. (01.11.2020).

IISS (2017). Armed Conflict Survey 2017. *techreport*. International Institute for Strategic Studies.

ILO, International Labour Organization (2019). *Implementing the ILO Indigenous and tribal peoples convention No. 169: Towards an inclusive, sustainable and justice future.*

Iyer, Lakshmi and Petia Topalova (2014). Poverty and crime: evidence from rainfall and trade shocks in India.

Jacques, Scott (2010). The Necessary Conditions for Retaliation: Toward a Theory of Non-violent and Violent Forms in Drug Markets. *Justice Quarterly* 27(2), pp. 186–205.

Jima-González, Alexandra and Miguel Paradela-López (2020). Indians in Pensamiento Gonzalo: The Influence of 20th-Century Peruvian Intelligentsia on Shining Path's Ideology. *SAGE Open* 10(4), p. 2158244020982990.

— (2021). The Negative Impact of Shining Path on Indigenous Mobilization in Peru: An Approach from Political Opportunity and New Social Movements Theories. *Latin American Perspectives* 48(6), pp. 194–209.

Juhn, Chinhui, Gergely Ujhelyi, and Carolina Villegas-Sanchez (2013). Trade Liberalization and Gender Inequality. *American Economic Review* 103(3), pp. 269–73.

Kalyvas, Stathis N (2006). The logic of violence in civil war. Cambridge University Press.

Kapstein, Ethan B (2017). Seeds of stability: land reform and US foreign policy. Cambridge University Press.

Kent, Robert B (1993). Geographical dimensions of the Shining Path insurgency in Peru. *Geographical Review*, pp. 441–454.

Kranton, Rachel E (2016). Identity economics 2016: Where do social distinctions and norms come from? *American Economic Review* 106(5), pp. 405–09.

Kulher, Ulrich (1975). Cambio cultural dirigido en los Altos de Chiapas. *México, INE.*

La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert Vishny (1999). The quality of government. *The Journal of Law, Economics, and Organization* 15(1), pp. 222–279.

Lang, James (1975). Conquest and commerce: Spain and England in the Americas. New York: Academic Press.

Laplante, Lisa J and Kimberly Theidon (2007). Truth with consequences: Justice and reparations in post-Truth Commission Peru. *Hum. Rts. Q.* 29, p. 228.

Leon, Gianmarco (2012). Civil conflict and human capital accumulation the long-term effects of political violence in perú. *Journal of Human Resources* 47(4), pp. 991–1022.

Lichter, Andreas, Max Löffler, and Sebastian Siegloch (2021). The long-term costs of government surveillance: Insights from stasi spying in East Germany. *Journal of the European Economic Association* 19(2), pp. 741–789.

Lipscomb, Molly, A Mushfiq Mobarak, and Tania Barham (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics* 5(2), pp. 200–231.

Lustig, Nora and Mariano Tommasi (2020). El COVID-19 y la protección social de los grupos pobres y vulnerables. *PNUD (abril).*

Marcia, James E et al. (1980). Identity in adolescence. *Handbook of adolescent psychology* 9(11), pp. 159–187.

Marks, Amy Kerivan, Laura A Szalacha, Meaghan Lamarre, Michelle J Boyd, and Cynthia García Coll (2007). Emerging ethnic identity and interethnic group social preferences in middle childhood: Findings from the Children of Immigrants Development in Context (CIDC) study. *International Journal of Behavioral Development* 31(5), pp. 501–513.

Mason, T David (1998). " Take Two Acres and Call Me in the Morning": Is Land Reform a Prescription for Peasant Unrest? *The Journal of Politics* 60(1), pp. 199–230.

Michalopoulos, Stelios and Elias Papaioannou (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica* 81(1), pp. 113–152.

— (2014). National institutions and subnational development in Africa. *The Quarterly journal of economics* 129(1), pp. 151–213.

Miguel, Edward and Mary Kay Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of public Economics* 89(11-12), pp. 2325–2368.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti (2004). Economic Shocks and Civil Conflict: An Instrumental Variables Approach. *Journal of Political Economy* 112(4), pp. 725–753.

Montalvo, José G and Marta Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *American economic review* 95(3), pp. 796–816.

Montalvo, Jose G and Marta Reynal-Querol (2017). Ethnic diversity and growth: Revisiting the evidence.

Montenegro, Raul A and Carolyn Stephens (2006). Indigenous health in Latin America and the Caribbean. *The Lancet* 367(9525), pp. 1859–1869.

Morrison, Andrew R (1993). Violence or economics: what drives internal migration in Guatemala? *Economic Development and Cultural Change* 41(4), pp. 817–831.

Murdock, George Peter (1967). Ethnographic atlas: a summary. *Ethnology* 6(2), pp. 109–236.

Murphy, Tommy E and Martín A Rossi (2016). Land Reform and Violence: Evidence from Mexico. *Journal of Economic Behavior & Organization* 131, pp. 106–113.

— (2020). Following the Poppy Trail: Origins and Consequences of Mexican Drug Cartels. *Journal of Development Economics* 143, p. 102433.

Nunn, Nathan (2012). Culture and the historical process. *Economic History of Developing Regions* 27(sup1), S108–S126.

Osorio, Javier (2015). The Contagion of Drug Violence: Spatiotemporal Dynamics of the Mexican War on Drugs. *Journal of Conflict Resolution* 59(8), pp. 1403–1432.

Parkes, Colin Murray, Joan Stevenson-Hinde, and Peter Marris (2006). Attachment across the life cycle. Routledge.

PCP, Partido Comunista del Perú [Sendero Luminoso] (1976). *El problema campesino y la revolución*. URL: https://cedema.org/digital_items/635 (visited on 05/12/2022).

Phinney, Jean S (1989). Stages of ethnic identity development in minority group adolescents. *The Journal of Early Adolescence* 9(1-2), pp. 34–49.

Phinney, Jean S (1990). Ethnic identity in adolescents and adults: review of research. *Psychological bulletin* 108(3), p. 499.

Reuter, Peter (2009). Systemic Violence in Drug Markets. *Crime, Law and Social Change* 52(3), pp. 275–284.

Robertson, Raymond (2000). Wage shocks and North American labor-market integration. *American Economic Review* 90(4), pp. 742–764.

Rohner, Dominic (2011). Reputation, group structure and social tensions. *Journal of Development Economics* 96(2), pp. 188–199.

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti (2013a). Seeds of distrust: Conflict in Uganda. *Journal of Economic Growth* 18(3), pp. 217–252.

— (2013b). War signals: A theory of trade, trust, and conflict. *Review of Economic Studies* 80(3), pp. 1114–1147.

Russo, Francesco Flaviano (2014). Cocaine: The Complementarity between Legal and Illegal Trade. *The World Economy* 37(9), pp. 1290–1314.

Schüller, Simone (2015). Parental ethnic identity and educational attainment of second-generation immigrants. *Journal of Population Economics* 28(4), pp. 965–1004.

SCT (1987). Volúmenes de tránsito local de 1987. Red carretera troncal. República Mexicana. 1st ed. México, D.F.: Secretaría de Comunicaciones y Transportes.

— (1988). El Transporte en México. Pasado, presente, futuro. 1st ed. México, D.F.: Secretaría de Comunicaciones y Transportes.

Servan-Mori, Edson, Pilar Torres-Pereda, Emanuel Orozco, and Sandra G Sosa-Rubí (2014). An explanatory analysis of economic and health inequality changes among Mexican indigenous people, 2000-2010. *International journal for equity in health* 13(1), pp. 1–8.

Shining Path (1988). *Documentos fundamentales [Fundamental Documents]*. URL: http://www.solrojo.org/pcp_doc/pcp_gd88.htm (visited on 05/12/2022).

Shreve, Ronald L. (1966). Statistical Law of Stream Numbers. *The Journal of Geology* 74(1), pp. 17–37.

Silva, Adriana Carolina and Douglas S Massey (2015). Violence, networks, and international migration from Colombia. *International Migration* 53(5), pp. 162–178.

Sobrino, Fernanda (2020). Mexican Cartel Wars: Fighting for the US Opioid Market. *mimeo*.

Stewart, Scott (2016). From Colombia to New York City: The Narconomics of Cocaine. *Business Insider (June 27)*. URL: https://www.businessinsider.com/from-colombia-to-new-york-city-the-economics-of-cocaine-2015-7?r=DE&IR=T.

Sviatschi, Maria Micaela (2022). Making a Narco: Childhood Exposure to Illegal Labor Markets and Criminal Life Paths. *Econometrica* 90(4), pp. 1835–1878.

Tabellini, Guido (2008). The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics* 123(3), pp. 905–950.

Tajfel, Henri and John C Turner (2004). The social identity theory of intergroup behavior. In: *Political psychology*. Psychology Press, pp. 276–293.

Tanck de Estrada, Dorothy (2005). Atlas ilustrado de los pueblos de indios. Nueva España 1800. El Colegio de Mexico, El Colegio Mexiquense, CNDPI, Fomento Cultural Banamex.

The White House (1990). National Drug Control Strategy. *techreport*. Office of National Drug Control Policy, Executive Office of the President.

— (1992). National Drug Control Strategy: Reclaiming Our Communities From Drugs and Violence. *techreport*. Office of National Drug Control Policy, Executive Office of the President.

— (2003). National Drug Control Strategy: 2003 - Data Supplement. *techreport*. Office of National Drug Control Policy, Executive Office of the President.

Trejo, Guillermo and Sandra Ley (2018). Why Did Drug Cartels Go to War in Mexico? Subnational Party Alternation, the Breakdown of Criminal Protection, and the Onset of Large-scale Violence. *Comparative Political Studies* 51(7), pp. 900–937.

UNODC (2010). Transnational Organized Crime Threat Assessment. *techreport*. United Nations Office on Drugs and Crime.

Von Glümer, Bodo (1896). Carta de Comunicaciones de los Estados Unidos Mexicanos. Formada de orden de la Secretaría de Comunicaciones y Obras Públicas, escala 1.1:000,000. The National Archives of Mexico.

Waters, Wendy (2006). Remapping Identities: Road Construction and Nation Building in Postrevolutionary Mexico. (M. Vaughan. & S. Lewis E., Eds.). Durham, NC: Duke University Press.

Watt, Peter and Roberto Zepeda (2012). Drug War Mexico: Politics, Neoliberalism and Violence in the New Narcoeconomy. Bloomsbury Publishing.

Weiner, Tim and Tim Golden (1993). Free-Trade Treaty May Widen Traffic In Drugs, U.S. Says. *The New York Times (May 24, p. 1)*. URL: https://www.nytimes.com/1993/05/24/world/free-trade-treaty-may-widen-traffic-in-drugs-us-says.html.

Wheat, Andrew (1990). Shining Path's" fourth sword" ideology. *Journal of Political & Military Sociology*, pp. 41–55.

Wood, Elisabeth Jean (2003). Insurgent collective action and civil war in El Salvador. Cambridge University Press.

The World Bank (2015). *Indigenous Latin America in the twenty-first century: the first decade.*

Xin Li, Sherry (2010). Social identities, ethnic diversity, and tax morale. *Public Finance Review* 38(2), pp. 146–177.

Yip, Tiffany (2014). Ethnic identity in everyday life: The influence of identity development status. *Child Development* 85(1), pp. 205–219.

# List of Applied Software

- R version 4.1.0 and RStudio 1.4.1717: Applied to clean data, create maps, and for the empirical analysis.
- Excel 2019: Applied for basic data cleaning and data storage.
- TeXstudio 3.0.4: Used to compile the drafts of the single papers and the final thesis.

# Eidesstattliche Erklärung nach § 8 Abs. 3 der Promotionsordnung vom 17.02.2015

Hiermit versichere ich an Eides Statt, dass ich die vorgelegte Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Aussagen, Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir dienachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich geholfen:

- -

Weitere Personen, neben den ggf. in der Einleitung der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Köln, February 28, 2023