École des Hautes Études en Sciences Sociales

EHESS Doctoral School

Doctoral School nº 465
Économie Panthéon Sorbonne

Economic Sciences

Daniel SÁNCHEZ-ORDÓÑEZ

Essays on Land Conflict, Productivity, and Inequality

Supervised by: Prof. Facundo ALVAREDO

Defense date: June 15, 2026

Jury:

Prof. Leticia ARROYO-ABAD *(Referee)*, CUNY - Queens College & Graduate Center
Prof. Denis COGNEAU, Paris School of Economics
Prof. Diego SÁNCHEZ-ANCOCHEA, University of Oxford
Prof. Emiliano TRAVIESO *(Referee)*, Carlos III University of Madrid
Prof. Juan VARGAS, Collegio Carlo Alberto – University of Turin

École des Hautes Études en Sciences Sociales

EHESS Doctoral School

École doctorale nº 465
Économie Panthéon Sorbonne

Sciences Économiques

Daniel SÁNCHEZ-ORDÓÑEZ

Essais sur le Conflit, la Productivité et l'Inégalité de la Terre

---

Thèse dirigée par : Prof. Facundo ALVAREDO

Date de soutenance : 15 juin 2026

Jury :

Prof. Leticia ARROYO-ABAD *(Rapporteur)*, CUNY - Queens College & Graduate Center
Prof. Denis COGNEAU, École d'Économie de Paris
Prof. Diego SÁNCHEZ-ANCOCHEA, Université d'Oxford
Prof. Emiliano TRAVIESO *(Rapporteur)*, Université Carlos III de Madrid
Prof. Juan VARGAS, Collegio Carlo Alberto – Université de Turin

Acknowledgements

Acknowledgements are never fair. I will surely miss some contributors.

I want to start by thanking my advisor, Facundo Alvaredo. It is no exaggeration to say that I would not have started, much less finished, this project without his support. As a frustrated Master's student, confronted with an academic world increasingly preoccupied with the measurable and the methodologically secure, I stumbled into Facundo's office. From our first meeting, he never minced words about the need for research to move knowledge forward. I am not sure whether this thesis does so, but his unwavering commitment to the scientific project has left an impression on me that I will happily carry forward. For this, and for ample support throughout this PhD, I am deeply appreciative.

I want to thank all the members of my jury for engaging with my work so generously. Denis Cogneau has provided helpful comments and advice since I took his class during the Master's at PSE. The four other members of my jury, Leticia Arroyo-Abad, Diego Sánchez-Ancochea, Emiliano Travieso, and Juan Vargas, all provided their time and comments over the past year. I could not be happier with their participation, as the breadth of their expertise has undoubtedly strengthened the thesis. I hope our conversations continue well into the future.

I want to particularly thank all my co-authors for the many years of collaboration that helped produce Chapters 2 and 3. Regarding Chapter 2, I want to thank Margarita Gáfaro, Ana-María Ibáñez, and María Camila Ortiz for such a delightful and productive collaboration. Working with all three of you gave me an ideal benchmark for how teams should be organized and led with efficiency, respect, and intellectual exchange. For Chapter 3, a large thanks is in order to Luis Bauluz, Yajna Govind, and Filip Novokmet. Being my first co-authorship, it was a collaboration that brought many new lessons, but working with three such kind people always made it a pleasure.

Next, I want to thank the institutions, colleagues, and friends who participated in one way or another. I want to thank the Econ History Group at PSE for helping instill in me the great historical tradition of the EHESS. I hope you all keep passing down the heritage for as long as possible. The World Inequality Lab became a (somewhat late) home filled with many bright faces. It is a challenge to list the countless people whom I have crossed paths with thanks to the lab. Thanks to Álvaro, Alice, Gastón, Mark, and Óscar for many conversations and laughs over the years. A special shout-out to Ignacio and Rowaida for wonderfully positive friendships at many moments when I needed them and for an endless array of conversations. While embattled at times, PSE did house me for many years, and hence I had the chance to know many bright minds. For having been constant sparring partners from the first years

of the PhD, as well as (semi) office-mates, I thank Juan Camilo and Baptiste. Due to my own stubbornness, I did not fully integrate into the so-called "Development Group" until an unexpected trip to México late in the PhD, but I couldn't be happier to count Andrea, Anna, Billy, Héctor, Maya, María, Paola, and Sebastián as friends more than colleagues. The Political Economy Project group filled the first two years of the PhD with a world of knowledge and stakes that I wish professional academia offered more often. Nonetheless, our long conversations on the intricacies of T1, T2, and the beginnings of T3 will remain some of the finest memories of the Covid years. Hopefully, some more are yet to come. Hence, I want to sincerely thank Facundo, Ingrid, Mauricio, Marc, Pilar, and Simone for the serious engagement and countless discussions.

Next comes a smattering of individuals who helped in numerous ways, whether by being companions at the dimmest of conferences or by providing advice and support during my travels. You were all wonderfully unexpected additions along a long journey: Andrés, César, Marco, Pierre, Raphaël, and Timo. Particular thanks are due to Juliana Jaramillo for her generosity with her time and efforts, and for introducing an ignorant economic historian of Colombia to a wealth of data and people, not to mention a lasting friendship. Arriving at the end, I want to thank the many friends who, in different ways, accompanied me throughout the last couple of years: Catherine, Isabelle, Iván, Juliana, Kit, Lou, Maggie, Muñoz, Parra, Paul, Pedro, Rico, Romain, Salomé, Santiago, Sarah, and Sofía. A special thanks to Mateo for his patience during the summer of 2025, and to Álvaro and David for their unexpected help in the final stretch. The sounds of Brian, Franco, Gustav, Roberto, and Philip deserve a special mention.

Finally, my sincerest thanks go to my family, Ana, Ernesto, Andrés, Lena, and Lynn-Alexandra, for decades of constancy, and to Zoé for more than her fair share of patience, a great deal of belief, and a steadfast desire to experience joy, all of which made the day-to-day that much more bearable.

Summary

This thesis consists of three essays on the role of land in structuring conflict, productivity, and inequality in the agricultural sector. This thesis starts by analyzing the case study of Colombia, then moves to the Latin American region before broadening further to the entire developing world.

Chapter 1 examines Colombia's two largest civil conflicts, the Thousand Days War (1899–1902) and *La Violencia* (1948–1964), and the First Globalization coffee boom that took place between them. The chapter yields two sets of results. First, the geography of conflict *incidence* persists across centuries: municipalities that experienced battles during the Thousand Days War are significantly more likely to see *La Violencia* break out half a century later, with partisan identity acting as the channel of transmission. Second, conflict *intensity* does not persist: nineteenth-century conflict measures fail to predict the number of violent deaths during *La Violencia*. In a horse-race against all leading explanations from the historiography, coffee production is the strongest correlate of violence intensity. Once armed-group presence is accounted for, the direct effect of coffee on violence drops to near zero, consistent with armed groups serving as the channel through which coffee rents translate into conflict.

Chapter 2 gauges the contribution of agricultural dynamics to income inequality in Latin America and the Caribbean, using harmonized agricultural census and labor survey microdata for nine countries. Theil decompositions show that agricultural income inequality contributes between 11% in Uruguay and 58% in Bolivia to total income inequality, with an unusually large between-sector component. The chapter explores the role of farm size in shaping agricultural incomes: across countries, regions with higher average farm size have higher average agricultural household income and agricultural wages, consistent with the inverse relationship between farm size and land yields and the positive relationship between farm size and output per worker. The prevalence of very small farms in much of Latin America therefore carries a country-level cost: low agricultural household income contributes meaningfully to overall income inequality.

Chapter 3 revisits the measurement and economic role of land inequality in the developing world. Existing cross-country evidence relies on agricultural censuses, which measure operational holdings by area, treat land as homogeneous, and exclude the landless. The chapter constructs a harmonized household-level database on land ownership, land value, income, credit, and wealth across seventeen developing countries. Three findings emerge.

First, land *value*, not land area, explains household income, access to credit, and wealth among agricultural households. Second, once land value and the landless are incorporated, land inequality is at least as concentrated as overall wealth, with a Gini coefficient of around 0.77. Cross-country differences in land inequality shrink under value-based measures, reflecting a robust inverse relationship between landholding size and value per hectare. Third, the chapter introduces a measure of land-based asset poverty: the share of households below a US$1,000 PPP threshold ranges from under 15% in China and Vietnam to over 60% in Malawi and Ethiopia, yet the aggregate land value required to close these gaps is often modest relative to total land wealth. Land-based deprivation is widespread in incidence but modest in aggregate magnitude.

Keywords: land inequality, conflict, agriculture, coffee, Latin America, developing countries, farm size, income distribution, land value, poverty

JEL codes: N56, O13, O15, O54, Q15, J43, Q12

Résumé

Cette thèse est composée de trois essais sur le rôle de la terre dans la structuration du conflit, de la productivité et de l'inégalité dans le secteur agricole. La thèse commence par l'analyse du cas colombien, puis élargit le propos à la région latino-américaine avant d'étendre au monde en développement dans son ensemble.

Chapitre 1 examine les deux plus grands conflits civils de Colombie, la Guerre des Mille Jours (1899–1902) et *La Violencia* (1948–1964), ainsi que le boom caféier de la Première Mondialisation qui s'est produit entre ces deux évènements. Le chapitre livre deux séries de résultats. Premièrement, la géographie de l'*incidence* du conflit persiste à travers les siècles : les municipalités ayant connu des batailles pendant la Guerre des Mille Jours sont significativement plus susceptibles de voir *La Violencia* éclater un demi-siècle plus tard, l'identité partisane agissant comme canal de transmission. Deuxièmement, l'*intensité* du conflit ne persiste pas : les mesures de conflit du XIXe siècle ne prédisent pas le nombre de morts violentes durant *La Violencia*. Dans une régression incluant toutes les explications dominantes de l'historiographie, la production de café est le corrélat le plus fort de l'intensité de la violence. Une fois la présence de groupes armés prise en compte, l'effet direct du café sur la violence tombe à un niveau proche de zéro, ce qui est cohérent avec le rôle des groupes armés comme canal par lequel les rentes du café se traduisent en conflit.

Chapitre 2 évalue la contribution des dynamiques agricoles à l'inégalité des revenus en Amérique latine et dans les Caraïbes, à partir de microdonnées harmonisées de recensements agricoles et d'enquêtes sur l'emploi pour neuf pays. Les décompositions de Theil montrent que l'inégalité des revenus agricoles contribue entre 11 % en Uruguay et 58 % en Bolivie à l'inégalité totale des revenus, avec une composante inter-sectorielle inhabituellement élevée. Le chapitre explore le rôle de la taille des exploitations dans la formation des revenus agricoles : dans l'ensemble des pays, les régions où la taille moyenne des exploitations est plus élevée présentent également des revenus agricoles moyens et des salaires agricoles plus élevés, conformément à la relation inverse entre taille des exploitations et rendements de terre agricole et à la relation positive entre taille des exploitations et production par travailleur. La prévalence de très petites exploitations dans une grande partie de l'Amérique latine représente donc un coût au niveau national : les faibles revenus des ménages agricoles contribuent de manière significative à l'inégalité globale des revenus.

Chapitre 3 revisite la mesure et le rôle économique de l'inégalité de terre agricole dans le monde en développement. Les données internationales existantes reposent sur les

recensements agricoles, qui mesurent les exploitations par superficie, traitent la terre comme homogène et excluent les sans-terre. Le chapitre construit une base de données harmonisée au niveau des ménages couvrant la propriété de terre agricole, la valeur de la terre, le revenu, le crédit et le patrimoine dans dix-sept pays en développement. Trois résultats principaux émergent. Premièrement, c'est la *valeur* de la terre, et non sa superficie, qui explique le revenu des ménages, l'accès au crédit et le patrimoine parmi les ménages agricoles. Deuxièmement, une fois la valeur de terre agricole et les sans-terre intégrés, l'inégalité de terre agricole est au moins aussi concentrée que l'inégalité globale de patrimoine, avec un coefficient de Gini d'environ 0,77. Les différences internationales d'inégalité de terre agricole se réduisent lorsqu'on utilise des mesures fondées sur la valeur, reflétant une relation inverse robuste entre la taille des propriétés de terre agricole et la valeur par hectare. Troisièmement, le chapitre introduit une mesure de la pauvreté en actifs de terre agricole : la part des ménages en dessous d'un seuil de 1 000 dollars PPA varie de moins de 15 % en Chine et au Vietnam à plus de 60 % au Malawi et en Éthiopie, mais la valeur de terre agricole agrégée nécessaire pour combler ces écarts reste modeste par rapport à la valeur de terre agricole totale. La privation de terre agricole est donc largement répandue en incidence mais modeste en magnitude agrégée.

Mots-clés : inégalité de terre agricole, conflit, agriculture, café, Amérique latine, pays en développement, taille des exploitations, distribution des revenus, valeur de terre agricole, pauvreté, productivité

Codes JEL : N56, O13, O15, O54, Q15, J43, Q12

Contents

Acknowledgements v

Summary viii

Résumé xi

List of Tables xviii

List of Figures xx

General Introduction 1

1 (Un-)Persistent Conflict? First Globalization Coffee Boom in Colombia . . . 8

1.1 Introduction 8

1.2 Historical Context & Analysis 11

1.2.1 First Globalization and Coffee Boom, 1875-1955 11

1.2.2 From 19th Century Civil Wars to *La Violencia*, 1948-1965 15

1.3 Data & Measurement 20

1.3.1 Variables 20

1.3.2 Construction of Historical Geographic Units (HGU) 23

1.4 Does Conflict Persist Across Centuries? 24

1.4.1 Testing Persistence at the Municipality Level 24

1.4.2 Where Violence Erupts: Incidence Persists Across Centuries 25

1.4.3 How Intense Violence Becomes: Intensity Does Not Persist 26

1.4.4 How Partisan Identity Transmits Conflict Across Centuries 28

1.5 What Explains the Intensity of *La Violencia*? 31

1.5.1 Coffee Production and Violence Intensity 31

1.5.2 How Coffee Translates into Violence: Rents and Fighting Feasibility 33

1.5.3 Armed Groups as the Channel 35

1.5.4 Historical Evidence: Coffee Rents and Armed Group Operations . . . 37

1.6 Conclusion 40

2 Farm Size and Income Distribution of Latin American Agriculture **49**

2.1 Introduction 49

2.2 Data sources 53

2.2.1 Agricultural Census data . . . 53
2.2.2 Labor surveys . . . 55
2.2.3 Other controls . . . 57
2.3 Description of Agricultural Sector: Distribution of Land and Income . . . 58
2.3.1 Land Inequality: Context and Estimates . . . 58
2.3.2 Land Size Structure in LAC . . . 61
2.3.3 Income Inequality: Concepts and Estimates . . . 63
2.4 Land size, agricultural productivity, and income inequality . . . 68
2.5 Rural income, agricultural labor, and farm size . . . 71
2.5.1 Two facts about farm size and agricultural labor . . . 72
2.5.2 Heterogeneity across regions and average agricultural income . . . 76
2.5.3 Heterogeneity within regions and the within agricultural income inequality . . . 77
2.6 Conclusions . . . 79

3 Land Inequality in the Developing World . . . 86
3.1 Introduction . . . 86
3.2 Concepts and Data . . . 90
3.2.1 Data Sources and Database Construction . . . 90
3.2.2 Definitions and Measurement . . . 91
3.2.3 Inequality Measurement Concepts . . . 92
3.2.4 Data Validation . . . 93
3.3 Agricultural households: land, income, and credit . . . 94
3.4 Land Wealth Inequality . . . 97
3.4.1 From Land Area to Land Wealth . . . 97
3.4.2 The Inverse Value-Size Relationship . . . 99
3.5 Land Value, Asset Poverty, and Minimum Endowments . . . 102
3.6 Conclusion . . . 106

A Appendix to Chapter 1 . . . 111
A.1 Supplemental Tables & Figures . . . 111
A.1.1 Coffee Revenue per Farm: Construction and Sources . . . 116
A.1.2 Coffee's Share of Agricultural Value . . . 119
A.1.3 Historical Evidence: Bandit Groups and the Coffee Economy . . . 120
A.2 Robustness . . . 126
A.2.1 Persistence and Anti-Persistence . . . 127
A.2.2 Coffee and Violence . . . 142
A.2.3 Mechanisms . . . 152
A.3 Data & Measurement . . . 153

A.3.1 Data Sources & Variables . . . 154
A.3.2 Measurement Error . . . 160
A.3.3 Historical Geographic Units . . . 164

B Appendix to Chapter 2 . . . 180
B.1 Description of the data . . . 180
B.2 Derivations of the theoretical framework . . . 182
B.3 Additional Figures and Tables . . . 184

C Appendix to Chapter 3 . . . 196
C.1 Main Robustness Checks . . . 196
C.1.1 Census–Survey and Alternative Survey Comparisons . . . 196
C.1.2 Alternative Definition of Landless Households . . . 200
C.2 Robustness for Section 3.3 . . . 202
C.2.1 Quantile Effects of Land Ownership . . . 202
C.2.2 Alternative Treatment of Zero Income . . . 203
C.2.3 Cross-Country Heterogeneity . . . 205
C.2.4 Main Specification Variants . . . 206
C.2.5 Non-Response Analysis . . . 207
C.2.6 Poverty Lines . . . 209
C.2.7 Alternative Wealth Measures . . . 210
C.2.8 Landless Analysis . . . 212
C.3 Land Wealth Inequality . . . 214
C.3.1 Elasticities, Area Inequality, and the Covariance Decomposition . . . 214
C.3.2 Land use composition and land-area dispersion . . . 218
C.3.3 Crop versus Animal Income Mix and Land Area Dispersion . . . 219
C.4 Land Value, Asset Poverty, and Minimum Endowments . . . 221
C.4.1 Land as a Share of Household Wealth . . . 221
C.4.2 Income Poverty and Land Poverty: Alternative Specifications . . . 222
C.4.3 Robustness: Land Poverty and Redistribution Shares . . . 225
C.5 Definitions . . . 227
C.5.1 Ownership . . . 227
C.5.2 Income Aggregate . . . 228

List of Tables

1.1 Persistence of Conflict Incidence . . . 25
1.2 Persistence of Conflict Intensity . . . 27
1.3 Liberal Voting and Conflict Persistence . . . 29
1.4 Short-Run Determinants: Violent Deaths . . . 33
1.5 Conceptual Framework: Economic Incentives and Fighting Feasibility . . . 34
1.6 Mean Violence: Any Coffee vs None × Three Proximity Groups . . . 34
1.7 Bandit Location: Coffee and Armed Group Presence . . . 36

2.1 Average land size and input use . . . 56
2.2 Rural population and agricultural sector by country . . . 58
2.3 Land Inequality Estimates: LAC, Asia, MENA, SSA . . . 60
2.4 Average Farm Size per Population Share (ha) . . . 62
2.5 Rural-Urban Theil Decomposition . . . 64
2.6 Agricultural - Non. Agricultural Theil Decomposition . . . 64
2.7 Agricultural Sector Regional Theil Decomposition . . . 66

3.1 Database Summary: Survey Sources and Key Characteristics . . . 91
3.2 Income, Wealth, and Loans Regressions . . . 95

A.1 Coffee Production by Department, Selected Years (sacks of 62.5 kg) . . . 111
A.2 Bandit Groups by Department, 1962 . . . 113
A.3 International Intentional Death Rates per 100,000 (Top 5 Countries) . . . 113
A.4 Coffee × Distance to Early Violence Hub: OLS and PPML . . . 115
A.5 Role of Bandits: Coffee × Armed Group Presence . . . 115
A.6 Coffee Revenue per Farm and Number of Farms by Department, Selected Years 118
A.7 Coffee vs. Other Agricultural Products by Colombian Department (1955) . . . 119
A.8 Robustness: Early Violence Components (LPM) . . . 127
A.9 Robustness: Alternative Intensity Measure . . . 128
A.10 Robustness: Extensive vs. Intensive Margin of Late Violence . . . 130
A.11 Robustness: Extensive vs. Intensive Margin — Alternative Measure . . . 131
A.12 Frontier Subsample: New Frontier (Antioquia, Caldas, Tolima, Valle) . . . 137
A.13 Frontier Subsample: Rest of Colombia . . . 138
A.14 Liberal 1856 and Early Violence: Reduced Form . . . 139

A.15 Liberal Voting and Violence: PPML (Antioquia) 141
A.16 Robustness: Alternate Measures of Conflict 143
A.17 Robustness: Alternate Coffee Measures 144
A.18 Robustness: Frontier Subsamples 146
A.19 Spatial Robustness of Coffee Effect (Standard Errors) 147
A.20 Spatial Unit Root Tests on Benchmark Residuals 148
A.21 Robustness: Winsorization of Outliers 151
A.22 Data Sources and Variables 154
A.23 Source Concordance: Police Reports (1962) vs. Guzman reports Clashes (1958–63) 162

B.1 Description of Agricultural Censuses by Country 192
B.2 Number and area of farms by farm size range 193
B.3 Definition of Rural Population by Country 194
B.4 Agricultural income and farm size 195
B.5 Agricultural wage and farm size 195
B.6 Interquartile range of agricultural income and farm size 195

C.1 Robustness: Alternative Treatment of Zero Income 203
C.2 Robustness: Income and Consumption Returns to Land 204
C.3 Robustness: Main Specification Variants 206
C.4 Robustness: Extensive Margin of Income Result: Poverty Lines 209
C.5 Robustness: Alternative Wealth Measures 210
C.6 Landless Analysis: Ownership, Income, and Agricultural Share 212
C.7 Shapley decomposition of the covariance effect 216
C.8 Country-level elasticity robustness checks 217
C.9 Ownership questions and definitions by country 227
C.10 Income aggregate components: Guatemala (2000) 229

List of Figures

1.1 GDP Growth and Coffee Export Dependence in Colombia, 1870–1970 . . . 12
1.2 The Two Colombias: Coffee Production and Land Shares by Frontier, 1874–1955 14
1.3 Long-Run Development by Historical Group: State Presence, Coffee Frontier, and Other Municipalities . . . 16
1.4 Intentional Deaths by Frontier Group, 1946–1960 . . . 20
1.5 Event Study: Liberal vs. Conservative Municipalities (Antioquia) . . . 30
1.6 OLS Predicted Violent Deaths by Coffee Status and Distance to Hub . . . 35
1.7 Average Marginal Effect of Coffee on Violent Deaths by Bandit Presence . . . 37
1.8 Revenue per Coffee Farm Relative to Agricultural Wage by Frontier (1924–1955) 38

2.1 Countries covered in the chapter . . . 54
2.2 Shares of farms by Population Groups . . . 62
2.3 Household Level Decomposition of Income Sources . . . 67
2.4 Aggregate Agricultural Productivity, Farm Size, and Rural Poverty . . . 72
2.5 Share of land and labor by farm size . . . 73
2.6 Labor intensity by farm size . . . 74
2.7 Output per worker and farm size . . . 75
2.8 Land yields and farm size . . . 75
2.9 Relationship between household agricultural income and farm size . . . 76
2.10 Relationship between interquartile range, 25th, and 75th percentile of agricultural income and farm size . . . 77
2.11 Relationship between interquartile of range of output per worker and interquartile of agricultural income . . . 78

3.1 Land Inequality Across Measures . . . 98
3.2 The inverse value–size relationship and land inequality . . . 100
3.3 Asset Poverty and Potential Redistribution . . . 104

A.1 Coffee Production Intensity by Municipality, 1944 . . . 112
A.2 Intentional Deaths by Department, 1946–1960 . . . 114
A.3 Threshold-Varying Persistence: Violent Deaths . . . 133
A.4 Threshold-Varying Persistence: Clashes . . . 134
A.5 Quantile Regression: Deaths > 0 Sample . . . 135

A.6 Quantile Regression: Deaths > 0 or Battles > 0 Sample 136
A.7 Balance Test: Covariates by Liberal Treatment (Antioquia) 141
A.8 Leave-One-Department-Out: Coffee Coefficient on Violent Deaths 145
A.9 Shape of the Coffee-Violence Relationship 150
A.10 PPML Predicted Violent Deaths by Coffee Status and Distance to Hub . . . 152
A.11 Violence Trends by Frontier Type: Deaths vs. Clashes 160
A.12 Violence Trends by Department: Deaths vs. Clashes 161
A.13 Source Concordance: Deaths (1962) vs. Clashes (1958–63) 163
A.14 HGU aggregation: the case of El Águila splitting from Ansermanuevo . . . 165

B.1 Individual Level Decomposition of Income Sources 184
B.2 Share of land and labor by total farm area 185
B.3 Output per worker and total farm area 186
B.4 Land yields and total farm area 187
B.5 Output per worker and farm size 188
B.6 Yields per hectare and farm size 189
B.7 Relationship between household agricultural income and total farm area . . 190
B.8 Relationship between agricultural income Gini and farm size Gini 191

C.1 Census and survey measures of land inequality 197
C.2 Land inequality: alternative country-specific surveys and DHS 198
C.3 Alternative Definitions of Landless Households 201
C.4 Quantile regression coefficients: landowner indicator on log income 202
C.5 Land Returns by Country: Main Specification 205
C.6 Non-Response and Standardized Land Returns 208
C.7 Effects of Land and Value on Wealth by Country 211
C.8 Landowner–Landless Gap by Country: Income and Agricultural Share . . . 213
C.9 Elasticities, land-area inequality, and price dispersion 215
C.10 Land use composition and the covariance mechanism 218
C.11 Crop–animal income mix and the land-inequality mechanism 220
C.12 Plot Wealth as a Share of Total Wealth, by Country 221
C.13 Income poverty by landholding group: within-survey gaps and robustness to alternative thresholds 223
C.14 Land poverty and potential redistribution: alternative thresholds 226

General Introduction

Almost 30 years ago, Atkinson (1997) argued for the importance of 'bringing income distribution in from the cold.' It seems his call was widely heard: inequality is back in vogue. Since the 2008 Global Financial Crisis, the distribution of income and wealth has become an outsized topic in public debate and in academic circles, driven by works documenting the historical trajectories, causes, and consequences of inequality in the developed world (Atkinson, 2015; Milanovic, 2016; Piketty, 2014), among many others. The diagnostics for the developed world are now somewhat clear: higher taxation, greater redistribution, stronger labor-market institutions, and more robust public investment in education, health, and social protection. Research on income and wealth inequality in the developing world has also progressed considerably (Alvaredo et al., 2025; De Rosa et al., 2024). But the developing world is not a poorer version of the developed one, and some of the central distributional questions it faces cannot be answered with the same approach.

Around 26% of the global population is employed in agriculture, but this share is overwhelmingly concentrated in the developing world, where the agricultural labor force in many countries constitutes well over half of the overall workforce (FAO, 2024a,b). Furthermore, up to 78% of the world's extreme poor are agricultural workers (de Janvry and Sadoulet, 2020). In the developed world, agricultural land long ago ceased to figure as an important share of aggregate wealth or as a significant source of employment (Piketty and Zucman, 2014, p. 1259, 1297, 1300): it can reasonably be set aside when studying distributional questions. In the developing world, it cannot. Hundreds of millions of people depend on agricultural land as their primary source of income, their principal store of wealth, and their only form of economic security. In the Global South, the agricultural sector either still dominates economic life or its recent transformation continues to shape pressing questions of development. This thesis takes this observation as its starting point. It studies how land and agricultural structures shape inequality, poverty, productivity, and conflict across the developing world.

Chapter 1 analyzes how an agricultural commodity boom can reshape the long-run geography of civil conflict. The chapter focuses on Colombia's two largest civil wars, the Thousand Days War (1899–1902) and La Violencia (1948–1964), and on the coffee boom of the First Globalization that took place between them. I ask whether the geography of violence persisted across these two conflicts or whether coffee, by creating new concentrations of appropriable wealth across the Colombian countryside, displaced the locations of conflict entirely. The key intuition is that coffee rents drew armed actors into regions that had not previously been centers of violence, disrupting a long history of uniquely bipartisan civil wars.

The chapter assembles a municipality-level dataset that covers the universe of coffee and conflict data for Colombia between 1899 and 1964. Conflict intensity is measured in counts (deaths, armed clashes, bandit locations) rather than the binary indicators used in previous quantitative work on this period. These counts come from two independently compiled archival sources, the field reports of Guzmán et al. (1964) and the criminality reports of the National Police, which are cross-validated against each other. Coffee production data span 1923 to 1960 and are drawn from federation bulletins and national agricultural censuses. Because Colombian municipalities have been repeatedly split, merged, and renamed over this period, the analysis constructs historical geographic units that allow consistent spatial comparison across more than half a century of administrative change.

The empirical analysis proceeds in two steps. The first tests whether conflict persists across Colombia's two largest civil wars. Conflict incidence does persist: municipalities that experienced battles during the Thousand Days War are 8.8 percentage points more likely to see La Violencia erupt half a century later, with partisan identity, traced back to Liberal vote shares from 1856, acting as the channel of transmission. An event study around the Gómez presidency in 1950 confirms this mechanism. However, conflict intensity does not persist at all: nineteenth-century conflict measures have no predictive power over death counts during La Violencia. The second step asks what explains the variation in intensity. In a regression that includes all the dominant explanations from the historiography (state capacity, political competition, land tenure, agrarian tensions), coffee production emerges as the single strongest correlate, with a standardized coefficient more than double the next significant predictor. Once armed group presence is accounted for, coffee's direct effect drops to near zero, consistent with armed groups serving as the channel through which coffee rents translated into violence.

These findings speak to several literatures. In the study of conflict persistence (Besley and Reynal-Querol, 2014; Dincecco et al., 2019; Fearon and Laitin, 2014), the chapter provides within-country evidence that persistence and anti-persistence operate simultaneously: partisan identity transmits conflict propensity across generations, but a large economic transformation, in this case the coffee boom of the First Globalization, creates rents in new places and reshapes the geography of violence entirely. In the natural resources and conflict literature (Berman and Couttenier, 2015; Collier and Hoeffler, 2004; Dube and Vargas, 2013; McGuirk and Burke, 2020), it offers a case in which the arrival of a new commodity, rather than a price shock to an existing one, fundamentally altered conflict trajectories. And in the historiography of Colombian violence (Arocha, 1979; Bergquist, 1986; Chacón et al., 2011; Oquist, 1980; Sánchez et al., 2001), it provides the first systematic quantitative test of why coffee production was central to the intensity of La Violencia, a fact long recognized by historians but never examined against competing explanations with comprehensive data.

Chapter 2, co-authored with Margarita Gáfaro, Ana-María Ibáñez, and María Camila Ortiz, revisits one of the oldest debates in agricultural and development economics: the inverse farm

size-productivity relationship (Berry and Cline, 1979). This longstanding literature contends with one of the strongest empirical facts in the field: small farms are more productive per hectare than large ones. The finding has been used to argue for redistributive land policy for decades. While much of the modern literature debates whether the relationship holds and what drives it, this chapter asks a different question: what does it imply for income distribution? Small farms may achieve higher yields, but larger farms have higher labor productivity, and it is labor productivity that determines household income. We empirically test whether locations with larger farm sizes exhibit different rates of income inequality, and what role the agricultural sector plays in shaping overall inequality across Latin America.

To answer these questions, we harmonize agricultural census microdata from nine countries and merge them with nationally representative household employment surveys at the regional level. The censuses provide farm-level information on cultivated area, crop production, and labor inputs; the surveys provide household income disaggregated by source: own-farm income, agricultural wages, and non-agricultural earnings. To make output comparable across crops and countries, we construct revenue-based productivity measures using FAO producer prices in constant 2011 dollars.

Theil decompositions show that agricultural income inequality accounts for between 11 percent of total inequality in Uruguay and 58 percent in Bolivia, with an average of roughly 35 percent across the nine countries. A large share of this comes from the gap between agricultural and non-agricultural household incomes, not only from dispersion within the agricultural sector itself. At the regional level, we find that locations with larger average farm sizes also exhibit higher average agricultural incomes and higher agricultural wages, after controlling for soil quality and rainfall. Within regions, output per worker rises with farm size across all countries in the sample, as larger farms have higher labor productivity even where smaller farms achieve higher yields per hectare. The combination of extreme land inequality and a large share of farms that are too small to sustain high labor productivity thus has direct consequences for income distribution at the national level.

These results speak to the large literature on income inequality in Latin America (Ferreira et al., 2013; Gasparini et al., 2008; Lustig et al., 2013), which has focused on education, labor markets, and fiscal policy as the primary drivers, while evidence on the contribution of agricultural dynamics to overall income inequality remains scarce (Lagakos, 2020). The chapter also extends the debate on farm size and agricultural productivity (Berry and Cline, 1979; Foster and Rosenzweig, 2022; Sen, 1962) by documenting that the positive relationship between farm size and labor productivity translates directly into agricultural household incomes and, through them, into national income inequality.

Chapter 3, co-authored with Luis Bauluz, Yajna Govind, and Filip Novokmet, turns to a purely distributional question: how is agricultural land distributed across households in the developing world, and what dimensions of land matter most for the welfare of the farmers

who depend on it? Three out of four of the world's poorest billion individuals work in agriculture, yet there exist remarkably few systematic estimates of how land, their most important productive asset, is distributed at the ownership level. This chapter addresses this gap, establishing a new set of facts about the distribution, value, and economic role of agricultural land across seventeen developing countries.

To address these questions, we construct a harmonized household-level database covering land ownership, land value, income, credit, and wealth across seventeen developing countries in Africa, Asia, and Latin America, together accounting for roughly two-thirds of the developing world by population. The data come from household surveys rather than agricultural censuses, which is a departure from the existing cross-country literature on land inequality (Deininger and Squire, 1998; Frankema, 2010). This allows three things that census data cannot: measuring land by value rather than only by physical area, capturing ownership rather than operational holdings, and including landless households, on average 35 percent of the agricultural population, who are excluded by construction from census sampling frames.

The empirical results are organized around three sets of findings. First, land value, not land area, is what matters for household welfare: among agricultural landowners, the elasticity of income with respect to land value is 0.445, and once value is included, the coefficient on land area turns negative. The same pattern holds for wealth and the size of outstanding loans. Second, when land inequality is measured by value and includes the landless, the Gini coefficient reaches approximately 0.77, at least as concentrated as the global distribution of wealth. Cross-country differences in land inequality also narrow substantially under value-based measures, reflecting an inverse relationship between the size of landholdings and their value per hectare. Third, we introduce a measure of land asset poverty: the share of agricultural households below a threshold of 1,000 PPP dollars in land value ranges from under 15 percent in China and Vietnam to over 60 percent in Malawi and Ethiopia. The aggregate land value needed to close these gaps is modest relative to total land wealth in most countries. Land deprivation is widespread in incidence but small in aggregate magnitude.

These findings speak to the study of the economic lives of the poor (Banerjee and Duflo, 2007), extending it to their most important rural asset, and to the cross-country measurement of land inequality (Deininger and Squire, 1998; Easterly, 2007; Frankema, 2010), where moving from area to value and from landowners to the full agricultural population yields a substantially more concentrated picture than census-based data have suggested. The chapter also contributes to the literature on asset poverty and poverty traps (Balboni et al., 2022; Banerjee and Newman, 1993) by quantifying both the incidence and the modest aggregate cost of land deprivation across the developing world.

References

Alvaredo, F., M. De Rosa, I. Flores, and M. Morgan (2025). The Inequality (or the Growth) We Measure: Data Gaps and the Distribution of Incomes. *Economía LACEA Journal 24*(1).

Arocha, J. (1979). *La violencia en el Quindío: determinantes ecológicos y económicos del homicidio en un municipio caficultor*. Ediciones Tercer Mundo.

Atkinson, A. B. (1997). Bringing Income Distribution in From the Cold. *The Economic Journal 107*(441), 297–321.

Atkinson, A. B. (2015). *Inequality: What Can Be Done?* Harvard University Press.

Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2022). Why do people stay poor? *The Quarterly Journal of Economics 137*(2), 785–844.

Banerjee, A. V. and E. Duflo (2007). The economic lives of the poor. *Journal of economic perspectives 21*(1), 141–167.

Banerjee, A. V. and A. F. Newman (1993). Occupational choice and the process of development. *Journal of Political Economy 101*(2), 274–298.

Bergquist, C. (1986). *Labor in Latin America: Comparative Essays on Chile, Argentina, Venezuela, and Colombia* (1st US - 1st Printing edition ed.). Stanford, California: Stanford Univ Pr.

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

Berry, R. A. and W. R. Cline (1979). *Agrarian Structure and Productivity in Developing Countries: A Study Prepared for the International Labour Office Within the Framework of the World Employment Programme*. Johns Hopkins University Press.

Besley, T. and M. Reynal-Querol (2014). The Legacy of Historical Conflict: Evidence from Africa. *American Political Science Review 108*(2), 319–336.

Chacón, M., J. A. Robinson, and R. Torvik (2011). When is Democracy an Equilibrium? Theory and Evidence from Colombia's La Violencia. *Journal of Conflict Resolution 55*(3), 366–396.

Collier, P. and A. Hoeffler (2004). Greed and grievance in civil war. *Oxford Economic Papers 56*(4), 563–595.

de Janvry, A. and E. Sadoulet (2020). Using agriculture for development: Supply- and demand-side approaches. *World Development 133*, 105003.

De Rosa, M., I. Flores, and M. Morgan (2024). More unequal or not as rich? Revisiting the Latin American exception. *World Development 184*, 106737.

Deininger, K. and L. Squire (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics 57*(2), 259–287.

Dincecco, M., J. Fenske, and M. G. Onorato (2019). Is Africa Different? Historical Conflict and State Development. *Economic History of Developing Regions 34*(2), 209–250. _eprint: https://doi.org/10.1080/20780389.2019.1586528.

Dube, O. and J. F. Vargas (2013). Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *The Review of Economic Studies 80*(4), 1384–1421.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics 84*(2), 755–776.

FAO (2024a). *Employment indicators 2000–2022. October 2024 update*. Number No. 92 in FAOSTAT Analytical Briefs. Rome, Italy: FAO.

FAO (2024b). *World Food and Agriculture – Statistical Yearbook 2024*. Number 2024 in FAO Statistical Yearbook – World Food and Agriculture. Rome, Italy: FAO.

Fearon, J. D. and D. Laitin (2014). Does Contemporary Armed Conflict Have 'Deep Historical Roots'?

Ferreira, F. H., J. Messina, J. Rigolini, L.-F. López-Calva, M. A. Lugo, and R. Vakis (2013). *Economic Mobility and the Rise of the Latin American Middle Class*. The World Bank.

Foster, A. D. and M. R. Rosenzweig (2022). Are there too many farms in the world? labor market transaction costs, machine capacities, and optimal farm size. *Journal of Political Economy 130*(3), 636–680.

Frankema, E. (2010). The colonial roots of land inequality: geography, factor endowments, or institutions? *The Economic History Review 63*(2), 418–451.

Gasparini, L., M. Horenstein, E. Molina, and S. Olivieri (2008). Income polarization in latin america: Patterns and links with institutions and conflict. *Oxford Development Studies 36*(4), 461–484.

Guzmán, G., O. Fals Borda, and E. Umaña Luna (1964). *La violencia en Colombia: Estudio de un proceso social. Tomo II* (Segunda edición ed.). Colección Actores de la violencia en Colombia. Bogotá, Colombia: Ediciones Tercer Mundo. Number: 11.

Lagakos, D. (2020). Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities? *Journal of Economic Perspectives 34*(3), 174–192.

Lustig, N., L. F. Lopez-Calva, and E. Ortiz-Juarez (2013). Declining inequality in latin america in the 2000s: The cases of argentina, brazil, and mexico. *World Development 44*, 129–141.

McGuirk, E. and M. Burke (2020). The Economic Origins of Conflict in Africa. *Journal of Political Economy 128*(10), 3940–3997.

Milanovic, B. (2016). *Global Inequality: A New Approach for the Age of Globalization*. Harvard University Press.

Oquist, P. H. (1980). *Violence, conflict, and politics in Colombia*. New York : Academic Press.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T. and G. Zucman (2014). Capital is Back: Wealth-Income Ratios in Rich Countries 1700–2010. *The Quarterly Journal of Economics 129*(3), 1255–1310.

Sen, A. K. (1962). An aspect of indian agriculture. *Economic Weekly 14*(4-6), 243–246.

Sánchez, G., D. Meertens, and A. Hynds (2001). *Bandits, Peasants, and Politics: The Case of ""La Violencia"" in Colombia*. Austin: University of Texas Press.

Chapter 1

(Un-)Persistent Conflict? The Effects of the First Globalization Coffee Boom in Colombia

1.1 Introduction

Colombia is a violent country and has been for a long time. This is due in no small part to the low-intensity armed conflict that began with the formation of the FARC in 1964 and lasted until their demobilization in 2016. That conflict was itself rooted in the bloodiest civil war in Colombian history, *La Violencia* (1948–1964), which claimed an estimated 200,000 lives. But Colombian violence long predates *La Violencia*. As (Mazzuca and Robinson, 2009) note, "*Colombia's nineteenth century was politically chaotic even by Hispanic American standards: the record includes nine national civil wars, dozens of local revolts and mutinies, material destruction equivalent to the loss of several years of economic output, and at least 250,000 deaths due to political violence.*" If conflict has been a defining feature of Colombia for the past two hundred years, have the determinants of that conflict persisted across time, or have they evolved?

This paper investigates this question by examining Colombia's two largest civil conflicts (the Thousand Days War, 1899–1902, and *La Violencia*, 1948–1964) and the coffee boom that took place between them. Coffee went from a negligible export in the 1870s to the country's central commodity by the 1930s, accounting for up to 80% of total exports and concentrating in a newly settled agricultural frontier. I compile a comprehensive municipality-level dataset that covers, to my knowledge, the universe of conflict and coffee data for this period, combining

newly digitized archival data with extensive secondary sources spanning more than a century. I use this data to test whether nineteenth-century conflict predicts mid-twentieth-century violence, and to identify what factors best account for the large variation in violence intensity across municipalities during *La Violencia*.

The paper yields two sets of results. First, the geography of conflict *incidence* persists across centuries: municipalities that experienced battles during the Thousand Days War are significantly more likely to see *La Violencia* break out half a century later. Partisan identity is most strongly associated with this persistence. Liberal strongholds in the nineteenth century remained Liberal into the mid-twentieth century and experienced more violence during the early phase of *La Violencia*, when the conflict was driven primarily by state-sponsored partisan repression under the Conservative presidency of Laureano Gómez (1950–1953). The cross-sectional evidence, however, may reflect omitted municipal characteristics that correlate with both partisanship and conflict. To address this, I exploit annual violence data from 99 municipalities in the department of Antioquia to estimate an event study around Gómez coming to power. Violence in Liberal municipalities increases relative to Conservative municipalities after 1950, with no differential trend beforehand, consistent with partisan identity driving the onset of conflict rather than pre-existing differences between municipalities.

Second, conflict *intensity* does not persist. None of the nineteenth-century conflict measures predicts the number of violent deaths during the later phase of *La Violencia*, a result that holds across alternative outcome measures, estimators, and subsamples. The variation in intensity across municipalities is large (death rates in the most affected areas exceeded those of any country in the world at the time), yet historical conflict patterns cannot account for it. In a horse-race specification that includes all leading explanations from the historiography, coffee production is the strongest correlate of violence intensity, with an association more than twice the magnitude of the only other significant predictor (Liberal vote share). All other competing explanations (state capacity, political competition, and agrarian tensions) are statistically insignificant.

I interpret this result through a framework that combines two ingredients: coffee rents provide the economic incentive for armed activity, while proximity to existing conflict hubs provides the organizational capacity to exploit those rents. Where both conditions are met, violence should be most intense; where either is absent, it should be lower. The evidence is consistent with this logic. The association between coffee and violence is strongest near early *La Violencia* hubs and decays with distance: at the hub itself, coffee-producing municipalities average significantly more violent deaths than non-producers, and this gap closes gradually as distance increases. Coffee municipalities are also significantly more likely to host armed group leaders, and the marginal effect of coffee on violence is near zero where armed groups are absent but large where they are present. Once armed group presence is accounted for, coffee's direct association with violence drops to near zero, consistent with armed groups

serving as the channel through which coffee rents translate into conflict.

I show that conflict location persists across centuries in Colombia, transmitted through partisan identity, but that the intensity of violence follows a different logic. Violence during *La Violencia* intensified most in municipalities on a new coffee frontier, where rents from a recently arrived cash crop sustained the armed groups that drove this violence. The geography of conflict, these results suggest, is shaped less by where violence has been than by where the resources to sustain it are located.

These findings contribute to three literatures. First, I add a subnational case to the small but growing literature on conflict persistence. Existing work has documented that historical conflict predicts contemporary conflict across regions in Africa, Asia, and the Middle East (Besley and Reynal-Querol, 2014; Dincecco et al., 2019; Fearon and Laitin, 2014). These studies have relied primarily on cross-country or cross-regional comparisons. I contribute municipality-level evidence from within a single country, which allows me to do two things the existing literature has not always been able to do: identify a mechanism for both persistence and for anti-persistence. The results also speak to the broader persistence and anti-persistence debates in economic history (Arroyo-Abad and Maurer, 2021; Cioni et al., 2021; Conley and Kelly, 2025; Voth, 2021). These debates have centered on whether historical patterns endure or reverse over time. I offer a case in which both forces operate simultaneously: persistence is common, but a large economic transformation, such as a First Globalization commodity boom can create new rents in new places and alter the trajectory of conflict.

Second, I contribute to the literature on natural resources and conflict. A large body of careful, causally identified work has established that natural resources of different types can drive or reduce violence through opportunity cost and rapacity mechanisms (Berman and Couttenier, 2015; Collier and Hoeffler, 2004; Dube and Vargas, 2013; McGuirk and Burke, 2020). The focus of this literature, however, is entirely contemporary, and it therefore takes existing resource industries as given. I study instead how the arrival of a new natural resource industry in a previously marginal area can bring violence by expanding the frontier of conflict. The coffee boom created, for the first time, economic rents large enough to sustain armed activity in regions that had no prior history of conflict, expanding the frontier of violence into the new agricultural frontier. This result connects to a related literature on the First Globalization and its effects on developing economies, which has documented a wide range of consequences including changes in inequality (Arroyo-Abad, 2013; Williamson, 1999, 2002), education (Musacchio et al., 2014), and labor coercion (Saleh, 2024). To my knowledge, a link between the First Globalization and the trajectory of conflict has not been explored. I add that the commodity booms of this period could also change the geography of violence, by creating industries whose rents would fuel armed activity for decades.

Third, I contribute to the historiography of Colombian violence. Historians have long noted that violence was more extreme in coffee regions. (Arocha, 1979), (Oquist, 1980),

(Sánchez et al., 2001), and most importantly (Bergquist, 1986a) have provided accounts linking coffee and violence, but these are based on narrative histories or case studies of individual municipalities rather than systematic analysis of how and why coffee mattered relative to other theories. The two main quantitative studies on *La Violencia* are (Chacón et al., 2011) and (Chacón and Sánchez, 2003). The first focuses exclusively on the early period and the role of political competition. The second provides correlational analysis of both periods and finds similar overarching patterns, but does not connect the competing theories within a unified framework, does not identify coffee as the principal factor, and does not provide a generalized account of why the geography of violence changed over time. I am able to use newly digitized count data (from (Roldán, 2002), (Policía Nacional de Colombia, 1965), and (Guzmán et al., 1964)), which permits analysis of the intensity of violence rather than relying on binary indicators alone, building a more accurate picture of the determinants of violence across both periods. Coffee production emerges as the strongest predictor of violence intensity, conditional on all competing explanations. This is a major empirical fact that any account of *La Violencia* must explain, and this paper provides a first systematic attempt to do so.

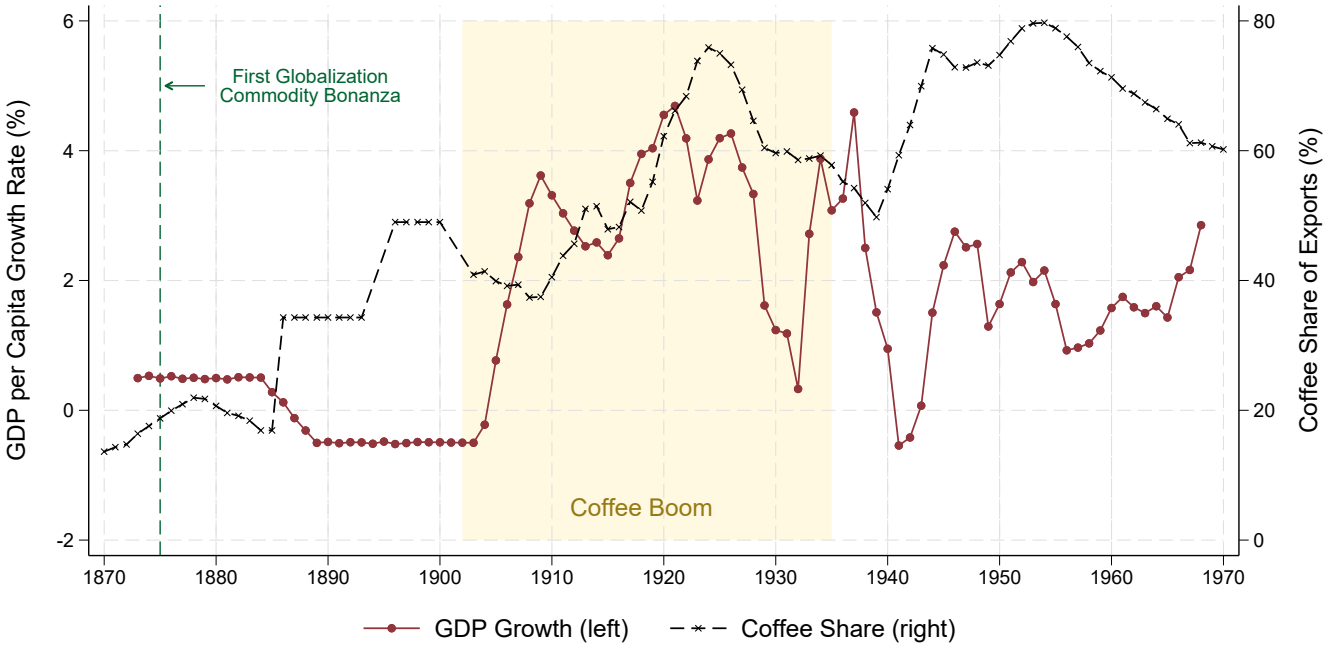
The remainder of the paper proceeds as follows. Section 1.2 presents the historical context: the coffee boom, the nineteenth-century civil wars, and *La Violencia*. Section 3.2 describes the data. Section 1.4 tests whether conflict persists across centuries and identifies partisan identity as the channel. Section 1.5 tests whether coffee production explains the intensity of *La Violencia* and explores the mechanisms. Section 3.6 concludes.

1.2 Historical Context & Analysis

1.2.1 First Globalization and Coffee Boom, 1875-1955

After independence in 1821, Colombia was among the poorest countries in Latin America (Bulmer-Thomas, 2003). Three Andean mountain ranges divided the country into isolated regions, inhibiting economic integration and state development (Safford and Palacios, 2002). The main export, gold, remained at colonial production levels despite a growing population, severely limiting revenue and economic advancement (Bushnell, 1993; Deas, 1982). For the first half-century of the republic, Colombia had poor internal connections, negligible global market integration, and no viable source of national wealth. This began to change with the onset of the First Globalization, as surging international demand for tropical commodities opened new possibilities for peripheral economies like Colombia's. Throughout the nineteenth century, entrepreneurs experimented with various cash crops (tobacco, quinine, and indigo) in what (Ocampo, 1984) characterizes as a "commodity bonanza" marked by repeated boom-and-bust cycles. Yet none of these crops proved durable enough to anchor a national export sector, and it was not until the late 1880s that coffee emerged as the leading export, a position it maintained for nearly a century.

Figure 1.1. GDP Growth and Coffee Export Dependence in Colombia, 1870–1970



Note: Five-year moving averages. The left axis plots GDP per capita growth; the right axis plots coffee's share of total exports. The shaded region marks the coffee boom period.
Sources: GDP per capita from Maddison Project Database (Bolt and van Zanden, 2025). Coffee share of total exports: (Ocampo, 2011, p.357) for 1870–1905; (Urrutia-Montoya et al., 2000, p.12) for 1905–1970.

Figure 1.1 plots GDP per capita from the Maddison Project Database (Bolt and van Zanden, 2025) alongside coffee's share of total exports (Urrutia-Montoya et al., 2000) to illustrate the scale of this transformation. During the last quarter of the nineteenth century, Colombian entrepreneurs were actively searching for a viable export base, yet the economy stagnated: GDP per capita growth hovered near zero, and although coffee's export share began rising toward the end of the century, volumes remained unstable and the crop had not yet established clear dominance. This changed after 1900, when political stability and increased foreign demand for coffee[1] ushered in the first period of sustained economic growth in Colombian history (Bergquist, 1986b). As coffee's export share surged past 60% by 1920, GDP per capita growth climbed to 3–5% annually. The coffee economy's emergence is usually credited for pushing the Colombian economy into "modern" economic growth (Arango, 1977; Bejarano, 1979, 1987; Palacios, 2002; Vázquez, 1955). The visible downturns in the series correspond to the World Wars and the Great Depression, but the broader pattern is straightforward: by 1925, coffee reached 75% of total export value and maintained a 60–80% share until the 1970s. Colombia became, in the words of (Palacios, 2002, p.225), the country "*where coffee most dominated the export sector, followed by Ethiopia, Rwanda, Uganda, and Haiti*."

[1] (Jiménez, 1995) argues that coffee in particular saw a great surge in global demand because of the Second Industrial Revolution, especially in the United States, where laborers' need for energy during strict work schedules, combined with their integration into a consumer goods economy, fueled mass consumption of stimulant beverages.

Coffee's expansion in Colombia had a distinct spatial dimension: the boom was geographically uneven, concentrating in regions that had been economically marginal throughout the nineteenth century. At the outset of the First Globalization, the primary coffee-producing departments[2] were Cundinamarca, Norte de Santander, and Santander, regions that had persisted from colonial-era hacienda models or risen immediately after independence (Palacios, 2002, Chs. 1,3). I refer to these departments as the *old frontier*. Following the turn of the century, a new frontier of coffee production emerged in previously marginal regions to the west and southwest: Antioquia, Caldas,[3] Valle del Cauca, and Tolima.[4] These departments, peripheral to Colombia's historical exporting sector, became the center of the coffee economy during the first half of the twentieth century. This geographical typology is not new. Luis Eduardo Nieto Arteta, Colombia's first economic historian, captured this spatial bifurcation in his depiction of *"Dos Colombias"* [Two Colombias]: one modern, entrepreneurial, and integrated with global markets through coffee, and the other stagnant, trapped in older colonial-feudal economic structures (Nieto Arteta, 1958).[5] The effects of this coffee-led transformation on the new frontier were large, sustained, and have been the subject of extensive scholarship.[6] Appendix Figure A.1 plots the spatial distribution of 1944 coffee production (sacks of 62.5 kg) across municipalities, with the new frontier outlined in green and the old frontier in red, illustrating the concentration of production in the western departments.

Figure 1.2 documents the scale of this spatial reversal. Panel (a) plots the share of national coffee production by frontier group from 1874 to 1955. In 1874, the old frontier accounted for roughly 95% of all coffee produced in Colombia, with the new frontier contributing less than 5%. By 1913, the two groups had converged to near parity, and from 1924 onward, the new frontier consistently produced 70–80% of the national total. Panel (b) turns to coffee land allocation using national coffee census data available from 1924. By this date, the new frontier already held approximately 67% of all coffee land in the country, a share that remained stable through 1955 even as total national coffee land nearly tripled, from roughly 200,000 to 500,000 hectares. The displacement was thus not a gradual rebalancing but a wholesale reversal: the regions that had been economically irrelevant in the nineteenth century became

[2]The closest equivalent to a U.S. county or French canton is a Colombian municipality. These municipalities belong to departments, the next highest level of administration, comparable to a U.S. state or French département.

[3]This region separated in the 1960s into three distinct departments: Caldas, Quindío, and Risaralda. Historiographically, it is common to call the three departments "Viejo Caldas," but for consistency, I continue to call it Caldas.

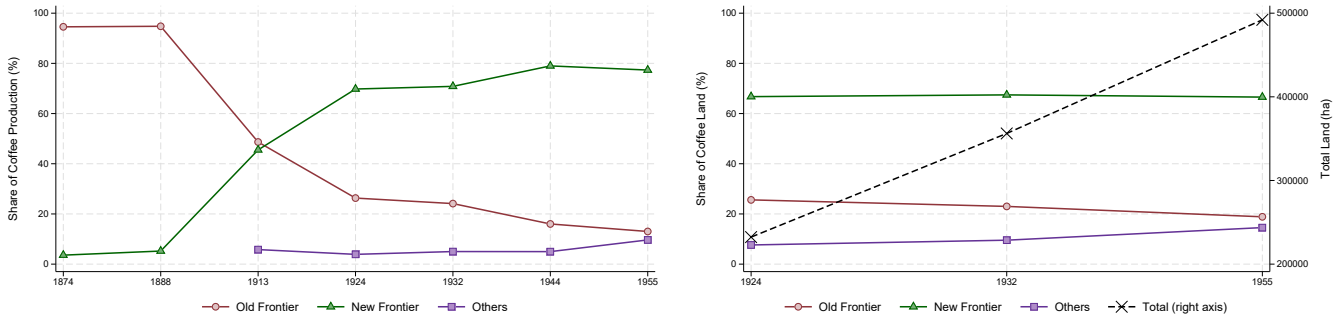
[4]The one exception to the peripheral characterization is the gold industry in Antioquia. However, this industry was more directly connected with mining and early industry than with agriculture; see (Brew, 1977).

[5]Despite critiques of Nieto Arteta's sometimes exaggerated and over-optimistic social interpretations (see (Cataño, 2012; Restrepo, 2001)), his fundamental insight regarding the spatial bifurcation of the Colombian economy remains true.

[6]Research covers migration (Parsons, 1968), coffee's role in industrialization (Arango, 1977; Vázquez, 1955), large coffee estates (Deas, 1978; Jimenez, 1989), socio-political history (Palacios, 2002), and its centrality in modern economic growth (Bejarano, 1979, 1987). Recent econometric work includes studies on land conflicts (Sánchez et al., 2010), education (Fuentes-Vásquez and España-Eljaiek, 2023), industrialization and human capital (Uribe-Castro, 2021), and colonization processes (Nugent and Robinson, 2010).

the heart of Colombia's most important industry.

Figure 1.2. The Two Colombias: Coffee Production and Land Shares by Frontier, 1874–1955



(a) Share of Coffee Production

(b) Share of Coffee Land

Note: "Old Frontier" comprises Cundinamarca, Norte de Santander, and Santander. "New Frontier" comprises Antioquia, Caldas, Tolima, and Valle del Cauca. "Others" includes Bolívar, Boyacá, Cauca, Huila, Magdalena, and Nariño.
Sources: Panel (a): Author's calculations based on (Beyer, 1947) (p.388), (Arango, 1977) (Cuadro no. II.8), (Palacios, 2002) (p.21), and (Ocampo, 2011) (p.279–280, 357, 369) for 1874-1913 and (Federación Nacional de Cafeteros de Colombia, 1933, 1945; Monsalve, 1927; United Nations. Economic Commission for Latin America, 1958) for 1924-1955 . See Table A.1 in the Appendix for raw production values and construction details. Panel (b): Author's calculations based on national coffee censuses (Federación Nacional de Cafeteros de Colombia, 1933; Monsalve, 1927; United Nations. Economic Commission for Latin America, 1958). Total coffee land in hectares on right axis.

Long-Run Effect of Coffee Boom on *New Frontier*

A large body of work in the political economy of development documents how colonial-era activities shape modern outcomes.[7] A common finding in this literature is that early state presence has path-dependent consequences: places where colonial institutions took root have a large effect on outcomes today. In Colombia specifically, (Faguet et al., 2024) show that municipalities with colonial *encomiendas* have better long-run outcomes, reinforcing the view that initial state formation carries lasting advantages.

Less well understood is what happens when a large commodity boom drives internal reorganization and the emergence of wealth into regions that had no prior state presence. The Colombian coffee expansion offers a setting to examine this: the new frontier departments had been economically and institutionally marginal throughout the colonial period, yet were transformed within a few decades by the coffee economy. To assess the scope of this transformation, I compare long-run outcomes across three mutually exclusive groups of municipalities: those with documented state presence by 1794 (García Jimeno, 2005), those in the new frontier with any coffee cultivation, and all remaining municipalities (no early state presence and no coffee).

[7] See, among others, (Dell, 2010) on the mining *mita* in Peru; (Bruhn and Gallego, 2012) on colonial activities across Latin America; (Dell and Olken, 2020) on the Dutch Cultivation System in Java; (Lowes and Montero, 2021) on concessions in the Congo; and (Arias and Flores-Peregrina, 2024) on colonial haciendas and public goods provision.

Figure 1.3 presents normalized group means (each bar represents a group's average relative to the national mean) across four historical periods. In the colonial panel, the pattern is as expected: municipalities with early state presence have population levels well above the national average, while coffee frontier locations and the rest of the country are indistinguishable and far below. The coffee boom panel captures the effect of the First Globalization previously discussed: new frontier municipalities see a sharp increase across demographic and economic indicators, narrowing the gap with state-presence locations. Compared to colonial centers, coffee regions have undistinguishable literacy, public employees per capita, while now exhibiting higher revenues per capita and population during the peak of the coffee boom. By the modern outcomes panel, coffee frontier municipalities have converged with the historically advantaged state-presence group. This pattern is consistent with what Colombian historiography has long emphasized: the coffee economy was the principal engine of twentieth-century development in these regions, generating a re-reversal of fortune from initial marginality to economic prominence.

The violence panel, however, presents a different picture. Wars in the nineteenth century (battles during the civil conflicts that preceded the coffee boom) concentrate in state-presence municipalities, consistent with these locations' role as centers of political power and nascent state formation. Yet violence during *La Violencia* (1948–1965), the most violent internal conflict in Colombian history, concentrates in the coffee frontier. The same regions that experienced rapid economic convergence also became the epicenter of mid-twentieth-century violence. Understanding the history of conflict in post-independence Colombia is thus necessary to understand the geographical shift in violence.

1.2.2 From 19th Century Civil Wars to *La Violencia*, 1948-1965

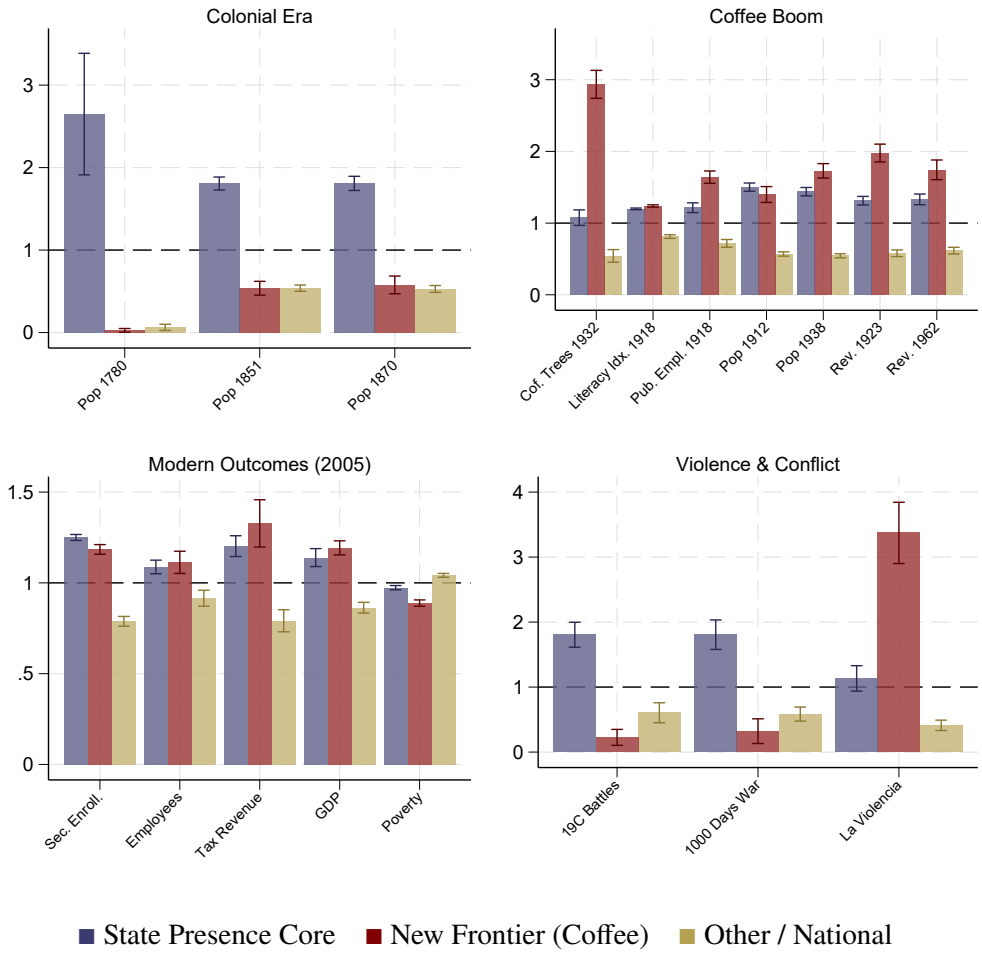
Nineteenth Century Civil Wars

Colombia's challenging geography weakened any consistent process of state formation, fueling decades of political instability (Mazzuca, 2021). This instability was channeled through a partisan divide between the Liberal and Conservative parties, which dominated Colombian politics from independence onward (Bushnell, 1993)[8]. The two parties' competition culminated in two principal periods of partisan control: the First Liberal Republic (1863–1885) under Liberal leadership and the Regeneraci'on (1886–1900) led by the Conservatives (Delpar, 1981; Park, 1985). The polarized political system and lack of monopoly of violence led to numerous series of revolts, insurrections, and civil wars throughout the century.[9] As one contemporary politician observed, the absence of viable national industries had made civil war and politics the country's principal occupations (Wallis, 1919, p.51).

[8] *Liberal* and *Conservative* are terms that arose more clearly toward the end of the 1840's. At the outset of Independence, the combination of support for (i) federalist government and (ii) separation of church and state, would entail support for *Liberals* while the opposite would signal support for *Conservatives*.

[9] Major conflicts occurred in 1839–1842, 1851, 1854, 1860, 1876–1877, 1884–1885, and 1895.

Figure 1.3. Long-Run Development by Historical Group: State Presence, Coffee Frontier, and Other Municipalities



Note: Each bar represents a group's mean divided by the national mean; the dashed line at 1.0 indicates the national average. Error bars show ±1 standard error. "State Presence Core" includes municipalities with documented colonial state presence by 1794. "New Frontier (Coffee)" includes municipalities in Antioquia, Caldas, Tolima, and Valle del Cauca with coffee cultivation. "Other" includes all remaining municipalities. *n* (SP Core / New Frontier / Other) = 377 / 114 / 545. The four panels correspond to: Colonial period variables, Transition (coffee boom era), Modern Outcomes, and Violence.
Sources: Author's calculations. See data section for variable definitions and sources.

This cycle of partisan conflict culminated in the Thousand Days' War (1899–1902), the deadliest civil war in Colombian history up to that point. During the end of the century's Conservative government, the Liberal Party had been largely shut out of formal power, holding just one legislative seat during the 1890s (Jaramillo Castillo, 1991). This sort of exclusionary practice led to widespread insurrection along party lines lasting three years. Battles involved large partisan armies in confrontations that decided clear lines of war (Palau, 1900). While the Conservative government nominally exited the war as the victor, the entire Colombian territory suffered a great demographic and economic shock due to the war.[10]

Yet the war's resolution brought a lasting institutional break. The end of the war produced

[10]For example, the war famously resulted in the secession of Panama from Colombia, aided by U.S. intervention.

a bipartisan agreement which coalesced around an export-import coalition oriented toward (coffee) export-led growth (Bergquist, 1986b; Jiménez, 1985). Both (Fergusson and Vargas, 2023) and (Mazzuca and Robinson, 2009) emphasize this pattern: the absence of credible power-sharing had sustained nineteenth-century conflict, and its establishment after the Thousand Days' War initiated a period of political stability hitherto unseen in Colombian history.

La Violencia, 1948–1965

The bipartisan settlement that followed the Thousand Days' War held. For nearly half a century, Colombia experienced no partisan warfare of the kind that had defined the nineteenth century. Other kinds of social tensions persisted (land conflicts between colonizers and large landowners (LeGrand, 1986), labor disputes[11]), but these did not escalate into organized Liberal-Conservative armed conflict.

This stability ended on April 9, 1948, with the assassination of Jorge Eli'ecer Gait'an, the popular presidential candidate. The immediate aftermath, widespread rioting in Bogotá known as *El Bogotazo*, is conventionally taken as the starting point of *La Violencia*. The "official story" (Oquist, 1980) holds that Gait'an's assassination triggered a wave of partisan warfare between Liberals and Conservatives that escalated through 1953, was temporarily contained under the military dictatorship of Gustavo Rojas Pinilla (1953–1956), and re-surged until the formation of a bipartisan National Front coalition in 1958 which would alternate the government between the two parties every four years during a period of sixteen years. By 1964, the conflict had evolved into organized guerrilla warfare, marked by the formation of the FARC.

A large body of regional historiography paints a different picture.[12] Most accounts date the true outbreak of sustained violence not to the Bogotazo but to November 27, 1949, when Conservative hardliner Laureano G'omez assumed the presidency.[13] What followed was systematic repression of the Liberal Party through both national and local state apparatus. Unlike the nineteenth-century civil wars, this violence was not fought between organized armies along river valleys and strategic cities. It was hyper-localized: Conservative state officials, police, and informal mercenaries carried out targeted repression at the municipal level, often to settle local disputes over patronage and revenue under the cover of partisan conflict (Ortiz Sarmiento, 1995; Roldán, 2002). However, the underlying dynamic still resembled

[11] A well known example is the 1928 banana workers' strike in Ci'enaga, Magdalena, in which the Colombian army fired on striking United Fruit Company workers, made famous by One Hundred Years of Solitude (García Márquez, 1967).

[12] See, among others, (Arocha, 1979; Henderson, 2010; Ortiz Sarmiento, 1995; Roldán, 2002; Sánchez et al., 2001). For example, (Roldán, 2002, p. 10) argues, the official account obscures a violence that was "remarkably limited in scope and far more varied in impulse" than the generalized partisan narrative suggests.

[13] Mariano Ospina P'erez, president since 1946, had shut down Congress in 1949. The Liberal candidate subsequently refused to participate in elections, citing a lack of safety guarantees, leaving G'omez unopposed (Henderson, 2010).

the nineteenth-century pattern: the exclusion of one party from governance by another and subsequent violent repression. Hence, the outbreak of *La Violencia* followed largely partisan biases: Liberal hotspots were targets of state repression. The response to this initial conflict, however, was less organized and more diffuse.

Resistance appeared progressively following Conservative repression. The main (non-state) armed actors were not armies but small local groups of varied character. During the initial stages of *La Violencia*, these groups took up various roles. Liberal *chusmas* were early Liberal guerrilla groups, often with the purpose of protecting localities and advancing local party interests. Conservative *contra-chusmas* arose as a local response, in order to retaliate against the *chusmas*. For-hire mercenary groups also emerged, either working for large landowners or wealthy patrons[14] or for their own economic gain. Armed peasant groups, outside the confines of regular partisan politics, began to slowly form, eventually transforming into the two groups that would survive the end of the civil war, the FARC (Fuerzas Armadas Revolucionarias de Colombia) and ELN (Ej'ercito de Liberaci'on Nacional).[15] The continued proliferation of these groups led to the military coup of Rojas Pinilla in 1953, which represented the beginning of another bi-partisan power sharing agreement, later concretizing in the 1958 National Front.

Two major amnesty efforts occurred during these years (1954-1958) (Delgado, 2016). However, while some of the initially purely partisan bandits broke up, many groups continued operations or even formed, particularly in locations where conflict dynamics from the start of *La Violencia* had not been quelled. The groups that remained, whether still political in purpose or not, could no longer rely on formal political patronage, even at local government levels (Sánchez et al., 2001, p.21,174). Groups increasingly utilized economic extortion as the principal source of reliable funding. Hence, in 1954, contemporaries started commonly identifying the phenomenon of *bandolerismo*[16] or banditry (Betancourt, 1990; Sánchez et al., 2001).

This surge in banditry concentrated geographically in the coffee belt (the new frontier). We can track this empirically using newly digitized archival data from (Policía Nacional de Colombia, 1965), National Front government police records tracking the universe of known bandit groups identified at the time. Table A.2 presents this data, aggregated to department/frontier level. All the bandit leaders are found in coffee producing regions.[17]

[14] A well known-example is the so-called *p'ajaros* of Valle de Cauca.

[15] Both founding figures of each group, Manuel Marulanda V'elez (Tirofijo) and Fabio V'asquez Castaño, are among the bandits presented in Table A.2. V'asquez Castaño came from the second-highest coffee-producing municipality in 1944 and Tirofijo from a municipality in the top 10% of producers. Both come from the same region, the highest coffee-producing department in Colombia, Quind'io (one of the three within Viejo Caldas).

[16] During this period, the word "bandoleros" was used to describe any armed group, Conservative or Liberal; hence, I use the word bandit to describe both sides as well (Sánchez et al., 2001, p. 68).

[17] It is worth explaining the two departments in "Others"—Cauca and Huila. While I do not include them in the new frontier typology, since historiographically they are usually taken as separate cases and their coffee expansion starts around 30 years after the new frontier, by 1955, the two departments are the 6th and 8th highest coffee producing departments (United Nations. Economic Commission for Latin America, 1958). For all intents and purposes, they could be included in the new frontier classification and the analysis would not change.

The coffee-intensive new frontier holds over 70% of all bandit groups with more than 100 total, while only accounting for 37% of the country's population according to the 1964 census (Departamento Administrativo Nacional de Estadística (DANE), 1967).

Early vs. Late Period *La Violencia* Data

Conventional periodization of *La Violencia* categorizes early-period violence (1948–1953) as distinct from late period violence (1958–1964) (Chacón et al., 2011; Chacón and Sánchez, 2003). As explained previously, conflict in the first period behaved more along the lines of partisan state repression and consequent resistance, while the later period involved a greater number of local bandit groups, broadly defined. Data on the intensity of the conflict across time, however, is lacking. Detailed municipal count data for the whole Colombian territory is only available for specific years starting in 1958, the late-period. Thus, it is worthwhile to discuss how representative of the entire *La Violencia* period, each sub-period is.

I combine two sources to examine this: (i) (Gutiérrez Tovar, 1961), a digitized archival source documenting department level intentional deaths during 1946 to 1960; and (ii) excess death estimates for the whole Colombian territory produced from population census-based demographic analysis from (Romero-Prieto and Meisel-Roca, 2019)[18] for the period of 1946-1964. Figure 1.4 plots intentional deaths by frontier group from (Gutiérrez Tovar, 1961) on the left axis, ending in 1960, and the national-level death estimates from (Romero-Prieto and Meisel-Roca, 2019) on the right axis. Three principal observations are in order. First, with regard to the by-frontier numbers, once Laureano Gomez took power in 1949, the new frontier death counts outpaced the rest of the country by a significant margin, the difference peaking in the late period. Appendix Figure A.2 disaggregates these trends by department, confirming that the pattern is driven by Tolima, Caldas, Valle, and Antioquia individually rather than by any single department within the group. Second, late-period violence exhibits higher levels of intensity than all early-period years. Third, the valleys and peaks of violence across time are remarkably consistent between both sources, considering one is an archival source and the other is built ex-post from population censuses. Table A.3 in the Appendix reports international intentional death rates from (United Nations, 1961): Colombia's rate of 34 per 100,000 in 1960 was the highest in the world and falls within the bounds estimated by (Meisel-Roca and Romero-Prieto, 2017). The implied death rate using the (Gutiérrez Tovar, 1961) data for 1960 with 1964 census population is 27 per 100,000, the same order of magnitude, confirming that these independent sources bracket a consistent range of violence

[18] *La Violencia* death counts have varied greatly across time. Starting with ex-president Alfonso L'opez Pumarejo proclaiming 50,000 deaths in 1951, many figures have been circulated based on little more than speculation. (Romero-Prieto and Meisel-Roca, 2019) compile a list of 32 publications with estimates often ranging between 150,000 and 300,000. The calculations by (Oquist, 1980) (Ch. 1) have been particularly influential in shaping the notion that early-period violence was more intense than late-period. (Romero-Prieto and Meisel-Roca, 2019) use reliable census records to produce demographic estimates of the national death toll across time in order to bound the number at a national level. While still very high, their estimates are much lower than the previous estimates.

intensity. The available aggregate intentional death data across time hence supports the historical analysis offered before.

Figure 1.4. Intentional Deaths by Frontier Group, 1946–1960



Note: "New Frontier" comprises Antioquia, Caldas, Tolima, and Valle del Cauca. "Old Frontier" comprises Cundinamarca, Norte de Santander, and Santander. Vertical lines mark key political events. *Sources:* Author's calculations based on (Gutiérrez Tovar, 1961). National trend from (Romero-Prieto and Meisel-Roca, 2019).

1.3 Data & Measurement

Before moving to the empirical analysis, I provide a broad overview of main data sources and variables utilized. The principal new data in this paper are newly digitized measures of conflict intensity for the Thousand Days' War and La Violencia and the universe of coffee census records spanning from 1923 to 1960. These data are complemented with other conflict data from secondary sources, data related to other explanations for *La Violencia* in the historiography, and a host of demographic, geographic, and historical variables. For a more complete description and in-depth discussion of variable construction, please refer to Appendix A.3.1.

1.3.1 Variables

Conflict Data: Measures of conflict are drawn from newly digitized sources and secondary sources and can be separated into two broad historical periods: the nineteenth century and the mid-twentieth century.

The nineteenth century saw a wide series of civil conflicts culminating in the largest civil war at the end of the century, the Thousand Days War, the main focus of historical conflict. For this conflict, I digitize (Jaramillo Castillo, 1991), who provides a comprehensive list of

battles fought during the Thousand Days War (1899–1902), recording the municipality of each engagement. Second, broader coverage of civil conflict across the century is obtained from (Grueso, 1950), who compiles information on national-level civil wars and localized armed disputes between political factions from 1820 to 1885. Third, I incorporate data on land-related conflicts from (LeGrand, 1986), who documents legal disputes between squatters and landowners from 1827 to 1900, and 1900 to 1931.

The mid-twentieth century saw a period of intense civil conflict known as *La Violencia* (1948–1963). To capture this era at the municipal level, I rely on two main sources that together offer the most detailed account of mid-century violence available: the two classic[19] reports by (Guzmán et al., 1962, 1964) and the official police statistics published in the *Revista Criminalidad* (Ministerio de Guerra, Policía Nacional, 1959; Policía Nacional, 1962; Policía Nacional de Colombia, 1965). The Guzm'an, Fals Borda, and Umaña reports [henceforth, Guzman reports] are based on extensive fieldwork and testimonies, documenting political killings and rebel activity across rural Colombia. The *Revista Criminalidad* [henceforth, Criminality Reports], published by the National Police Department, provides systematic annual crime statistics, including homicide counts and intelligence reports on bandit groups, across municipalities. While the existing literature has relied on binary indicators of violence presence across time (Chacón, 2004; Chacón et al., 2011; Fergusson et al., 2020; Lopez-Uribe and Sanchez Torres, 2024), I use both sources to build new continuous measures of conflict intensity measured through various dimensions. The importance of measuring the intensity of conflicts has been highlighted in the Colombian context by (Deas, 2011), who criticizes the use of binary outcomes of conflict: "*...conflicts could have been large or small... Without much more about the nature of each dispute, it is unsafe and unconvincing to treat them as a uniform variable in a speculative exercise.*" (p.196). This is particularly the case for *La Violencia* given that there was extremely large geographical variation in death rates as discussed by (Romero-Prieto and Meisel-Roca, 2019). For example, according to (Gutiérrez Tovar, 1961), the four departments of the Caribbean Region exhibited 30 times fewer deaths than the rest of the country, with death rates per 100k in the 5-15 range at their peak, while the coffee regions exhibited rates in the 50-150 range at their peak.[20] Count variables allow us to try to explain this extremely large range.

From the first Guzman report, I digitize a dummy for whether a municipality is mentioned as experiencing violence in the early period (1948–1953). From the second report, I extract a count of total conflict events (clashes) by municipality for the period 1958–1963.

[19] These reports were seminal at the time they were published (Henderson, 2010, p.4-6). They can be seen as the first instance in a long history of "truth commissions" in Colombia. Today, the principal organization carrying out this endeavor is the *Centro Nacional de Memoria Hist'orica* (CNMH), the institute in charge of collecting data on the Colombian conflict.

[20] For reference, in today's terms, this variation of death rate is equivalent to going from the intentional death rate per 100k of the United States to twice the death rate of the most violent country in the world (https://data.unodc.org/datareport/hom-victim).

From the *Revista Criminalidad*, I digitize: (i) a dummy for early violence (from the 1959 volume); (ii) the total number of homicides from *La Violencia* in 1962 (from the 1962 volume); and (iii) the number of bandit leaders in 1962 (from the 1965 volume).

The skeptical reader might be worried about the implicit biases surrounding the compilation of violence data during a period of civil conflict. As argued in the literature (Chacón et al., 2011), the Guzman Commission adopted a particularly third-party approach to the analysis of violence. These reports, as well as the National Police reports, come from a specific moment in Colombian political history when both historic parties—Liberals and Conservatives—agreed to a 15-year power-sharing agreement to temper the endemic conflict seen during the civil war. Both of these reasons assuage concerns about systematic coding along party lines.

However, beyond partisan bias, violence statistics might simply suffer from measurement error geographically, that is, different regions having more or less of under-reporting. While this will always be a concern with this sort of data, I try to examine the possible extent of measurement error by leveraging the fact I have digitized two different, independently compiled municipal level archival sources on conflict, the violent deaths in 1962 (Policía Nacional de Colombia, 1965) and the clashes for 1958-1963 (Guzmán et al., 1964) as well as a department level panel for the 1946-1960 period (Gutiérrez Tovar, 1961). First, I verify that the time trends at the department level are similar across the clashes data and the department level panel. They track remarkably well. Second, I test whether the relative rank of intensity among municipal observations tracks among both municipal sources. The Spearman rank correlation between the two sources is $\rho = 0.639$ ($p < 0.001$), and over 60% of municipalities in the top decile of one source also appear in the top decile of the other. The high degree of similarity between two completely different sources is reassuring, even if eliminating all concern of measurement error is impossible with historical data. Appendix A.3.2 presents the results for both of these analyses.

Coffee Data: Coffee variables are obtained from a combination of agricultural reports, national coffee censuses, and bulletins from the *Federación Nacional de Cafeteros* (Departamento Administrativo Nacional de Estadística and Charry Lara, 1962; Federación Nacional de Cafeteros de Colombia, 1933, 1943, 1945; Monsalve, 1927; Sociedad de Productores de Café and Sociedad de Agricultores de Colombia, 1923). The main variable is municipal coffee production measured in 62.5 kg sacks, with the 1944 census serving as the baseline measure. I also use the coffee area in 1960 and the number of coffee trees in 1932 as alternative measures.

Competing Theories: The Colombian historiography has proposed several explanations for *La Violencia*. I collect data on the leading alternatives to test them against coffee production.

State capacity. (Oquist, 1980) argues that violence arose from a breakdown of state authority. I proxy pre-conflict state capacity with two measures: municipal revenue in 1923 and the number of state employees in 1918, from the 1918 population census.

Political competition. (Chacón et al., 2011) argue that politically competitive municipalities were more likely to experience violence during the early years of *La Violencia*. I measure political competition as 1 − |Conservative share − Liberal share| from the 1946 presidential election.

Partisan identity. The Colombian historiography documents that state-sponsored partisan repression targeted Liberal municipalities (Chacón and Sánchez, 2003; Roldán, 2002). I measure partisan identity through the Liberal vote share in the 1946 presidential election. For a deeper historical measure, I also use Liberal vote share in the 1856 presidential election.

Agrarian tensions. (LeGrand, 1986) and (Lopez-Uribe and Sanchez Torres, 2024) emphasize the role of land disputes in fueling violence. I include land conflicts from 1900 to 1931 from (LeGrand, 1986) and a measure of land dispossession at the municipal level for 1914–1946 from (Lopez-Uribe and Sanchez Torres, 2024).

Historic, Demographic, and Geographic Controls: I complement the conflict, coffee, and competing theory variables with three sets of controls that capture deep historical state capacity, demographic structure, and geographic characteristics.

Historic controls. Encomienda assignment is a dummy for whether a municipality had a colonial *encomienda*, a forced-labor institution in which the Spanish Crown assigned indigenous tributary populations to Spanish settlers. These data are from (Faguet et al., 2024), who digitized a 1560 colonial survey. Colonial state presence in 1794 is a dummy for the existence of Crown employees and fiscal institutions at the municipal level, from (García Jimeno, 2005), based on the original colonial source of (Durán y Díaz, 2012).

Demographic controls. Log municipal population from national population censuses (1851, 1905, or 1938, depending on the analysis) and the share of agricultural employment in 1938, from the 1938 census.

Geographic controls. Altitude, terrain ruggedness (Nunn and Puga, 2012), distance to the departmental capital, and latitude and longitude. These are derived from the Universidad de Los Andes CEDE dataset.

1.3.2 Construction of Historical Geographic Units (HGU)

Colombian municipalities have undergone frequent administrative reorganization since the late nineteenth century, primarily through subdivision, but also through name changes and mergers. This creates a fundamental problem for longitudinal analysis: if a municipality splits between two data periods, matching observations by name across periods compares non-equivalent geographic areas, biasing any estimated relationship. To address this, I construct Historical Geographic Units (HGUs) following the aggregation procedure of (Horan and Hargis, 1995). A crosswalk links each pair of adjacent data years, recording every name change, split, and merger. When a municipality in a later period is identified as having split from a parent

municipality, it is aggregated back into the parent, so that the same geographic area is observed in both periods. Figure A.14 in the appendix illustrates this procedure with a concrete example from the data. In each empirical analysis, the HGU is constructed at the level of the oldest variable used, since aggregating to the earliest period maximizes the number of consistent units available and preserves cross-sectional variation. The crosswalk draws on the 1918 population census, the 1924 coffee census (Monsalve, 1927), the IGAC *Diccionario Geográfico de Colombia*, and from 1962 onward the official DANE registry of municipal changes. A fuller discussion of the HGU construction procedure, together with the complete list of documented name changes, splits, and mergers, is provided in Appendix A.3.3.

1.4 Does Conflict Persist Across Centuries?

A prominent claim in the Colombian historiography is that mid-twentieth-century violence was rooted in the partisan conflicts of the nineteenth century. This section tests the persistence claim at the municipal level, distinguishing between two margins: the *incidence* of conflict (whether violence occurs at all) and its *intensity* (how many die). I find that incidence persists across centuries but intensity does not, and that partisan identity is the channel through which incidence persistence operates.

1.4.1 Testing Persistence at the Municipality Level

I test whether nineteenth-century conflict predicts twentieth-century *La Violencia* using Harmonized Geographic Units constructed at the 1892 level. The persistence claim has two testable implications: that historically violent municipalities are more likely to experience *La Violencia* at all (the *extensive margin*), and that they experience more intense violence when it occurs (the *intensive margin*). Testing both matters because a conflict that recurs in the same places but for different reasons would show extensive-margin persistence without intensive-margin persistence. The analysis proceeds in two steps. First, I estimate persistence on the extensive margin using a linear probability model:

$$\text{Conflict}_i^{20\text{th}} = \beta\, \text{Conflict}_i^{19\text{th}} + X_i'\gamma + \delta_d + \varepsilon_i, \tag{1.1}$$

where $\text{Conflict}_i^{19\text{th}}$ is a dummy for the relevant nineteenth-century conflict measure, δ_d are department fixed effects, and X_i includes colonial state presence (1794), encomienda assignment, altitude, ruggedness, distance to the department capital, and log population. Second, I estimate a count model for the number of violent deaths or clashes using Pseudo-Poisson Maximum Likelihood (PPML) with the same controls. If the persistence claim holds, both the extensive and intensive margins should show positive, significant coefficients: municipalities with a history of conflict should be both more likely to experience *La Violencia* and more severely affected by it.

1.4.2 Where Violence Erupts: Incidence Persists Across Centuries

Does nineteenth-century conflict predict where *La Violencia* breaks out? I estimate this as a linear probability model in which both sides of the regression are binary: the dependent variable equals one if a municipality experienced any early violence (1949–1953), and the independent variable is a dummy for the relevant nineteenth-century conflict measure. Table 1.1 reports the results.

Panel A tests whether nineteenth-century conflict predicts early *La Violencia*. The central result is that the Thousand Days War, the largest and most recent nineteenth-century conflict, significantly predicts twentieth-century violence: municipalities that experienced Thousand Days battles are 8.8 percentage points more likely to see early *La Violencia* (column 2). Nineteenth-century land conflicts show an even stronger association at 20.5 percentage points (column 3), and a composite indicator for any nineteenth-century conflict yields 11.0 percentage points (column 4). Earlier civil war battles on their own are not significant (column 1), suggesting that persistence is strongest for the most proximate and politically salient conflicts.

Panel B extends the question further back within the nineteenth century. Earlier civil war battles predict participation in the Thousand Days War itself by 11.6 percentage points (column 5). The locations where partisan warfare occurred early in the century are the same locations that fought in its largest war, and, a half-century later, where *La Violencia* would later commence.

Table 1.1. Persistence of Conflict Incidence

	Panel A: Early Violencia (Dummy)				**Panel B: 1000 Days War (Dummy)**	
	(1)	(2)	(3)	(4)	(5)	(6)
Civil wars battles (dummy)	0.029				0.116***	
	(0.035)				(0.041)	
1000 Days battles (dummy)		0.088**				
		(0.040)				
Land conflicts 19th c. (dummy)			0.205***			0.076
			(0.076)			(0.071)
Any 19th c. conflict				0.110***		
				(0.034)		
Dept. FE	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y
Pop. 1905	Y	Y	Y	Y		
Pop. 1870					Y	Y
HGU level	1892	1892	1892	1892	1892	1892
N	651	651	651	651	651	651

Notes: OLS (LPM). All independent variables enter as dummies. Robust standard errors (HC1) in parentheses. State controls: colonial state presence dummy (1794) and encomienda dummy. Geographic controls: altitude, ruggedness, distance to dept. capital. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Decomposing the composite early violence indicator into its two components (Criminality Reports and Guzman Reports) confirms the pattern. The Thousand Days battles significantly

predict Criminality Reports incidence (7.4 percentage points), and land conflicts predict both components. These results are in Appendix Table A.8.

1.4.3 How Intense Violence Becomes: Intensity Does Not Persist

The previous result established that past conflict predicts *where* violence breaks out. I now ask a different question: Does past conflict predict *how intense* violence becomes? As explained before, variation in intensity during *La Violencia* was large, so understanding what is driving high violence locations is essential to understanding the conflict.

I focus on the late period of *La Violencia* for a specific reason. As the aggregate data in Figure 1.4 showed, this later period was in fact more intense than the early one, and the nature of the conflict had transformed: violence was increasingly driven by small armed groups operating under the bipartisan National Front government rather than by the original partisan repression. If the conflict's character has shifted from partisan warfare to something else, we should expect the intensity of late-period violence to be disconnected from nineteenth-century partisan conflict. The dependent variable is violent deaths in 1962, estimated via Pseudo-Poisson Maximum Likelihood (PPML) to account for the skewed nature of count data.

Table 1.2 reports the results. None of the three nineteenth-century conflict measures (civil war battles, Thousand Days battles, or land conflicts) predicts the intensity of late-period violence. All point estimates are small and statistically insignificant. The two results together (persistent incidence but null intensity) tell a coherent story. The outbreak of *La Violencia* follows the geography of past partisan warfare, but the scale of the later conflict cannot be explained by historical patterns. Something else determined why certain municipalities experienced far more deaths than others. This is the variation that Section 1.5 will explain through coffee production.

A null result requires extra scrutiny: it could reflect insufficient power, measurement error, or a particular specification choice rather than a genuine absence of persistence. I subject the finding to multiple robustness checks, each addressing a distinct concern. The first replaces the outcome measure to rule out a measurement artifact. The next two decompose the count model to test whether persistence hides in particular margins or quantiles of the violence distribution. The last splits the sample by region to check for geographic heterogeneity. Full results are in Appendix A.2.

The null might be specific to the death-count measure from the police statistics. Replacing violent deaths with clashes from Guzman Reports (1958–1963), an independently compiled source covering armed confrontations rather than deaths, yields the same null result (Appendix Table A.9).

The PPML specification pools all municipalities, combining the question of *whether* violence occurs with *how much*. Decomposing the count model into its extensive and intensive

Table 1.2. Persistence of Conflict Intensity

	Deaths 1962 (PPML)		
	(1)	(2)	(3)
Civil wars battles (count)	0.056		
	(0.046)		
1000 Days battles (count)		-0.024	
		(0.060)	
Land conflicts 19th c. (dummy)			0.115
			(0.390)
Dept. FE	Y	Y	Y
State controls	Y	Y	Y
Geographic controls	Y	Y	Y
Pop. controls	Y	Y	Y
HGU level	1892	1892	1892
N	651	651	651

Notes: Pseudo-Poisson Maximum Likelihood (PPML). Continuous battle counts as independent variables; land conflicts 19th c. always enters as a dummy. Robust standard errors (HC1) in parentheses. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. Land conflicts 19th c. always enters as a dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

components reveals that Thousand Days battles do predict the probability of *any* late-period deaths (13.2 percentage points, significant; Appendix Table A.10). However, conditional on experiencing any violence, past conflict has no predictive power for the number of deaths. How to explain this disconnect? If past conflict predicts only very few deaths then it would be evidence of persistence into low-violence municipalities. In this case, the extensive margin would appear persistent without explaining meaningful variation in intensity. A threshold-varying exercise confirms this interpretation. I estimate a series of regressions in which the dependent variable is a binary indicator for deaths exceeding threshold k, varying k from low to high values. The coefficient shrinks monotonically toward zero as k rises, becoming insignificant well before reaching the upper tail (Appendix Figures A.3 and A.4). Hence, if the goal is understanding why some regions of Colombia had such extremely high levels of conflict, past conflict does not provide explanatory power.

Persistence might be concentrated at particular quantiles of the conditional death distribution. Quantile regressions from the 10th to the 90th percentile show the coefficient on past conflict is near zero and insignificant at every quantile (Appendix Figures A.5 and A.6).

The null might reflect regional heterogeneity. Restricting the sample to the new frontier (Antioquia, Caldas, Tolima, and Valle, the departments where violence was most concentrated) yields uniformly null estimates (Appendix Table A.12). The rest of Colombia shows a weak positive association only for land conflicts (Appendix Table A.13). The coffee frontier is entirely disconnected from nineteenth-century conflict patterns in terms of intensity.

1.4.4 How Partisan Identity Transmits Conflict Across Centuries

Liberal Voting as the Channel of Incidence Persistence

The previous subsection established that the *locations* of conflict persist across centuries but that the *scale* of violence does not. This raises a question: what mechanism carries conflict propensity from one century to the next? A natural candidate is partisan identity. Colombia's Liberal and Conservative parties were the organizing axis of both nineteenth-century civil wars and mid-twentieth-century violence. If municipalities that were Liberal in the 1850s remained Liberal in the 1940s, and if the outbreak of *La Violencia* was driven by partisan persecution, then partisan affiliation is the channel through which conflict locations persist.

I test this hypothesis in two steps, reported in Table 1.3. The first step (Panel A) asks whether Liberal identity in the nineteenth century independently predicts the extensive and intensive margins of the Thousand Days War (1899–1902). Liberal vote share in 1856[21] emerges as a strong predictor: a municipality that voted entirely Liberal is 14.4 percentage points more likely to experience a Thousand Days battle (column 1) and sees 0.333 more battles (column 2).

The second step (Panel B) shifts the question forward by a century and changes the empirical design in an important way. The goal is no longer to estimate the effect of past conflict on later conflict, which was established in Section 1.4. Instead, the goal is to test whether partisan identity itself predicts the outbreak of *La Violencia*. This shift in focus changes the role of each variable. Liberal vote share in 1946 becomes the variable of interest, and past conflict measures (Thousand Days battles, land conflicts) enter as controls. Because partisan identity is now the treatment rather than a potential confounder, I can additionally include the competing theories of *La Violencia* as pre-treatment controls: state capacity (municipal revenue 1923, state employees 1918), land tenure (land conflict 1900-1931 and land dispossession 1914–1946), and political competition (1946). The question this design answers is whether partisan identity retains predictive power for the outbreak of violence even after conditioning on all other proposed causes. It does: the coefficient on Liberal vote share in 1946 is 36.6 percentage points for any early violence (column 3), 32.1 percentage points for Guzman Reports violence (column 4), and 28.9 percentage points for Criminality Reports (column 5), all significant at the 1% level.

A concern with the Panel B results is that Liberal vote share in 1946 may itself reflect pre-violence conditions. If municipalities already experiencing tensions voted more Liberal as a result, the direction of causality would be reversed: violence would cause Liberal voting, not the other way around. To address this concern, I replace the 1946 Liberal vote share with Liberal vote share in 1856, a measure of partisan identity recorded a full century before *La*

[21] This year is used as it is the first direct election in the country's history, and later elections in the century are no longer representative at the municipal level since Colombia goes through a series of constitutional reforms, which do not hold direct national level elections until the twentieth century.

Table 1.3. Liberal Voting and Conflict Persistence

	Panel A: 19th Century		**Panel B: Early Violence (1946–53)**		
	(1) 1000 Days (Dummy)	(2) 1000 Days (Count)	(3) Any Early Violence (Dummy)	(4) Early Violence Guzman Rep. (Dummy)	(5) Early Criminality (Dummy)
Liberal share 1856	0.144***	0.333**			
	(0.053)	(0.166)			
Liberal share 1946			0.366***	0.321***	0.289***
			(0.055)	(0.055)	(0.044)
Dept. FE	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y
Geo. controls	Y	Y	Y	Y	Y
Pop. 1851	Y	Y			
Pop. 1938			Y	Y	Y
Past conflict	Y	Y	Y	Y	Y
Competing theories	N	N	Y	Y	Y
HGU level	1892	1892	1892	1892	1892
N	651	651	638	638	638

Notes: HC1 robust SEs in parentheses. State controls: state presence 1794 dummy, encomienda dummy. Geo. controls: altitude, ruggedness, distance to dept. capital. Past conflict: civil wars battles (count) in Panel A; 1000 Days battles (dummy), land conflicts 1827–1900 (dummy) in Panel B. Competing theories: log revenues 1923, state employees 1918, political competition, land conflicts 20th c., land dispossession, share of agricultural employment 1938. Liberal share 1856 = 1− share Ospina. Liberal share 1946 = (votes Turbay + votes Gaitán) / total votes 1946. The decrease in observations from 651 to 638 is due to 13 observations recording 0 votes in the 1946 elections. Coding these municipalities as Liberal or Conservative would be arbitrary, so they are dropped. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Violencia, and regress the early violence indicators directly on this deeper measure. The logic is straightforward: if a municipality's partisan alignment in the 1850s predicts its experience of violence in the 1940s and 1950s, this relationship cannot be driven by endogeneity from mid-twentieth-century conditions, because those conditions did not yet exist when the 1856 vote was cast. If, on the other hand, the 1856 measure has no predictive power, the Panel B results may reflect contemporaneous confounding rather than persistent partisan identity. Liberal vote share in 1856 does predict early violence: 14.0 percentage points under baseline controls, attenuating to 10.9 percentage points when the full set of controls is added. Both estimates are statistically significant (Appendix Table A.14). Partisan affiliation persists across generations and transmits conflict propensity from the nineteenth to the twentieth century.

Event Study: Did Gómez's Presidency Trigger Partisan Violence?

The cross-sectional evidence for liberal voting may reflect omitted municipal characteristics that correlate with both partisan identity and conflict propensity. To address this concern, I exploit fine-grained data from (Roldán, 2002) covering 99 Antioquia municipalities with annual violence counts for the period surrounding *La Violencia*. I estimate an event study that compares violence in Liberal-majority and non-Liberal-majority municipalities before and after Laureano Gómez came to power in 1950, an episode that triggered partisan persecution disproportionately directed at Liberal municipalities. I estimate the following specification:

$$y_{it} = \alpha_i + \gamma_t + \sum_{t \neq t_0} \delta_t \, Z_{i,t} + \varepsilon_{it}, \tag{1.2}$$

where y_{it} denotes violence in municipality i in year t, α_i are municipality fixed effects, and γ_t are year fixed effects. The coefficients δ_t trace out the difference in violence between Liberal-majority and non-Liberal-majority municipalities in each year relative to the omitted baseline year, 1950. The identifying assumption is that, absent Gómez's accession and the ensuing partisan repression, violence in Liberal and non-Liberal municipalities would have evolved in parallel over time. Under this assumption, the pre-treatment coefficients provide a test of differential trends prior to treatment, while the post-1950 coefficients capture the dynamic effect of the political shock on violence in Liberal municipalities.

Figure 1.5 presents the estimates. The pre-treatment coefficients are small and statistically indistinguishable from zero: before 1950, Liberal and non-Liberal municipalities exhibited similar violence trajectories. After 1950, violence rises sharply and persistently in Liberal municipalities relative to non-Liberal municipalities. Complementing this event-study evidence, a PPML regression of total deaths over 1949–1953 on Liberal vote share in 1946 yields a statistically significant coefficient of 3.655, confirming the cross-sectional relationship within Antioquia (Appendix Table A.15). A balance test further shows that Liberal and non-Liberal municipalities are comparable on observable pre-conflict characteristics (Appendix Figure A.7).

While these results are only for one department, and consequently, we cannot extrapolate causality to other regions, they provide strength to the cross-section regressions linking partisan identity and the onset of violence during *La Violencia*.

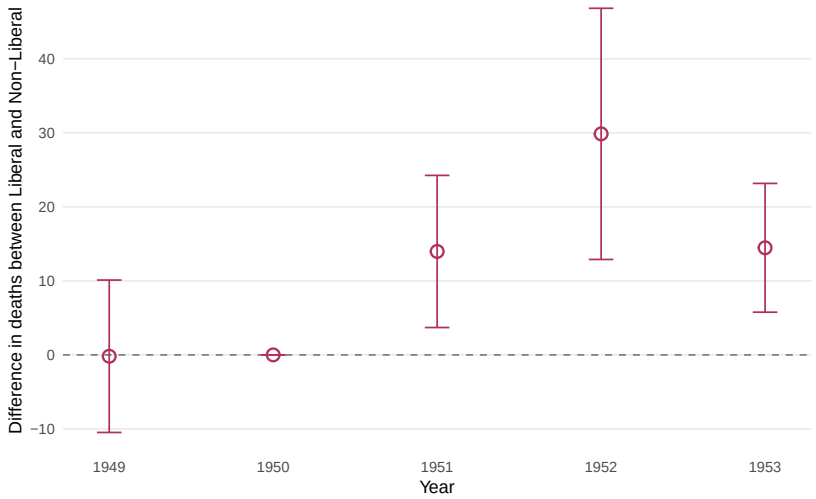


Figure 1.5. Event Study: Liberal vs. Conservative Municipalities (Antioquia)

Notes: Each point plots the estimated δ_t coefficient from Equation 1.2. The dependent variable is annual violence counts from (Roldán, 2002) for 99 Antioquia municipalities. The treatment variable is a liberal-majority indicator interacted with period dummies. The omitted period is 1950 (Laureano Gómez coming to power). Vertical bars show 95% confidence intervals.

1.5 What Explains the Intensity of *La Violencia*?

The previous section established that the *locations* of conflict persist across centuries, but the *scale* of violence does not. Something other than historical conflict patterns must explain why certain municipalities experienced far more intense violence during *La Violencia* than others. The historical analysis pointed to coffee cultivation as a candidate: municipalities with larger coffee sectors offered higher appropriable rents, and the coffee frontier is where violence was most concentrated. This section tests whether coffee production is the primary correlate of variation in conflict intensity, relative to the leading alternative explanations from the historiography.

1.5.1 Coffee Production and Violence Intensity

Rather than testing each theory in isolation, I include coffee production alongside all competing explanations simultaneously. This horse-race design asks: which factor is the strongest predictor of violence intensity, conditional on the others?

The specification includes coffee production (sacks in 1944) alongside the leading alternative explanations from the historiography. Partisan identity enters through Liberal vote share in 1946, capturing the role of partisan repression documented in the previous section. State capacity is measured through municipal revenue in 1923 and state employees in 1918, testing whether violence concentrated in areas where the state was weakest. Political competition in the 1946 elections tests whether closely contested municipalities experienced more violence. Agrarian tensions are captured through two variables: land dispossesion (1914–1946) and land conflicts (1900–1931), reflecting the argument that violence was rooted in disputes over land access. In total, six competing variables enter alongside coffee.

All continuous regressors are standardized by dividing by two standard deviations and the outcome is divided by one standard deviation, following (Gelman, 2008), so that coefficients of binary and continuous regressors are comparable. The baseline specification includes department fixed effects, colonial state controls (state presence 1794, encomienda), geographic controls (altitude, ruggedness, distance to departmental capital), and population controls (log population 1938, agricultural employment share).[22] The specification is:

$$\text{Violence}_i = \beta\,\text{Coffee}_i + T_i'\gamma + G_i'\theta + \delta_d + \varepsilon_i, \tag{1.3}$$

where T_i collects the competing theories, and G_i includes geographic, population, and colonial state controls. I estimate both OLS on the standardized outcome and PPML on raw death counts to account for the count nature of the dependent variable.

[22] This section uses Harmonized Geographic Units constructed at the 1918 level rather than the 1892 level used in Section 1.4, because the competing theories require variables (such as state employees) that are only available for municipalities as they existed in 1918.

Table 1.4 reports the results. In the OLS specification on the full sample (column 1), the standardized coffee coefficient is 0.783, statistically significant at the 1% level: a two-standard-deviation increase in coffee production is associated with a 0.783 standard deviation increase in violent deaths, conditional on all controls and department fixed effects.

Among the competing theories, only partisan identity has significant predictive power. Liberal vote share in 1946 enters with a coefficient of 0.322, also significant at the 1% level. This is expected: the early phase of *La Violencia* was fundamentally partisan, and it would be surprising if partisan identity did not matter for late-period violence as well. What is notable is that coffee production, absent from the incidence regressions in Section 1.4, emerges as the dominant predictor of violence *intensity*, with a coefficient more than twice the magnitude of the Liberal vote share.[23] Municipal revenue, state employment, political competition, land dispossession, and land conflicts are all statistically insignificant.

The result is not driven by the contrast between municipalities that produce coffee and those that do not. Restricting the sample to coffee-producing municipalities (column 2) yields a coefficient of 0.711, still significant at the 1% level. The PPML estimates on raw death counts confirm the pattern in both the full sample (column 3: 0.381) and among producers (column 4: 0.480), both significant at the 1% level.

Robustness. The coffee result is subjected to four sets of robustness checks, each detailed in Appendix A.2.2.

The result holds across alternative outcome measures and alternative coffee measures. Replacing violent deaths with clashes from Guzman Reports (1958–1963), deaths per 100,000, or a binary any-deaths indicator yields significant coffee coefficients in all cases (0.766, 0.310, and 0.130, respectively). Using coffee area in 1960 (0.936) or coffee trees in 1932 (0.485), both statistically significant, in place of 1944 production confirms the finding is not an artifact of the specific coffee measure (Appendix A.2.2).

A concern with geographic data is that one region may drive the entire result. A leave-one-department-out analysis shows the coffee coefficient remains stable regardless of which department is dropped. Splitting the sample into the new frontier (Antioquia, Caldas, Tolima, Valle) and the rest of Colombia yields significant coefficients in both subsamples (0.740 and 0.325 respectively), though the effect is larger in the coffee heartland (Appendix A.2.2).

Spatial dependence has been identified as potential driver of spurious regressions (Conley and Kelly, 2025). To address this, I re-estimate the main benchmark estimation with Conley standard errors (Conley, 1999) at bandwidths from 50 to 500 km. All estimations preserve significance at the 1% level at all cutoffs. A spatial unit root test (Becker et al., 2025) confirms

[23] Adding coffee production to the incidence regressions from Section 1.4 does not change those results: Liberal vote share remains the dominant predictor of conflict *incidence*, and coffee is insignificant. Coffee matters for intensity specifically in the later period.

Table 1.4. Short-Run Determinants: Violent Deaths

	OLS (Standardized)		**PPML (Counts)**†	
	(1) Full Sample	(2) Coffee > 0	(3) Full Sample	(4) Coffee > 0
Coffee 1944	0.783***	0.711***	0.381***	0.480***
	(0.238)	(0.245)	(0.116)	(0.108)
Liberal 1946	0.322***	0.421***	1.703***	1.713***
	(0.092)	(0.136)	(0.422)	(0.472)
Revenue 1923	0.092	0.091	0.023	0.021
	(0.263)	(0.296)	(0.029)	(0.029)
State empl. 1918	-0.015	0.092	-0.002	-0.001
	(0.041)	(0.106)	(0.001)	(0.001)
Pol. comp. 1946	-0.007	-0.156	-0.047	-0.086
	(0.090)	(0.146)	(0.345)	(0.384)
Land dispos. 1914–46	0.074	0.059	0.036	0.029
	(0.069)	(0.109)	(0.026)	(0.029)
Land conflict 1900–31	0.020	0.040	0.033	0.073
	(0.119)	(0.160)	(0.102)	(0.107)
Dept. FE	Y	Y	Y	Y
State controls	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y
N	741	382	590	345
R^2 / Pseudo R^2	0.402	0.388	0.718	0.702
Within R^2 / Pseudo R^2	0.169	0.180	0.486	0.519
HGU level	1918	1918	1918	1918

Notes: DV: Violent deaths 1962. Cols. (1)–(2): OLS on standardized DV; continuous IVs divided by 2 SD for comparability. Cols. (3)–(4): PPML on raw counts. HC1 robust SEs. Liberal, pol. comp. enter raw. State controls: state presence (1794), encomienda. Geo.: altitude, ruggedness, dist. capital. Pop.: log pop. 1938, share agricultural 1938. † PPML coffee coefficient $\times 10^{-5}$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

that the OLS residuals are stationary, ruling out a spurious spatial regression (Appendix A.2.2).

Finally, the relationship between coffee and violence is not driven solely by a few extreme observations. A binscatter of violent deaths on log coffee shows that deaths roughly double from the lowest to the penultimate coffee bin, and the top bin (which accounts for roughly half of all coffee production) concentrates the highest death counts. Winsorizing the outcome at the 99th, 95th, 90th, and 80th percentiles reduces the coefficient to 0.593 and then to 0.407, but it remains significant at every threshold. Rents appear to bite hardest at the top of the coffee distribution, but the relationship extends well below it (Appendix A.2.2).

1.5.2 How Coffee Translates into Violence: Rents and Fighting Feasibility

A Framework

Coffee production is associated with violence intensity, but how does coffee translate into conflict? A natural framework combines two ingredients: economic incentives and fighting feasibility. Coffee rents provide the resources that make armed activity profitable, while

proximity to existing conflict hubs provides the organizational capacity and manpower to exploit those rents. Where both conditions are met (high coffee and short distance to an early *La Violencia* hub), violence should be most intense. Where either is absent, violence should be lower. Table 1.5 summarizes this logic.

Table 1.5. Conceptual Framework: Economic Incentives and Fighting Feasibility

	Economic Incentives (Coffee Rents)	
	Low / None	High
Fighting Feasibility		
High (Near conflict hub)	Medium	**High conflict**
Low (Far from conflict hub)	Low conflict	Medium

Notes: Economic incentives proxied by coffee production (sacks, 1944). Fighting feasibility proxied by distance to municipalities with early La Violencia (1949–53).

Descriptive Evidence

Table 1.6 provides descriptive evidence for this interaction. I classify municipalities into three proximity groups based on their distance to the nearest early violence hub (1949–1953): at the hub itself (distance = 0), near (0–30 km), and far (> 30 km). The near/far threshold of 30 km corresponds to the median distance among non-hub municipalities (29.6 km). Within each group, I compare coffee-producing and non-producing municipalities. Among hub municipalities, those with coffee average 8.99 violent deaths (80% experiencing any violence) compared to 1.73 deaths (39%) for those without coffee. Among nearby municipalities, the gap narrows but persists: 2.44 deaths versus 0.25. Among far municipalities, the gap is smaller still: 0.76 versus 0.03. Coffee amplifies violence at every distance, but the amplification is strongest where fighting is already feasible.

Table 1.6. Mean Violence: Any Coffee vs None × Three Proximity Groups

	Coffee (1944)	
	No coffee	Any coffee
Proximity to hub		
At hub (dist = 0)	1.73 39% w/ viol., N = 56	8.99 80% w/ viol., N = 142
Near (0 < dist ≤ 30 km)	0.25 11% w/ viol., N = 136	2.44 57% w/ viol., N = 136
Far (> 30 km)	0.03 2% w/ viol., N = 167	0.76 23% w/ viol., N = 104

Notes: Near/Far split at median distance among non-hub municipalities (29.6 km). N = 741..

The Coffee Effect Decays with Distance from Conflict Hubs

To test this interaction formally, I estimate an OLS specification that includes a coffee dummy, distance to the nearest early violence hub (in km), and their interaction, alongside the full set of controls and department fixed effects:

$$Y_i = \beta_1 \,\text{Coffee}_i + \beta_2 \,\text{Dist}_i + \beta_3 \,(\text{Coffee}_i \times \text{Dist}_i) + X_i'\gamma + \delta_d + \varepsilon_i \tag{1.4}$$

The interaction term β_3 captures how the coffee effect changes with distance. At distance zero, coffee-producing municipalities have 2.197 more deaths than non-producers (statistically significant). Each additional kilometer reduces the coffee effect by 0.022 deaths (also significant). Figure 1.6 plots the predicted deaths by coffee status across the 0–100 km range: the gap between coffee and non-coffee municipalities is wide near conflict hubs and closes gradually with distance. Table A.4 in the appendix reports the full regression, including a PPML specification; Figure A.10 in Appendix A.2.3 shows the PPML margins.

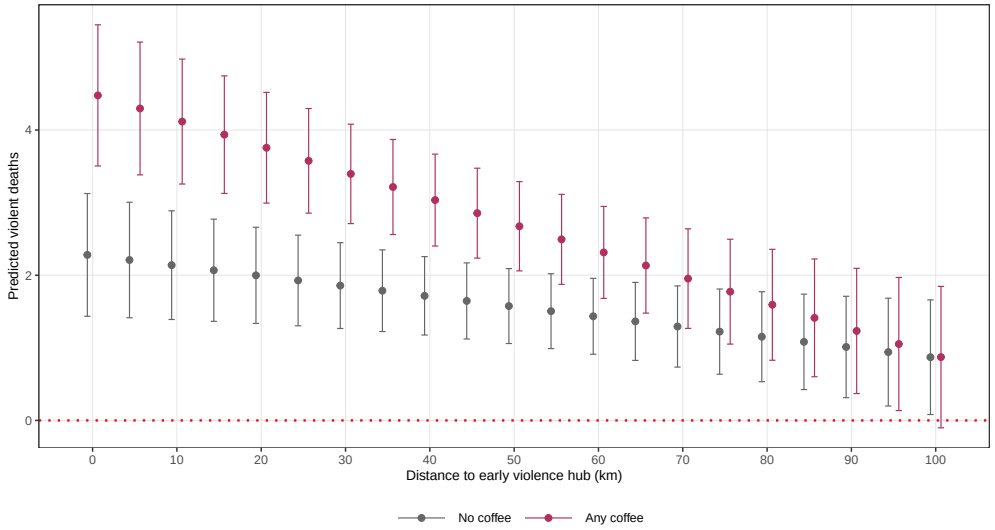


Figure 1.6. OLS Predicted Violent Deaths by Coffee Status and Distance to Hub

Notes: Predicted violent deaths (1962) from OLS estimation of Equation 1.4. Margins evaluated at 0–100 km in 5 km steps. Whiskers show 95% confidence intervals. Controls: competing theories, state controls, geographic controls, population controls, and department FE. $N = 741$.

1.5.3 Armed Groups as the Channel

Coffee Municipalities Attract Armed Groups

If coffee rents finance armed groups, coffee-producing municipalities should attract more bandit groups described in Section 1.2.2. I test this by replacing the dependent variable in the main specification (Table 1.4) with measures of armed group presence from Policía Nacional de Colombia (1965). Table 1.7 reports three specifications. A two-standard-deviation increase in coffee production raises the probability of hosting a bandit leader by 18.0 percentage points

(column 1), increases the leader count by 0.614 standard deviations (column 2), and yields a PPML coefficient of 0.713 (column 3). All three estimates are statistically significant at the 1% level. Coffee municipalities are not only more violent, but they are also where armed groups locate.

Table 1.7. Bandit Location: Coffee and Armed Group Presence

DV:	(1) LPM (Std) Has leader	(2) OLS (Std) Leader count	(3)† PPML Leader count
Coffee 1944	0.180***	0.614***	0.713***
	(0.051)	(0.174)	(0.177)
Dept. FE	Y	Y	Y
Competing theories	Y	Y	Y
State controls	Y	Y	Y
Geo. controls	Y	Y	Y
Pop. controls	Y	Y	Y
N	741	741	426
R^2 / Pseudo R^2	0.234	0.274	0.409
Within R^2 / Pseudo R^2	0.097	0.175	0.318
HGU level	1918	1918	1918

Notes: Cols. (1)–(2): OLS/LPM, coffee divided by 2 SD. Col. (3): PPML. † PPML coffee coefficient $\times 10^{-5}$. HC1 robust SEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Does Coffee Operate Through Armed Groups?

The previous results show that coffee is associated with both violence and bandit presence. A stronger test asks whether coffee's effect on violence operates *through* armed groups. I interact coffee production with a three-category measure of bandit presence. The reference group consists of municipalities with no leader and no neighboring municipality with a leader. The second group includes municipalities whose neighbor hosts a leader[24]. The third group consists of municipalities that host their own bandit leader. The specification is:

$$\begin{aligned} Y_i = {} & \beta_1 \,\text{Coffee}_i + \beta_2 \,\text{Neighbor}_i + \beta_3 \,\text{Leader}_i \\ & + \beta_4 \,(\text{Coffee}_i \times \text{Neighbor}_i) + \beta_5 \,(\text{Coffee}_i \times \text{Leader}_i) \\ & + X_i'\gamma + \delta_d + \varepsilon_i \end{aligned} \tag{1.5}$$

where Coffee is measured in 10,000 sacks (mean among producers $\approx$ 13,000). The coefficients β_4 and β_5 capture how the marginal effect of coffee on violence changes with bandit proximity. If coffee generates violence *through* armed groups, the marginal effect of coffee should be near zero where bandits are absent and large where they are present. I estimate this in both OLS and PPML.

[24] (Chacón and Sánchez, 2003) documented that neighbor bandit groups had an effect on the extensive margin of late-violence.

Figure 1.7 plots the average marginal effect of coffee on violent deaths for each bandit group. In the OLS, the marginal effect of coffee is near zero in municipalities without any bandit presence, rises modestly where a neighbor has a leader, and is large and statistically significant where the municipality itself hosts a leader (2.420). The PPML estimates follow the same qualitative pattern (near zero without bandits, larger and significant with own bandit leader), but the marginal effects are roughly an order of magnitude smaller. This decrease in magnitude decreases confidence in the exact level of magnitude of the OLS estimates, but does provide suggestive evidence for the mechanism.

What is more robust across both estimators is the direct role of armed groups. In the full regression (Appendix Table A.5), the direct coffee coefficient drops to near zero and is no longer significant for the first time, once bandit presence is included, while having a bandit leader remains large and significant in both specifications (3.398 in OLS, 2.118 in PPML). Armed group presence absorbs much of coffee's explanatory power for violence, consistent with bandits serving as a mediating channel through which coffee rents translate into conflict.

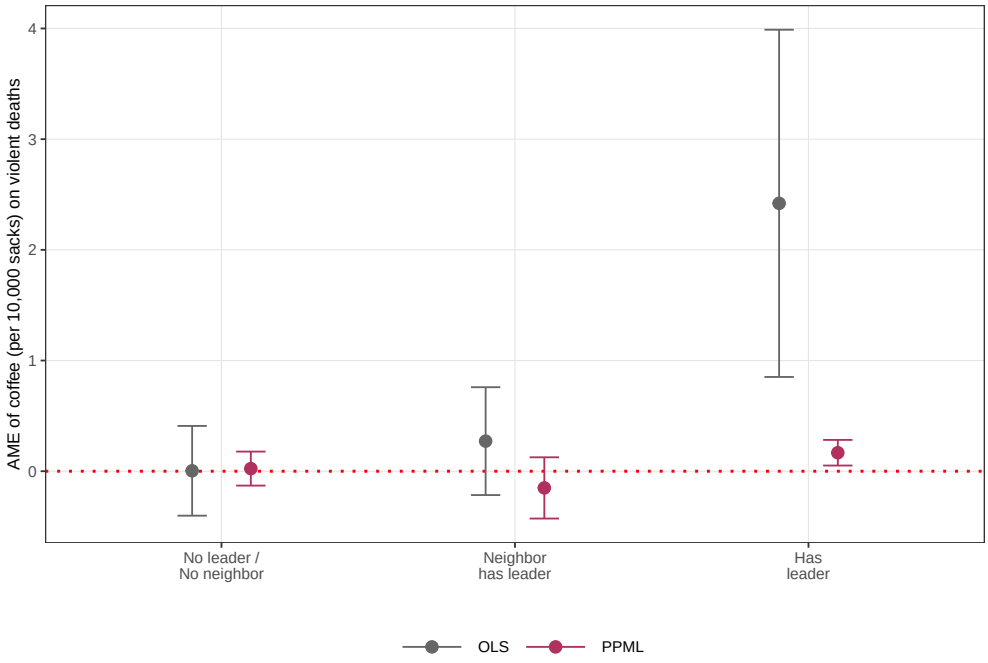


Figure 1.7. Average Marginal Effect of Coffee on Violent Deaths by Bandit Presence

Notes: Average marginal effect of coffee (10,000 sacks) on violent deaths (1962) by armed group presence. Left panel: OLS. Right panel: PPML. Whiskers show 95% confidence intervals. Leader groups: (0) no leader and no neighbor with a leader ($N_{\text{OLS}} = 452$), (1) neighbor has a leader ($N_{\text{OLS}} = 198$), (2) has its own leader ($N_{\text{OLS}} = 91$). Controls: competing theories, state controls, geographic controls, population controls, and department FE.

1.5.4 Historical Evidence: Coffee Rents and Armed Group Operations

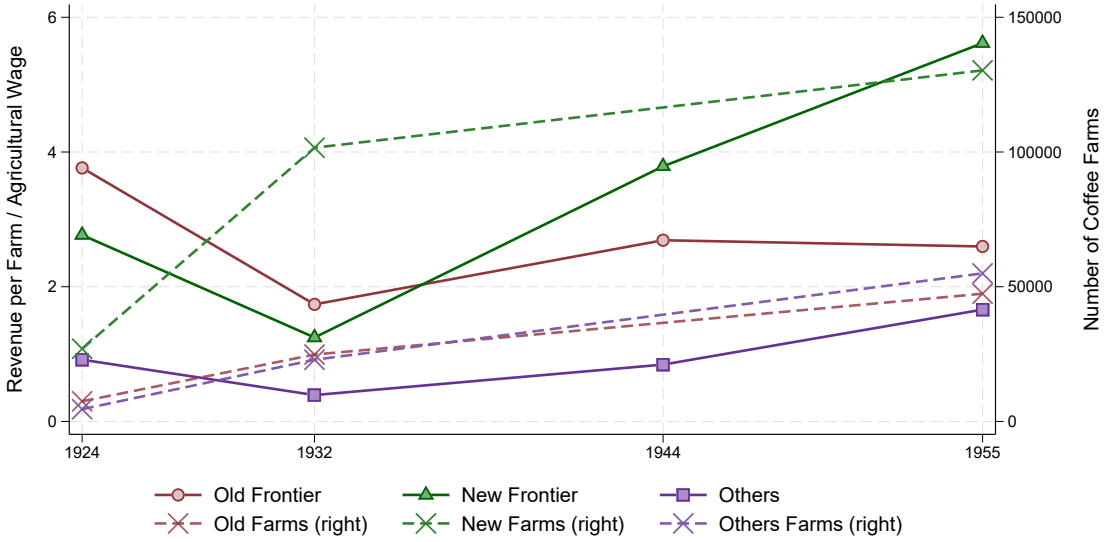
The previous sections established that coffee production is the strongest correlate of violence intensity and that armed groups mediate this association. This section asks how large the economic stakes were and how armed groups actually operated on the ground.

How Valuable was Coffee?

How appropriable was coffee in practice? A back-of-the-envelope calculation helps convey the magnitudes involved. A typical hectare of coffee land contained roughly 1,500 trees and produced around 600 kilograms of coffee per year, equivalent to approximately 10 sacks of 62.5 kg each. At prevailing prices by the 1950s, those 10 sacks were worth around 1.8 times the annual agricultural wage.[25] The median municipality in the new frontier had over 1,600 hectares of coffee land. The pool of rents available for appropriation in any given coffee municipality was large.

Figure 1.8 confirms this at the farm level, plotting the ratio of nominal coffee revenue per farm to the standard annual agricultural wage between 1924 and 1955. By the end of the period, a single new frontier coffee farm generated revenue equivalent to 5.6 annual wages, more than double the ratio in 1924. The old frontier ratio remained flat at roughly 2.6× throughout. The divergence accelerated after 1932, as international coffee prices rose and the new frontier expanded its planted area. Appendix Table A.6 reports the department-level detail and Appendix A.1.1 documents the construction methodology.

Figure 1.8. Revenue per Coffee Farm Relative to Agricultural Wage by Frontier (1924–1955)



Notes: Solid lines (left axis): ratio of nominal coffee revenue per farm to the standard annual agricultural wage. Dashed lines (right axis): total number of coffee farms per frontier group. Revenue per farm = (total production in 62.5 kg sacks × price per sack) / number of coffee farms. Agricultural wage = daily agricultural wage × 305 workdays. See Appendix A.1.1 for sources and construction details.

Coffee's dominance extended beyond individual farms. Appendix Table A.7 reports the gross farm value of all agricultural products by department in 1955, using estimates from (United Nations. Economic Commission for Latin America, 1958). Coffee alone accounted for 78% of total agricultural value in the New Frontier and over 60% nationally.

[25] Put differently, a single sack of coffee was worth roughly one-fifth of an annual wage for an agricultural laborer.

How Bandits Operated in Coffee Country

The quantitative evidence shows that coffee municipalities hosted more armed group leaders and that the coffee-violence association operates through these groups. The historiography provides a detailed account of how this operated on the ground. I summarize the main patterns here; Appendix A.1.3 collects the primary and secondary source passages on which this account is based.

The bandit groups of *La Violencia* were not a homogeneous phenomenon. They spanned the political spectrum and changed character over time. In the early phase, groups were closely tied to the Liberal and Conservative parties, serving as instruments of partisan warfare. After the 1953 military coup and the 1958 National Front agreement, formal partisan patronage dissolved (Sánchez et al., 2001). Groups that survived increasingly relied on local economic networks rather than party structures to sustain themselves. As (Sánchez et al., 2001) document, the financing of armed groups evolved through three stages: initial bankrolling by sympathetic landowners, followed by forced contributions through extortion (*boleteo*), and finally alliances with sharecroppers who found it advantageous to cooperate with the groups controlling their area.

Coffee was central to this economic sustenance. The most direct channel was theft: armed groups timed their operations to coincide with harvest periods, stealing processed coffee from warehouses, intercepting shipments on rural roads, or harvesting abandoned farms directly (Ortiz Sarmiento, 1995; Roldán, 2002). Contemporary police reports noted that "the appearance of these gangs coincides with the coffee harvests" and that "coffee pickers and dryers are murdered, and the harvest is stolen" (Policía Nacional, 1959). As (Oquist, 1980) observed, coffee in violent areas "had to be commercialized in an armed fashion or it might be stolen several times on the way to market," though he noted that the violence "did not represent national economic losses, but it did imply changes in the appropriation of the economic benefits of the crop. The coffee always got to market" (p.229).

Beyond direct theft, armed groups embedded themselves in the rural economy through the sharecropping system. In regions where landowners had fled or been killed, sharecroppers were the only actors with a continuous physical presence on the farms. Armed groups negotiated arrangements in which farm managers (*mayordomos*) handed over the landowner's share of the harvest in exchange for protection from confiscation (Roldán, 2002). Village merchants served as intermediaries who moved stolen coffee into the export chain at half the going price (Ortiz Sarmiento, 1995). This created an illicit economy in which estate managers, sharecroppers, armed groups, and middlemen each had an interest in maintaining the system. Violence in this context could take the form of outright pillaging and killing, particularly in areas of opposing partisan alignment where peasants and laborers were targeted to clear territory, or it could function as a more stationary arrangement of taxation and rent-sharing in areas with supportive local networks. In either case, coffee rents provided the economic base

that sustained armed group operations during the later phase of *La Violencia*, as "booty in the form of coffee is easy to expropriate and market, [and] *bandoleros* could not only survive but even become profitable" (Sánchez et al., 2001, p.30).

These armed groups were not, however, purely economic actors. As (Sánchez et al., 2001) document, the economic viability that coffee rents provided thrust many bandit leaders onto the local political scene. Several had themselves been victimized farmers from the early years of *La Violencia* who took up arms in response to partisan persecution, only to find themselves, after the National Front, hunted by the very parties that had originally organized them. Within their areas of control, these leaders negotiated local disputes, provided protection, and at times furnished public goods ordinarily associated with the state. They were, in short, important political actors at the municipal level, sustained by the coffee economy that made their continued operation possible.

1.6 Conclusion

Contemporary Colombia is known for a conflict defined by non-state armed groups operating across its territory. This mode of violence is a far cry from the partisan civil wars of the nineteenth century, fought between the two established parties of Colombian politics. How does a country move from one type of conflict to the other? This paper offers one piece of the answer by examining the two largest civil conflicts in Colombian history and the coffee boom that took place between them. I find that the locations where conflict occurs persist across centuries, transmitted through partisan identity, but that the intensity of violence follows a different logic: coffee production is the strongest predictor of how many people die, conditional on all competing explanations from the historiography. The evidence is consistent with coffee rents sustaining the armed groups that concentrated violence in the new agricultural frontier.

These findings suggest a broader interpretation. In a setting of persistent state weakness, the nature of conflict is shaped by which actors can sustain themselves over time. In the nineteenth century, the absence of widespread economic resources meant that only the two established political parties had the capacity to wage war. The coffee boom changed this. By planting easily appropriable rents across vast agricultural regions, it allowed smaller non-state armed groups to enter the conflict and sustain themselves independently of the old partisan infrastructure. These groups came from all sides: Liberal guerrillas, Conservative paramilitaries, and eventually revolutionary movements, all with objectives more complex than simple grievance or greed. The expanded economic geography of Colombia permitted, for the first time, the viability of these groups.

The next fifty years of Colombian conflict will be shaped by completely different forces, particularly the arrival of a second (illicit) commodity boom in the form of cocaine. However, the pattern identified here suggests that the relationship between appropriable rents and

non-state armed activity did not begin with the drug trade. The coffee boom of the early twentieth century had already established the conditions under which small armed groups could sustain themselves through natural resource rent extraction, decades before cocaine entered the picture. Understanding how commodity booms alter the geography and character of conflict is relevant well beyond the Colombian case.

References

Arango, M. (1977). *Café e Industria, 1850-1930*. C. Valencia Editores.

Arias, L. M. and D. Flores-Peregrina (2024). Colonial agricultural estates and rural development in twentieth-century Mexico. *Economic History of Developing Regions 39*(2), 105–144.

Arocha, J. (1979). *La violencia en el Quindío: determinantes ecológicos y económicos del homicidio en un municipio caficultor*. Ediciones Tercer Mundo.

Arroyo-Abad, L. (2013). Persistent Inequality? Trade, Factor Endowments, and Inequality in Republican Latin America. *The Journal of Economic History 73*(1), 38–78.

Arroyo-Abad, L. and N. Maurer (2021). History Never Really Says Goodbye: A Critical Review of the Persistence Literature. *Journal of Historical Political Economy 1*(1), 31–68.

Becker, S. O., P. D. Boll, and H.-J. Voth (2025). Spatial Unit Roots in Regressions: A Practitioner's Guide and a Stata Package.

Bejarano, J. A. (1979). *El Régimen Agrario de La Economía Exportadora a La Economía Industrial*. La Carreta.

Bejarano, J. A. (1987). El despegue cafetero (1900-1928). In *Historia Económica de Colombia*, pp. 195–232.

Bergquist, C. (1986a). *Labor in Latin America: Comparative Essays on Chile, Argentina, Venezuela, and Colombia* (1st US - 1st Printing edition ed.). Stanford, California: Stanford Univ Pr.

Bergquist, C. W. (1986b). *Coffee and Conflict in Colombia, 1886-1910*. Duke University Press.

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

Besley, T. and M. Reynal-Querol (2014). The Legacy of Historical Conflict: Evidence from Africa. *American Political Science Review 108*(2), 319–336.

Betancourt, D. E. (1990). *Matones y cuadrilleros: Origen y evolución de la violencia en el occidente colombiano, 1946-1965* (1. ed edition ed.). Bogotá, Colombia: Tercer Mundo Editores.

Beyer, R. C. (1947). *The Colombian Coffee Industry: Origins and Major Trends, 1740-1940*. Ph. D. thesis, University of Minnesota.

Bolt, J. and J. L. van Zanden (2025). Maddison-style estimates of the evolution of the world economy: A new 2023 update. *Journal of Economic Surveys 39*(2), 631–671.

Brew, R. (1977). *El Desarrollo Económico de Antioquia Desde La Independencia Hasta 1920*, Volume 38 of *Archivo de La Economía Nacional*. Tall. Gráf. del Banco de la República.

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

Bulmer-Thomas, V. (2003). *The Economic History of Latin America since Independence* (2 ed.). Cambridge Latin American Studies. Cambridge: Cambridge University Press.

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

Cataño, G. (2012). EL CAFÉ EN LA SOCIEDAD COLOMBIANA. *Revista de Economía Institucional 14*(27), 255–272.

Chacón, M. (2004). DINÁMICA Y DETERMINANTES DE LA VIOLENCIA DURANTE "LA VIOLENCIA" EN COLOMBIA.

Chacón, M., J. A. Robinson, and R. Torvik (2011). When is Democracy an Equilibrium? Theory and Evidence from Colombia's La Violencia. *Journal of Conflict Resolution 55*(3), 366–396.

Chacón, M. and F. Sánchez (2003). POLARIZACIÓN POLÍTICA Y VIOLENCIA DURANTE "LA VIOLENCIA"; 1946-1963.

Cioni, M., G. Federico, and M. Vasta (2021). Chapter 2 - The two revolutions in economic history. In A. Bisin and G. Federico (Eds.), *The Handbook of Historical Economics*, pp. 17–40. Academic Press.

Collier, P. and A. Hoeffler (2004). Greed and grievance in civil war. *Oxford Economic Papers 56*(4), 563–595.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1–45.

Conley, T. G. and M. Kelly (2025). The standard errors of persistence. *Journal of International Economics 153*, 104027.

Deas, M. (1978). A Colombian coffee estate: Santa Bárbara, Cundinamarca, 1870–1912. In I. Rutledge and K. Duncan (Eds.), *Land and Labour in Latin America: Essays on the Development of Agrarian Capitalism in the Nineteenth and Twentieth Centuries*, Cambridge Latin American Studies, pp. 269–298. Cambridge: Cambridge University Press.

Deas, M. (1982). The Fiscal Problems of Nineteenth-Century Colombia. *Journal of Latin American Studies 14*(2), 287–328.

Deas, M. (2011). Insecurity and economic development in Colombia in the 1st century of independence. *Revista de Historia Economica - Journal of Iberian and Latin American Economic History 29*(2), 183–212.

Delgado, J. E. (2016). *COUNTER-INSURGENCY IN COLOMBIA: THE ORIGINS AND EVOLUTION OF A MILITARY STRATEGIC TRADITION*. Unpublished PhD Thesis, KIng's College London.

Dell, M. (2010). The persistent effects of Peru's mining *Mita*. *Econometrica 78*(6), 1863–1903.

Dell, M. and B. A. Olken (2020). The development effects of the extractive colonial economy: The Dutch Cultivation System in Java. *Review of Economic Studies 87*(1), 164–203.

Delpar, H. (1981). *Red Against Blue: The Liberal Party in Colombian Politics, 1863 - 1899*. University of Alabama Press.

Departamento Administrativo Nacional de Estadística and A. Charry Lara (1962). *Directorio Nacional de Explotaciones Agropecuarias (Censo Agropecuario) 1960: Resumen Nacional*. Bogotá: DANE.

Departamento Administrativo Nacional de Estadística (DANE) (1967). *XIII Censo Nacional de Población y II Censo Nacional de Vivienda (15 de Julio de 1964)*. Bogotá, Colombia: Departamento Administrativo Nacional de Estadística.

Dincecco, M., J. Fenske, and M. G. Onorato (2019). Is Africa Different? Historical Conflict and State Development. *Economic History of Developing Regions 34*(2), 209–250.

Dube, O. and J. F. Vargas (2013). Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *The Review of Economic Studies 80*(4), 1384–1421.

Durán y Díaz, J. ([1794] 2012). *Estado General de todo el Virreynato de Santafé de Bogotá en el presente año de 1794*. Bogotá: Banco de la República.

Faguet, J.-P., C. Matajira, and F. Sánchez (2024). *Encomienda*, the colonial state, and long-run development in Colombia. *The Economic Journal*.

Fearon, J. D. and D. Laitin (2014). Does Contemporary Armed Conflict Have 'Deep Historical Roots'?

Federación Nacional de Cafeteros de Colombia (1933). *Boletín de Estadística de La Federación Nacional de Cafeteros, Número 5*, Volume Vol. I. Federación Nacional de Cafeteros de Colombia.

Federación Nacional de Cafeteros de Colombia (1943). *Boletín de Estadística de La Federación Nacional de Cafeteros, Número 24*, Volume Vol. II. Federación Nacional de Cafeteros de Colombia.

Federación Nacional de Cafeteros de Colombia (1945). *Boletín de Estadística de La Federación Nacional de Cafeteros, Número 26*, Volume Vol. VI. Federación Nacional de Cafeteros de Colombia.

Fergusson, L., A. M. Ibáñez, and J. F. Riaño (2020). Conflict, Educational Attainment, and Structural Transformation: La Violencia in Colombia. *Economic Development and Cultural Change 69*(1), 335–371.

Fergusson, L. and J. F. Vargas (2023). Colombia: Democratic but Violent? In F. Valencia Caicedo (Ed.), *Roots of Underdevelopment: A New Economic and Political History of Latin America and the Caribbean*, pp. 285–316. Cham: Springer International Publishing.

Fuentes-Vásquez, M. J. and I. España-Eljaiek (2023). Coffee tastes bitter: Education and the coffee economy in Colombia in the late nineteenth and early twentieth centuries. *European Review of Economic History 27*(2), 174–195.

García Jimeno, C. (2005). Colonial institutions and long-run economic performance in Colombia: Is there evidence of persistence? *Documentos CEDE No. 59 Octubre de 2005*.

García Márquez, G. (1967). *Cien años de soledad*. Buenos Aires: Editorial Sudamericana.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine 27*(15), 2865–2873.

Grueso, E. R. (1950). *Geografía guerrera de Colombia*. Impr. Bolivariana.

Gutiérrez Tovar, G. (1961). *Cinco Años de Criminalidad Aparente (1955-1959): Contribución al Estudio de La Criminalidad En Colombia*. Bogotá: Ministerio de Justicia - República de Colombia.

Guzmán, G., O. Fals Borda, and E. Umaña Luna (1962). *La Violencia En Colombia: Estudio de Un Proceso Social. Tomo I*. Number 10 in Colección Actores de La Violencia En Colombia. Bogotá, Colombia: Ediciones Tercer Mundo.

Guzmán, G., O. Fals Borda, and E. Umaña Luna (1964). *La Violencia En Colombia: Estudio de Un Proceso Social. Tomo II* (Segunda edición ed.). Number 11 in Colección Actores de La Violencia En Colombia. Bogotá, Colombia: Ediciones Tercer Mundo.

Henderson, J. D. (2010). *When Colombia Bled: A History of the Violencia in Tolima* (Illustrated edition ed.). University, Ala: University Alabama Press.

Horan, P. M. and P. G. Hargis (1995). County longitudinal template, 1840-1990.

Jaramillo Castillo, C. E. (1991). *Los guerrilleros del novecientos*. CEREC.

Jiménez, M. F. (1985). *The Limits of Export Capitalism: Economic Structure, Class and Politics in a Colombian Coffee Municipality, 1900-1930*. Harvard University.

Jimenez, M. F. (1989). Traveling Far in Grandfather's Car: The Life Cycle of Central Colombian Coffee Estates. The Case of Viota, Cundinamarca (1900-30). *The Hispanic American Historical Review 69*(2), 185.

Jiménez, M. F. (1995). From plantation to cup: Coffee and capitalism in the United States, 1830–1930. In *Coffee, Society, and Power in Latin America*, pp. 38–64.

LeGrand, C. (1986). *Frontier Expansion and Peasant Protest in Colombia, 1850-1936* (1st ed ed.). University of New Mexico Press.

Lopez-Uribe, M. D. P. and F. Sanchez Torres (2024). Ideology and Rifles: The Agrarian Origins of Civil Conflict in Colombia. *World Development 173*, 106387.

Lowes, S. and E. Montero (2021). Concessions, violence, and indirect rule: Evidence from the Congo Free State. *Quarterly Journal of Economics 136*(4), 2047–2091.

Mazzuca, S. (2021). *Latecomer State Formation: Political Geography and Capacity Failure in Latin America*. Yale University Press.

Mazzuca, S. and J. A. Robinson (2009). Political Conflict and Power Sharing in the Origins of Modern Colombia. *Hispanic American Historical Review 89*(2), 285–321.

McGuirk, E. and M. Burke (2020). The Economic Origins of Conflict in Africa. *Journal of Political Economy 128*(10), 3940–3997.

Meisel-Roca, A. and J. E. Romero-Prieto (2017). La mortalidad de la Guerra de los Mil Días, 1899-1902. *Cuadernos de Historia Económica* (15452).

Ministerio de Guerra, Policía Nacional (1959). *Criminalidad Colombiana 1959*. Bogotá: Departamento de Información Criminal y Estadística.

Monsalve, D. (1927). *Colombia Cafetera*. Barcelona: Artes Gráficas.

Musacchio, A., A. M. Fritscher, and M. Viarengo (2014). Colonial Institutions, Trade Shocks, and the Diffusion of Elementary Education in Brazil, 1889–1930. *The Journal of Economic History 74*(3), 730–766.

Nieto Arteta, L. E. (1958). *El Café En La Sociedad Colombiana*. Villegas.

Nugent, J. B. and J. A. Robinson (2010). Are factor endowments fate? *Revista de Historia Económica / Journal of Iberian and Latin American Economic History 28*(1), 45–82.

Nunn, N. and D. Puga (2012). Ruggedness: The Blessing of Bad Geography in Africa. *The Review of Economics and Statistics 94*(1), 20–36.

Ocampo, J. A. (1984). *Colombia y La Economía Mundial, 1830-1910. Economía y Demografía* (Ilustrada, reimpresa ed.). Siglo Veintiuno Editores.

Ocampo, J. A. (2011). *Colombia y la economía mundial 1830-1910*. Bogotá: Ediciones Uniandes.

Oquist, P. H. (1980). *Violence, Conflict, and Politics in Colombia*. New York : Academic Press.

Ortiz Sarmiento, C. M. (1995). *Estado y Subversión En Colombia: La Violencia En El Quindío Años 60*, Volume 4 of *Historia Contemporánea*. Cerec.

Palacios, M. (2002). *Coffee in Colombia, 1850-1970: An Economic, Social and Political History*. Cambridge University Press.

Palau, L. (1900). Cuadro sinoptico de los combates librados hasta la fecha en la presente guerra.

Park, J. W. (1985). *Rafael Nunez and the Politics of Colombian Regionalism, 1863-1886*. Baton Rouge: Louisiana State Univ Press.

Parsons, J. (1968). Antioqueno Colonization in Western Colombia, Revised Edition by James J. Parsons - Paper. https://www.ucpress.edu/books/antioqueno-colonization-in-western-colombia-revised-edition/paper.

Policía Nacional (1959). *Criminalidad Colombiana 1959*. Number No. 5. Bogotá: Editorial El Voto Nacional.

Policía Nacional (1962). *Criminalidad Colombiana 1962*. Number No. 5. Bogotá: Editorial El Voto Nacional.

Policía Nacional de Colombia (1965). *Estadística de Criminalidad En 1965*. Number No. 8. Bogotá: Editorial El Voto Nacional.

Restrepo, J. S. C. (2001). Nieto Arteta y el café en la sociedad colombiana. *Memoria y Sociedad 5*(10), 75–101.

Roldán, M. (2002). *Blood and Fire: La Violencia in Antioquia, Colombia, 1946-1953*. Duke University Press.

Romero-Prieto, J. E. and A. Meisel-Roca (2019). Análisis demográfico de la Violencia en Colombia. Technical report, Banco de la República, Bogotá, Colombia.

Safford, F. and M. Palacios (2002). *Colombia: Fragmented Land, Divided Society*. Oxford University Press.

Saleh, M. (2024). Trade, Slavery, and State Coercion of Labor: Egypt during the First Globalization Era. *The Journal of Economic History 84*(4), 1107–1141.

Sánchez, F., M. D. P. López-Uribe, and A. Fazio (2010). Land Conflicts, Property Rights, and the Rise of the Export Economy in Colombia, 1850–1925. *The Journal of Economic History 70*(2), 378–399.

Sánchez, G., D. Meertens, and A. Hynds (2001). *Bandits, Peasants, and Politics: The Case of ""La Violencia"" in Colombia*. Austin: University of Texas Press.

Sociedad de Productores de Café and Sociedad de Agricultores de Colombia (1923). *Revista Nacional de Agricultura. Órgano de La Sociedad de Productores de Café*, Volume Vol. 230. Bogotá: La Sociedad.

United Nations (1961). *Demographic Yearbook*, Volume 8. United Nations.

United Nations. Economic Commission for Latin America (1958). *Coffee in Latin America: Colombia and El Salvador*, Volume 1 of *Coffee in Latin America: Productivity Problems and Future Prospects*. United Nations.

Uribe-Castro, M. (2021). Caffeinated Development: Exports, Human Capital, and Structural Transformation in Colombia. *mimeo*.

Urrutia-Montoya, M. et al. (2000). Comercio exterior y actividad económica de Colombia en el siglo XX: Exportaciones totales y tradicionales. Borradores de Economía 163, Banco de la República.

Vázquez, L. O. (1955). *Industria y Protección En Colombia: 1810-1930*. ESF.

Voth, H.-J. (2021). Persistence – myth and mystery. In A. Bisin and G. Federico (Eds.), *The Handbook of Historical Economics*, pp. 243–267. Academic Press.

Wallis, J. M. Q. (1919). *Memorias autobiograficas, historico-politicas y de caracter social*. Tipografia Italo-orientale.

Williamson, J. G. (1999). Real wages, inequality and globalization in latin america before 1940. *Revista de Historia Economica - Journal of Iberian and Latin American Economic History 17*(S1), 101–142.

Williamson, J. G. (2002). LAND, LABOR, AND GLOBALIZATION IN THE THIRD WORLD, 1870–1940. *The Journal of Economic History 62*(1), 55–85.

Chapter 2

Farm Size and Income Distribution of Latin American Agriculture: New Perspectives on an Old Issue[1]

2.1 Introduction

The high and persistent income inequality in Latin America and the Caribbean (LAC) has coexisted with high poverty rates in rural areas, a widening rural-urban income gap, and a shrinking rural population. The rapid urbanization of the region has contributed to improve the condition of millions of people and placed the focus of public policies on cities and urban dynamics. Studies of inequality in LAC mostly ignore the contribution of rural dynamics on income inequality and, in particular, the role of the agricultural sector, implicitly assuming that low agricultural incomes and the large agricultural and non-agricultural income gap have little influence on current income inequality.

The purpose of this chapter is to gauge the contribution of agricultural dynamics on current inequality in LAC, and explore whether the interaction between farm size, agricultural productivity, and agricultural income may partially explain this contribution. Agricultural dynamics may contribute to inequality through two distinct mechanisms: inequality *between* agricultural and non-agricultural sectors, and inequality *within* agricultural areas.

We first measure the contribution of agricultural dynamics on income inequality for nine countries in LAC (Bolivia, Chile, Colombia, Ecuador, El Salvador, México, Paraguay, Perú,

[1]This chapter is based on joint work with Margarita Gáfaro, Ana-María Ibáñez, and María Camila Ortiz, published in *Oxford Open Economics*.

and Uruguay). Based on Theil decompositions, we document that agricultural inequality plays a rather important role in explaining total income inequality in the nine countries. The percentage of the contribution of agricultural income inequality in total income inequality ranges from 11 percent in Uruguay to 58 percent in Bolivia. The *between-component* explains around 30 percent of the total agricultural income inequality on average, while the *within agricultural* component explains the rest. Given that the *between-component* is typically small in Theil decompositions, our result documents that it is a significant contributing factor in our sample of countries, pointing to the relative importance of the agricultural and non-agricultural income gap. Nonetheless, the *within-component* plays also an important role to explain income inequality in LAC. The evidence presented by Gomez-Lobo and Oviedo (2023) in this compendium confirm this result, with a substantial contribution of the rural-urban income gap to total inequality in LAC. We present further decompositions to show that high inequality in the agricultural sector, the *within-component*, is not explained by differences *between* agricultural regions, indicating that inequality *within* the agricultural households in a region must be explained by other factors.

Dynamics in the agricultural sector hence largely explain the contribution of rural dynamics on high inequality in the nine countries. The sources of income for the poorest and richest agricultural households confirm this finding. The income of rich agricultural households comes from both own-farm and agricultural wages, yet these households heavily diversify across different sources of non-agricultural income. Across all countries, only households at the top of the distribution diversify into non-agricultural income. The poorest 10 percent, on the contrary, mostly depend on (low) own-farm income. Other than Chile and Uruguay, only a small number of these poor households rely on agricultural wages. The agricultural labor income is quite low for the poorest agricultural households and its share starts picking up for the entire sample around the 40th decile.

Because agricultural income (defined as household own-farm income and agricultural wages) largely determines welfare in the rural sector and income inequality, we explore the influence of farm size and agricultural production on agricultural income. Empirical evidence has documented a robust “inverse relationship" between farm size and land yields.[2] Policies to reduce rural poverty have thus placed a strong priority on small farms as a vehicle to increase agricultural income. However, this emphasis ignores a second empirical regularity: output per worker increases with farm size. Both empirical regularities can originate from frictions in agricultural labor and land markets, such as transaction costs in the labor market that increase the relative costs of hiring workers compared to family workers (Barrett et al., 2010; Feder, 1985; Foster and Rosenzweig, 2022; Lamb, 2003). The high costs from hiring labor pushes small farms to overuse family labor, leading to the negative relation between farm size and land yields, and the positive one between farm size and output per worker. A higher output

[2]The evidence on the “inverse relationship" is abundant. Some examples are (Sen, 1962) (Berry et al., 1979), (Feder, 1985), (Heltberg, 1998), (Restuccia et al., 2008), (Barrett et al., 2010) and (Julien et al., 2019).

per worker for large farms can give rise to a positive correlation between average farm size in a region and average agricultural income.

We focus on farm size, the direct input in the agricultural production function, rather than land ownership. Assessing land ownership inequality faces the data limitations documented by Carranza et al. (2023), in this compendium, when measuring wealth inequality. The link between income and land inequality depends on land ownership distribution and the functioning of land markets. Thin land markets in developing countries sustain a strong correlation between land ownership and farm size (Ciamarra, 2004). We provide an analytical framework showing that equalizing farm size can reduce aggregate agricultural income and does not eliminate income inequality if farmers are heterogeneous in their productivity levels.

While exhibiting some degree of heterogeneity, LAC is unique among developing regions for its high levels of land inequality and its complex structure in the farm size distribution. With a regional average Gini of 0.84, only a couple countries worldwide exhibit similar levels of land inequality. Some Latin American countries like El Salvador and Ecuador, exhibit a farm size distribution that is skewed towards smaller farms, while in countries like Chile and Uruguay, land is concentrated in large farms, resembling the farm size distribution of high-income countries (Adamopoulos and Restuccia, 2014).

Findings in this chapter confirm the positive correlation between average farm size and average agricultural income in the nine countries except for Chile. Whilst farm size and land yields are inversely correlated for relatively small farms, farm size is positively associated with output per worker. Within countries, regions with higher average farm size also have higher average agricultural household income and average agricultural wages, after controlling for land suitability and weather conditions. Moreover, for some countries we find a positive correlation between the dispersion of agricultural income and the dispersion of output per worker, while there is no systematic relationship between the dispersion in farm size and income inequality. Based on the predictions of a theoretical framework that we propose to guide our analysis, we interpret these results as evidence that factors deviating farm size from its efficient allocation may be relevant for income inequality both between households in agriculture and households in other sectors and within agricultural households.

This chapter contributes to two strands of the economics literature. First, it adds to the a vast literature on inequality in Latin American countries and the sources driving this inequality. Current research shows that high inequality levels in LAC are related to low intergenerational mobility and inequality of opportunities (Bourguignon et al., 2007; de la Torre et al., 2017; Ferreira et al., 2013; Neidhöfer et al., 2018), labor dynamics and regulations (e.g. changes in skills premiums, minimum wages and wage inequality) (Alvarez et al., 2018; Ferreira et al., 2018; Levy and Schady, 2013; Lustig et al., 2013, 2016; Maurizio and Vázquez, 2016; Messina and Silva, 2019; Rodríguez-Castelán et al., 2016), inequality in educational attainment (Bouillon et al., 2003; Cruces et al., 2014; Ferreira and Gignoux, 2014; Ferreira et al., 2008;

Neidhöfer et al., 2018), the redistributive effect of fiscal policies and social spending (Bucheli et al., 2014; Levy and Schady, 2013; Lustig, 2016; Lustig et al., 2014), regional inequality (Chauvin and Messina, 2020; Grajales et al., 2018), and institutional development (Acemoglu et al., 2007; Acemoglu and Robinson, 2012; Buonanno and Vargas, 2019; Gasparini et al., 2008). Evidence on the contribution of rural dynamics to overall income inequality is scarce. Ferreira et al. (2008) find that the shrinking income gap between urban and rural areas partially explained, among other factors, the fall on inequality in Brazil during the beginning of the 2000s. Bouillon et al. (2003) estimate that one fourth of the increase in México's income Gini between 1984 and 1994 was due to the deteriorating returns of living in rural areas. The findings in this chapter emphasizes rural dynamics, in particular, agricultural ones, play a non-trivial role in explaining income inequality in the nine countries we study, and points to the relation between farm size and agricultural productivity as a potential mechanism to explain low agricultural incomes.

Secondly, we contribute to the literature on the relation between farm size and agricultural productivity. An extensive economic literature has provided ample evidence on the relation between farm size and agricultural productivity, either land yields or output per worker (Barrett et al., 2010; Berry et al., 1979; Feder, 1985; Foster and Rosenzweig, 2022; Heltberg, 1998; Julien et al., 2019; Key, 2019; Rada and Fuglie, 2019; Restuccia et al., 2008; Sen, 1962; Sheng and Chancellor, 2019). Besides the impacts on aggregate agricultural productivity, the relation between farm size and labor productivity also plays a central role in determining the income of agricultural households. This chapter adds to this literature by documenting a strong positive association between average farm size, on the one hand, and average agricultural wages and own farm income for the countries studied. The findings suggest low labor productivity for small farms need to be tackled to reduce rural poverty as well as income inequality within the agricultural sector.

Our findings implicitly bears the question of the persistence of vast amounts of small and very-small farms in Latin America. In Bolivia, Colombia, Ecuador, El Salvador, México, and Perú, the bottom 25 percent of farms are less than or equal to 1 ha. Only Chile and Uruguay have a share of large farms (>50ha) of over 10 percent. Adamopoulos and Restuccia (2020), Adamopoulos and Restuccia (2014) and Chen et al. (2017) show that poorly designed policies and regulations lead to land misallocation and hence to a prevalence of small farm in developing countries. The evidence of this chapter highlights an additional cost of small farm size at the country level: the low income of agricultural households which may contribute to the high income inequality in the nine countries we explore.

The remainder of this chapter proceeds as follows. The next section describes the data we use for the analysis. In section 2.3, we present figures describing our sampled countries' land distribution and income distribution. Sections 2.4 and 2.5 present our conceptual framework for analyzing the relationship between farm size and income inequality and our results. We

close the chapter with some concluding remarks in section 2.6.

2.2 Data sources

Our empirical analysis covers nine countries: Bolivia, Chile, Colombia, Ecuador, El Salvador, México, Perú, Paraguay, and Uruguay (See map in Figure 2.1). We use two main sources of data: Agricultural Censuses and cross-sectional labor surveys. We selected the countries based on the availability of the Agricultural Censuses and the access to its microdata. In this section, we provide a general description of the data and the variables we created. In Appendix Section B.1, we provide further details about the construction of our main variables.

2.2.1 Agricultural Census data

Most data on land distribution, farm size, and agricultural output and inputs come from Agricultural Censuses.[3] To ensure a minimum level of comparability across countries, we first harmonize the data. This process poses several challenges. First, Agricultural Censuses are not updated frequently and thus the year is different in each country (see Appendix Section B.1). Second, the definition of what constitutes a farm and the inclusion criteria for farms and crops are not homogeneous across countries (see Table B.1). Third, the wording of questions and time periods for specific questions differ, rendering it difficult to build identically defined measures. Lastly, the censuses do not elicit information on farm gate prices or producer prices. The harmonization of the data is therefore one important contribution of this chapter.

Land variables include the size of private farms (in hectares), the number of hectares allocated to agricultural production, and the number of hectares allocated to livestock production. In our main results, we include only farms that report having agricultural land and we use the number of hectares allocated to agricultural production as a measure of farm size. In the appendix, we present the analysis including all farms and using total farmland as an alternative measure of farm size.

We create measures for agricultural inputs, besides land, which include: (i) number of total agricultural workers, and separated for family and hired workers; (ii) whether the farm has access to an irrigation system; and (iii) number of farms with tractors. We eliminate from the sample farms that do not report the number of workers in Bolivia (1.95%), Chile (26.24%), Ecuador (17.34%), El Salvador (2.16%), México (0.005%), Paraguay (8.34%), Perú (3.85%) and Uruguay (0.28%).[4] To measure labor inputs, we would ideally have the aggregate number of hours dedicated to agricultural activities on the farm, yet time use data is not collected on Agricultural Censuses. Hence, we calculate the yearly number of workers: total, family, and

[3] Except for Perú, where the census does not provide output data and we use the Agricultural Survey (2019), which is representative at the national level, to compute output and input variables.

[4] We verify that our main results hold if we include farms without workers in the analysis. Results available upon request.

Figure 2.1. Countries covered in the chapter



Notes: The map presents in colors the countries included in the analysis. The legend shows the census and surveys years and the number of regions within each country.

hired. In Appendix Section B.1 we describe in detail this calculation.

We use the information on agricultural output to calculate land yields and output per worker. A large percentage of farms produce more than one crop. Hence, for aggregating production at the farm level, we use revenue instead of the quantity produced. We use FAO producer prices at USD constant prices of 2011 to construct measures of farm revenue. This measures cover about, on average, 30% of crops included in all censuses across countries, which constitute 92% of total production (See Appendix Section B.1 for details).

Land yield and output per worker are defined as:

$$\text{Land yield}_{ij} = \frac{\sum_{k=1}^{N} p_k \times c_{ijk}}{\text{Total hectares}_{ij}} \tag{2.1}$$

$$\text{Output per worker}_{ij} = \frac{\sum_{k=1}^{N} p_k \times c_{ijk}}{\text{Total workers}_{ij}} \tag{2.2}$$

where p_k is the price per ton of crop k produced in farm i located in state j. Total hectares$_{ij}$ is the number of hectares cultivated for the crops with price information. Total workers$_{ij}$ represents the number of hired and family workers allocated to crops with price information, which we calculate as the total number of workers multiplied by the percentage of land cultivated on crops with price information.

Land size and the use of agricultural inputs vary widely across the countries of the sample (See Table 2.1). Nonetheless, countries can be grouped in three clusters, according to the share of small farms (below five hectares) and large farms (above or equal to 50 hectares) and the area covered by these farms (See Appendix Table B.2 for land distribution by land size ranges). Chile, Paraguay, and Uruguay, the first cluster, have the lowest share of small farms (41%, 40%, and 16%, respectively), the highest share of large farms (13%, 7%, and 57%), and thus larger average farm size – 70.7 ha in Paraguay, 87.8 ha in Chile, and 364.2 ha in Uruguay. Land intensity, measured as the number of hectares per worker, is larger than for most of the other countries: 3.5 for Chile, 6.8 for Paraguay, and 45.3 for Uruguay. In the second cluster of countries, Bolivia, Colombia, Ecuador, México, and Perú, the share of small farms is larger – ranging from 58.6 percent of total farms in Bolivia to 79.2 percent in Perú, whilst the share of farms above 50 hectares is small (1.8% in Perú, 3.8% in México, 4.5% in Colombia, 6.7% in Ecuador, and 8.5% in Bolivia). Subsistence farming is apparently more widespread in these countries, leading to higher labor intensity, albeit with a wide variances across the five countries. The conditions in El Salvador, the third “cluster" are akin to those of a subsistence rural economy. Farms have an average of 2.4 hectares and labor is used intensively (0.4 ha per worker). The percentage of cultivated land, 90 percent of total farm size, suggests over-exploitation, as a only small percentage of land is left for conservation and fallow.

2.2.2 Labor surveys

Data for the analysis of agricultural household income comes from cross-sectional labor surveys in each country. We use Sociométro-BID, IDB’s Data Lake for the harmonized household surveys which contains information on wages from the main and secondary occupations, labor and non-labor income, sector of occupation, firm size, rural/urban location, and other individual and household characteristics. We use the household survey of the same year than the Agricultural Census, except for Ecuador, México, and Paraguay, where the

Table 2.1. Average land size and input use

Variable	Bolivia	Chile	Colombia	Ecuador	El Salvador	México	Paraguay	Perú	Uruguay
Total farm size (hectares)	38.35	87.83	19.71	14.94	2.37	16.76	70.71	7.54	364.17
	(460,98)	(1,515.97)	(599.85)	(130.14)	(15.23)	(250.08)	(1,065.81)	(159.55)	(872.24)
Total number of workers	6.10	7.67	1.94	3.19	4.62	2.79	3.77	8.56	2.80
	(23.87)	(44.45)	(6.39)	(9.17)	(16.57)	(4.74)	(4.80)	(44.43)	(7.36)
Number of hired workers	2.99	6.06	1.15	0.91	4.30	1.28	1.17	6.60	1.25
	(20.55)	(51.76)	(6.42)	(9.16)	(98.89)	(4.66)	(4.67)	(44.44)	(7.37)
Number of family workers	3.10	2	0.79	2.27	1.41	1.72	2.59	2.04	1.54
	(10.04)	(1.18)	(0.80)	(1.51)	(1.26)	(1.52)	(1.79)	(1.42)	(0.98)
Hectares per worker	2.01	3.50	4.58	2.74	0.29	3.72	6.84	0.79	45.30
	(16.88)	(64.79)	(110.94)	(9.21)	(0.51)	(28.86)	(58)	(23.72)	(288.57)
Cultivated area (% total farm size)	46.33	41.07	74.88	43.40	89.86	84.70	51.87	62.61	24.85
	(0.33)	(0.32)	(0.34)	(0.33)	(0.22)	(0.27)	(0.27)	(0.34)	(0.26)
% Farms with irrigation	31.32	46.25	12.26	29.48	3.26	15.33	1.43	43.89	8.07
% Farms using tractors			0.78		4.09	36.95			
% Farms with own tractors	3.21	12.84		1.16			5.68	0.39	46.96
% Farms with rented tractors	35.91	24.72		0.16			13.71	22.66	

Notes: The sample includes all farms that report at least one worker. Farm size is measured as total farm area. Hectares per workers are calculated using cultivated hectares.
Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012) ; Uruguay - CGAXIX (2011).

earliest household surveys with the representativeness required for the analysis were conducted later (2007, 2016, and 2017, respectively).

For studying the link between farm size and agricultural income, we define agricultural income as the sum of household agricultural wages and household own farm income. We rely on two methods to construct own farm income. For Bolivia, Chile, Colombia, Ecuador, and Uruguay the surveys do not include a separate module for agricultural producers. Hence, we construct agricultural income by aggregating wages of all households members that report being employed in the agricultural sector and the income of those self-employed in the agricultural sector, assuming the latter are working on their own farm. We convert income variables to USD constant prices using 2011 as base year and Purchasing Power Parity (PPP) for the same year.

In El Salvador, México, and Perú, which have a separate module to estimate farm income, we calculate net farm income subtracting the costs of farm production from the reported gross farm revenue. We winzorize the highest and lowest one percent of this income for the econometric analysis. Since the surveys for El Salvandor and Perú also collect information for self-employed, we estimate agricultural income using both methods to gauge the difference in incomes from both modules. The household income calculated with the agricultural module is consistently lower than with the self-employed module: the former is 74 percent, and 72 percent of the latter for El Salvador and Perú.[5] Agricultural household income sums agricultural wages and own farm income for all household members.

The analysis in section 2.3 separate households into agricultural and non-agricultural

[5]The survey for Paraguay also includes an agricultural module. However, 54% of households report agricultural incomes equal to zero, while reporting costs. Given this, we use the self-employment module.

households. We define a household as agricultural when at least one household member works in the agricultural sector.

Merging agricultural variables with household income variables

We combine the information from the Agricultural Censuses with the information from the labor surveys by computing and merging aggregates at the lowest geographical level for which the labor surveys are representative. These correspond to *departamentos* (Bolivia, Colombia, Salvador, Paraguay, Perú and Uruguay), *provincias* (Ecuador), *regiones* (Chile), *entidad federativa* (México). In general, these geographical units are comparable to US states. For some countries, the surveys are not representative for all states, forcing us to drop these states from the analysis.[6] In most of the countries, the surveys are representative at the regional level, yet not at the region-rural level. Thus, agricultural wages and agricultural household income are at the state level, without disaggregating for urban and rural areas, unless noted. Nonetheless, most agricultural workers and agricultural producers in the nine countries live in rural areas.[7] In addition, this allows us to have better comparability as the definition for the rural population differs across country (see Table B.3).

Table 2.2 describes the conditions of the rural sector and its contribution to poverty and GDP. The share of the rural population in LAC is 19 percent, yet this masks high heterogeneity across countries. In our sample, the share of the rural population ranges from 4.6 percent in Uruguay to 38.1 percent in Paraguay. The contribution of agriculture to total GDP and employment is much larger than for OECD countries (1.4% and 5% respectively, per 2023 World Bank data) and varies widely across the sampled countries. The agricultural sector still employs a significant share of the population with wages well below non-agricultural wages. For example, in Ecuador, a country with a large share of employment in agriculture, the agricultural wage is 39 percent of the non-agricultural wage. Rural poverty is much higher than urban poverty in the all countries, except for Chile.

2.2.3 Other controls

We include in the analysis geographical controls to capture land quality and agro-climatic conditions. We construct these variables at the state level. Data for agricultural suitability comes from the FAO - GAEZ Data Portal which measures the agroclimatic potential yield for

[6] In Colombia, we drop Arauca, Casanare, Putumayo, San Andrés, Amazonas, Guainía, Guaviare, Vaupés and Vichada. In Ecuador, we eliminate Galápagos and undelimited zones (Las Golondrinas, La Concordia, Manga del Cura, El Piedrero); for the Amazon region, all states are considered as a single domain (Sucumbíos, Orellana, Napo, Pastaza, Morona Santiago, y Zamora Chinchipe); and we merge Santa Elena and Santo Domingo de los Tsáchilas to their original states, as for the 2000 Agricultural Census these states had not been created yet. In Paraguay, we drop Boquerón, Alto Paraguay and the capital district (Asunción).

[7] The percentage of agricultural workers/households living in rural areas in Bolivia, Chile, Colombia, Ecuador, El Salvador, México, Paraguay, Perú and Uruguay are, respectively: 91%/87%, 53%/ 51%, 83%/80%, 86%/80%, 74%/72%, 73%/69%, 90%/87,3%, 75%/70%, 67%/63%.

Table 2.2. Rural population and agricultural sector by country

Variable	Bolivia	Chile	Colombia	Ecuador	El Salvador	México	Paraguay	Perú	Uruguay
Rural population (%)	30.22	12.35	18.89	36.01	27.25	19.55	38.12	21.90	4.57
Employment in agriculture (%)	30.54	8.97	15.77	29.73	16.29	12.47	18.71	27.37	8.4
Agricultural GDP (% GDP)	12.21	3.98	6.41	8.80	5.14	3.39	9.99	6.85	6.45
Rural poverty rate (%)	5.6	0.4	11.8	8.4	2.5	6	1.9	11.1	0.3
Urban poverty rate (%)	0.3	0.8	3.4	1.3	0.7	2.3	0.5	0.8	0.1
Monthly agricultural wage (USD PPP)	666,66	341.02	496.04	114.44	124.05	259.75	660.53	355.22	659.65
	(399.81)	(263.28)	(299.83)	(91.10)	(62.37)	(204)	(420.31)	(301.89)	(644.17)
Yearly hh agricultural income (USD PPP)	5,619.15	6,717	5,467	2,205	1,298	3.111	8,615.53	3,315.97	12,935
	(10,006)	(9,567)	(6,351)	(3,282)	(1,309)	(6,137)	(38,229.10)	(4,620.48)	(16,765)
Agricultural wage ratio (% non-agricultural wage)	68,56	49.10	51.44	39.00	42.53	39,93	70,06	45.13	67,15

Notes: Monetary values are in USD constant prices using 2011. Urban and rural poverty rate at $2.15/day (2017 PPP). Wages and income variables use data for the same years of the household surveys in each country. Other indicators use 2019 data.
Sources: https://data.worldbank.org in year 2019 for % of rural population, employment in agriculture, agricultural GDP and rural and urban poverty rates. Own calculations for monthly agricultural wage, yearly household agricultural income and agricultural wage ratio based on: Bolivia - EH (2014); Chile - CASEN (2006); Colombia - GEIH (2014); Ecuador - ENEMDU (2007); El Salvador - EHPM (2008); México - ENIGH (2016); Paraguay - EPH (2017); Perú - ENAHO (2012); Uruguay- ECH (2011).

the main staple crops.[8] This variable measures average attainable yield of current cropland at grid levels of 5 x 5 arc minutes (approximately 11 x 11 km) for the period between 1981-2010. We construct the state level agricultural suitability index with a weighted average of the grid values.

We measure weather conditions using accumulated rainfall from ERA5 (Muñoz-Sabater 2019 and Muñoz-Sabater (2021)). The data have average monthly rainfall information for 11*11 km grids. To create a state-yearly accumulated rainfall, we calculate a weighted average, using the percentage of the grid located in each state, for the monthly rainfall at the state level and then aggregate it at the year level.

2.3 Description of Agricultural Sector: Distribution of Land and Income

2.3.1 Land Inequality: Context and Estimates

Land inequality and Latin America have a long and well-documented history together. Often highlighted as the pre-eminent example of long-term persistence of land inequality and its effects on modern economic and development outcomes (Easterly, 2007; Engerman et al., 2000; Frankema, 2010; North et al., 2000; Sokoloff and Engerman, 2000), Latin America usually holds a status as *sui-generis* with respect to the magnitude of both historical and present levels of land inequality. The typical story follows that during colonization of Latin America, geographic endowments of some sort (i.e. land suitable for cash crops) and/or a certain set of colonizer-imposed institutions induced an agrarian structure consisting of extremely unequal land ownership, leading to a development path dominated by unequal political, economic and social relationships, in one way or another leading to unequal and

[8] Staple crops included are maize, corn, rice, wheat, cassava, soybeans, potatoes, sorghum, sweet potato, yams, and plantain. We selected the main based on https://www.worldatlas.com/articles/most-important-staple-foods-in-the-world.html.

depressed modern economic outcomes (Eslava and Valencia, 2023).

Even beyond this well-known institutional story, most of the literature on land inequality focuses principally on its influence as a structuring factor in long-run development. Along these lines, land inequality has been theoretically and empirically linked to delayed industrialization (Adamopoulos, 2008), conflict and social unrest (Brockett, 1992, 2019), political participation and suppression (Baland and Robinson, 2008), and inter-generational mobility (Galán, 2018; Montero, 2022), to name a few. The other major focus of the literature has been to establish the reduced-form effect of *past* land inequality on *present-day* income inequality and growth (Birdsall and Londoño, 1997; Cipollina et al., 2018; Deininger and Olinto, 1999; Deininger and Squire, 1998; Frankema, 2005). There exists a dearth of estimates on present-day land inequality and, more importantly, almost no studies analyze present-day effect of land on income.

To address this gap, we start by providing an updated snapshot of land inequality estimates of Latin America, compared to other major developing regions. We perform general pareto interpolations (Blanchet et al., 2022) for 54 countries on tabulations of landholding sizes provided by the FAO World Programme for the Census of Agriculture (Nayo et al., 2019). These interpolations allow us to recover the entire distribution with high accuracy, based only on tabulated distributional data. FAO data provides the widest country coverage for land inequality estimates. This data only provides distributions of *operational holdings* and not *land ownership*, such that it might be the case that inequality of operational holdings underestimates the true distribution of land ownership. However, recent work has shown that census estimates of operational holding inequality and household survey estimates of land ownership inequality are highly correlated across all regions of the world including Latin America (Bauluz et al., 2020), mitigating concerns that FAO data does not accurately measure land inequality. Thin land rental markets can explain this strong correlation between the distribution of operational holdings and land ownership and, as we discuss in the next section, can generate distortions that deviate farm size from the efficient allocation.[9]

This allows us to calculate Gini coefficients for low and middle income countries based on consistent and comparable data. We present the results in Table 2.3. Latin America and the Caribbean (LAC) stands out with an average Gini of 0.84. The second most unequal region is the Middle Eastern and North-African (MENA) (0.73), followed by Asia (0.56) and Sub-Saharan Africa (0.51). Other than Saint Lucia, all of the LAC countries presented make up 18 of the 23 most land unequal countries. Furthermore, the even more staggering difference between regions is the average farm size in each country. Excluding Caribbean countries, only El Salvador and Guatemala have average farm holdings under 10ha. The rest of Latin American countries represent 12 of the 13 countries with highest average farm

[9] Gáfaro et al. (2012) find that using cadastral data of rural Colombia, *owner* Gini coefficients are very similar to *landholding* Gini coefficients as well.

size. Theoretically, land inequality does not necessarily imply a large average farm size and vice-versa. This combination is our first empirical observation of the specific nature of Latin American land dynamics. However, a deeper dive into their structure is required.

Table 2.3. Land Inequality Estimates: LAC, Asia, MENA, SSA

Region	Country	Gini Coefficient	Avg. Farm Size (ha)	Census year
Latin America and Carribean	Bolivia*	0.91	40.22	2013
	Brazil	0.86	67.81	2006
	Chile*	0.94	106.87	2006
	Colombia*	0.95	45.99	2014
	Costa Rica	0.82	26.74	2014
	Ecuador*	0.81	14.66	1999
	El Salvador**	0.82	2.36	2007
	Guatemala	0.85	4.52	2003
	Jamaica	0.81	1.62	2007
	México	0.85	16.85	2007
	Nicaragua	0.78	23.06	2011
	Panama	0.87	10.86	2011
	Paraguay*	0.94	107.61	2008
	Perú*	0.93	17.51	2012
	Puerto Rico	0.74	17.99	2012
	Saint Lucia	0.69	1.29	2007
	Uruguay*	0.77	365.27	2011
	Venezuela	0.87	63.81	2008
	Virgin Islands	0.85	10.85	2007
Average of Region:		**0.84**	**49.7**	
Asia	Bangladesh	0.52	.59	2008
	Cambodia	0.55	1.63	2013
	India	0.61	1.15	2010
	Indonesia	0.65	0.97	2013
	Kazakhstan***	0.90	394.33	2006
	Laos	0.44	2.41	2010
	Myanmar	0.49	2.56	2010
	Nepal	0.51	.66	2011
	Pakistan	0.64	2.59	2010
	Philippines	0.65	1.31	2012
	Thailand	0.51	3.15	2013
	Vietnam	0.63	0.75	2011
Average of Region:		**0.56**	**1.3**	
Middle East - North Africa	Algeria	0.65	8.74	2001
	Egypt	0.67	0.92	2009
	Iran	0.72	4.90	2014
	Jordan	0.82	3.33	2007
	Lebanon	0.69	1.40	2010
	Morocco	0.62	6.10	1996
	Oman	0.83	0.97	2012
	Qatar	0.91	11.91	2000
	Saudi Arabia	0.92	12.00	2015
	Tunisia	0.67	10.66	2004
	Turkey	0.59	6.10	2001
	Yemen	0.75	1.36	2003
Average of Region:		**0.73**	**5.7**	
Sub-Saharan Africa	Burkina Faso	0.45	4.40	2006
	Cabo Verde	0.52	1.07	2015
	Cote d'Ivoire	0.66	3.89	2001
	Ethiopia	0.52	1.03	2001
	Malawi	0.42	0.84	2006
	Namibia	0.38	2.88	1997
	Senegal	0.51	4.29	1999
	Seychelles	0.46	0.88	2011
	Tanzania	0.58	2.70	2007
	Uganda	0.62	2.16	1992
Average of Region:		**0.51**	**2.4**	

Notes: For all countries, micro and tabulations may not represent the same data. This arises from two main reasons: (i) the data trimming process which we perform on the micro data as described in the data sources section; and (ii) the micro-data we receive from statistical offices sometimes does not include the data on which the tabulations were created (for example: For Chile, the tabulations exclude forest and wooded regions while the micro data includes them. The inverse is true for Colombia, that is, micro-data don't include forest and wooded regions while the census tabulations do.). * **Ginis:** we also calculate the Gini based on agricultural census micro-data for the five countries in our sample which also had FAO tabulations: 0.91 (Bolivia); 0.91 (Chile); 0.90 (Colombia); 0.82 (Ecuador); 0.93 (Paraguay); 0.83 (Perú); 0.76 (Uruguay). Almost all Ginis from generalized pareto interpolations are very close in magnitude to micro-data estimates except for Peru with a decrease in the micro-based gini of 0.1. This is due to the trimming procedure discussed in the data section. * **Avg. Farm Size:** The sizeable difference in average size farm for both data sources for example in Chile (87 vs 106) and Colombia (19 vs 46) stems again from the same data discrepancies between both sources described above. ** Census Tabulations don't exist for El Salvador so micro-data from this paper was used to calculate this Gini and average farm size directly. *** Given the obvious status as an outlier, we do not include Kazakhstan in the averages for the region. If we do include it, the average Gini of Asia is 0.59 and the average farm size of Asia inflates up to 32.9 ha. Sources: Formatted FAO data used for generalized pareto interpolations was graciously provided by authors of (Bauluz et al., 2020). Original data stems of FAO WCA tabulations found in booklets for 2000 and 2010 rounds here: https://www.fao.org/world-census-agriculture/wcarounds/census/en/.

2.3.2 Land Size Structure in LAC

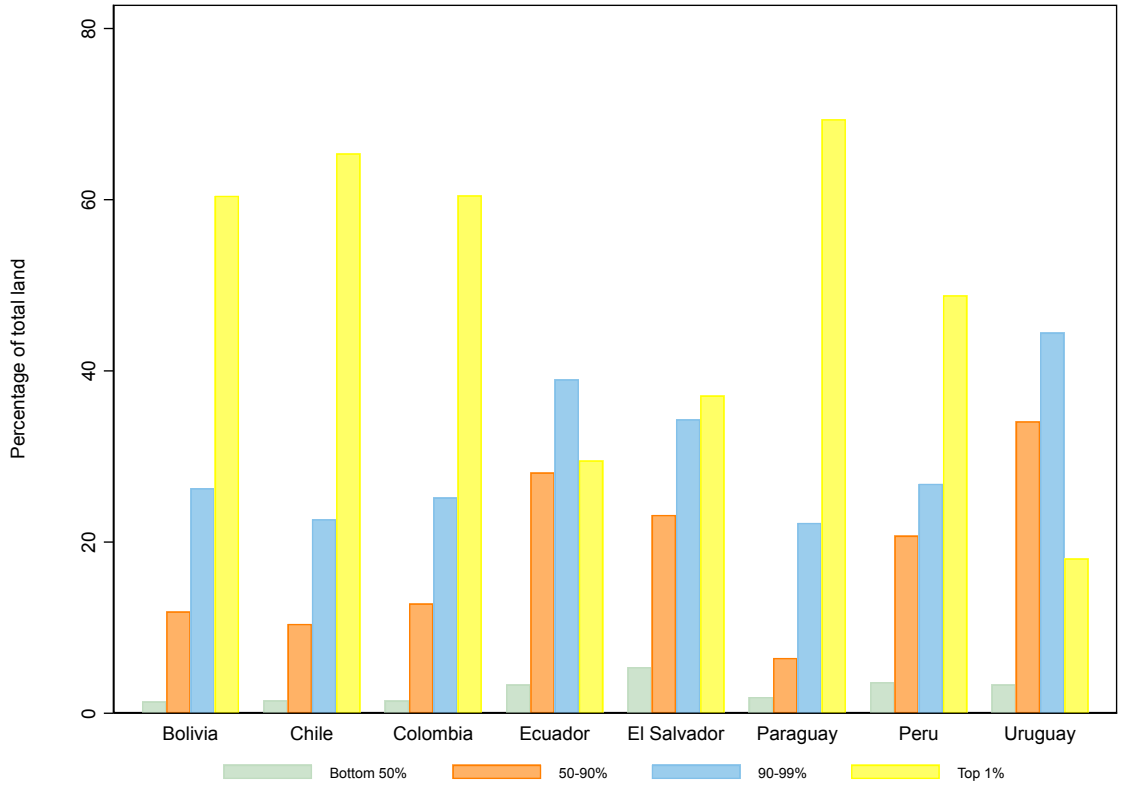
A common message of the income and wealth inequality literature is to emphasize understanding the entire distribution and not just a simple measure of inequality like a Gini coefficient. In particular, more and more attention is paid to what is happening at the top of the distribution (Alvaredo et al., 2013). In the income and wealth inequality literature, extreme accumulation at the top has mostly induced analysis searching to understand the mechanisms leading to said accumulation. As described in the previous section, research in land inequality has usually focused in understanding the long-run effects of historical land inequality, and little on the actual rates of accumulation today and their consequences. In this paper, we argue that we must also understand the different rates of accumulation at the top and bottom of the distribution as they are essential in understanding the productivity structure of agriculture sectors and hence the very rates of production of agricultural incomes and welfare.

The importance of going beyond simple inequality measures is evident when looking at Table B.2. Having an average farm size of 2.4 ha (El Salvador), or over 350 ha (Uruguay) makes an outsized difference in terms of land size structure and implies prevalence of certain types of structures such as primarily small holding (El Salvador), primarily medium-sized holdings complemented with very-large holdings (Colombia and México), or primarily very-large holdings (Uruguay). This changes in land structure directly imply different inputs, crop choice, and revenues. To better understand the extent of this land size heterogeneity across our sample, we break up the universe of farms into different population shares: bottom 50 percent, Middle 50-90 percent, 90-99 percent, and top one percent. We then analyze each share's portion in total land and average farm size.

Figure 2.2 plots the share of land per each population share, for each country. We see two main types arise: (i) one set of countries (Bolivia, Chile, Colombia, México, Paraguay, and Perú) can be categorized as *top-heavy* as their top one percent shares are all over 50 percent and their top 10 percent shares being over 70 percent, and; (ii) one set of countries (Ecuador, El Salvador, and Uruguay) can be categorized as *middle-heavy* as their Middle 50-90 percent shares are above 25 percent and most land is not concentrated only among the top one percent but among the whole top 10 percent (as evidenced by the similarity of 90-99% and top 1% shares). This classification, however, is not enough to understand the differences between countries.

Table 2.4 present average farm sizes per population share. Uruguay stands out as atypical in its structure. The average farm size of the bottom 50 percent is 10 times higher than the next highest country, and this higher order of magnitude is present across all population shares. This is being driven primarily by the reduced amount of total farms in the country, five times less than the next lowest country. We may contrast Uruguay with the other high average farm size countries, Chile and Paraguay. Compared to all countries other than Uruguay, these countries have few farms and high average size, evidenced by the large average size of the

Figure 2.2. Shares of farms by Population Groups



Notes: The sample includes all farms that report at least one worker. Farm size is measured as total farm area.
Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012); Uruguay - CGAXIX (2011).

bottom 50 percent for both countries and enormous size of the top 10 percent. While we see that average farm size is very large for all three countries, Uruguay exhibits much lower levels of accumulation due to small number of farms and larger sizes at the bottom of the distribution.

Table 2.4. Average Farm Size per Population Share (ha)

	Bottom 50%	50-90%	90-99%	Top 1%	# of Farms	Total Land	Avg. Size
Bolivia	1.04	11.31	115.52	2,298.54	844,104	32,373,152	38.35
Chile	2.63	22.85	220.94	5,744.15	218,487	19,190,198	87.83
Colombia	0.58	6.31	55.19	1,192.66	1,893,939	37,334,816	19.71
Ecuador	0.98	10.65	66.48	441.38	692,550	10,348,227	14.94
El Salvador	0.32	1.07	8.99	83.68	388,742	923,202	2.37
México	1.42	8.15	49.70	909.50	3,961,799	66,425,736	16.77
Paraguay	2.65	11.56	174.88	4879.67	261,428	18,486,282	70.71
Perú	0.54	3.83	24.81	348.42	2,123,973	16,016,540	7.54
Uruguay	24.44	310.25	1,799.94	6,548.42	44,653	16,261,345	364.71

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012) ; Uruguay - CGAXIX (2011).

Bolivia, Colombia, and México have a combination of many small-sized farms in the bottom 50 percent and very large farms for the top 10 percent. This coincides with the fact while having the same order of magnitude of total land as Chile and Paraguay, they have more than six times the total amount of farms. This implies the presence of a *dual* type of land size structure for Bolivia, Colombia and México with large amount of very small holdings and the presence of very large holdings accumulating the majority of the land. At first glance, Ecuador and Perú demonstrate very similar size patterns throughout the population shares. However, with similar total land sizes, Perú has more than three times the amount of farms, the main driver behind its much higher top 1 percent shares. Lastly, El Salvador stands alone as the sole *small-holding* example in our sample with high amount of total farms, very small farm size for most of them, with medium-size holdings (5-50 ha) up until the top one percent. Returning to the *top-heavy* and *middle-heavy* categories from above, we see how many differences in structure countries with similar top shares can exhibit.

This typology is not meant to be exhaustive, but it does provide a more precise idea of how different our sample of countries' land size structures are. This heterogeneity could easily be lost if one were to focus solely on land inequality estimates and particularly, on the very high inequality estimates across LAC. We need to remember that even the small-farm El Salvador has a higher Gini than the all but five of non-LAC countries (and a higher Gini than the country with the second-highest average farm size, Uruguay). This exercise thus intends to stress the necessity to move beyond simple measures of land inequality and emphasize the role of differences in land structure across countries.

2.3.3 Income Inequality: Concepts and Estimates

In section 2.4, we argue that there exists a relationship between farm size and productivity, and thus, a relationship between farm size and income. Before moving to this linkage, we explore how income is distributed in the rural sector, and more importantly, in the agricultural sector of our sample of countries. Income is defined at the household level unless stated otherwise.

Rural employment is composed of farm employment and non-farm employment. Given that we focus on land and agricultural production our primary interest is to understand agricultural income inequality and its contribution on total income inequality. However, given that we know that rural-urban gaps and rural inequality tend to play an important role in developing regions (Lagakos, 2020), we want to disentangle the specific role of the agricultural sector, and so we perform a rough check on the magnitude of the contribution of rural income inequality on total income inequality and compare it to the contribution of agricultural income inequality on total income inequality. We do this by performing a Mean Log Deviation (or Theil's L) decomposition of total household labor income across (i) urban and rural households, and (ii) agricultural and non-agricultural households in all countries in our sample.

The rural-urban results can be found in Table 2.5 while the agricultural-non-agricultural

results can be found in Table 2.6. We identify a sizeable amount of heterogeneity between sizes and gaps between agricultural and rural sectors. For example, we see that the rural sector is 50 percent bigger than the agricultural sector in Paraguay. An atypical case, Perú's agricultural sector is slightly bigger than its rural sector; an empirical observation mirrored by the World Bank data in Table 2.2. While relative sizes give us a first idea, we are interested in analyzing how different the decompositions for both sectors are, and later, what the agricultural-non-agricultural decomposition reveals about cross-country differences.

Table 2.5. Rural-Urban Theil Decomposition

	MLD_{Total}	$MLD_{Between}$	ELMO	MLD_{Rural}	Pop_{Rural}	MLD_{Urban}	Pop_{Urban}
Bolivia	0.56	0.08	0.26	0.74	0.34	0.36	0.66
Chile	0.50	0.01	0.08	0.51	0.13	0.48	0.87
Colombia	0.52	0.07	0.28	0.41	0.24	0.46	0.76
Ecuador	0.57	0.06	0.22	0.53	0.31	0.50	0.69
El Salvador	0.47	0.05	0.22	0.43	0.32	0.41	0.68
México	0.56	0.04	0.18	0.67	0.21	0.48	0.79
Paraguay	0.51	0.03	0.10	0.67	0.42	0.35	0.58
Perú	0.62	0.09	0.29	0.77	0.26	0.44	0.74
Uruguay	0.39	0.00	0.03	0.33	0.14	0.39	0.86

Notes: Includes all households with positive income. Rural definition is defined by each survey. The population sizes from Table 2.2 use World Bank data, while these estimates come directly from our labor surveys. Proportions are not identical, but the differences between rural are very similar.
Sources: Bolivia - EH (2014); Chile - CASEN (2006); Colombia - GEIH (2014); Ecuador - ENEMDU (2007); El Salvador - EHPM (2008); México - ENIGH (2016); Paraguay - EPH (2017); Perú - ENAHO (2012); Uruguay- ECH (2011).

Table 2.6. Agricultural - Non. Agricultural Theil Decomposition

	MLD_{Total}	$MLD_{Between}$	ELMO	$MLD_{Agricul}$	$Pop_{Agricul}$	$MLD_{Non-Agricul}$	$Pop_{Non-Agricul}$
Bolivia	0.56	0.06	0.20	0.87	0.31	0.34	0.69
Chile	0.50	0.02	0.12	0.38	0.16	0.50	0.84
Colombia	0.52	0.05	0.21	0.42	0.21	0.48	0.79
Ecuador	0.57	0.06	0.21	0.53	0.29	0.50	0.71
El Salvador	0.47	0.05	0.22	0.50	0.29	0.38	0.71
México	0.56	0.04	0.20	0.72	0.16	0.48	0.84
Paraguay	0.51	0.03	0.12	0.81	0.27	0.36	0.73
Perú	0.62	0.08	0.25	0.78	0.33	0.41	0.67
Uruguay	0.39	0.00	0.00	0.36	0.12	0.39	0.88

Notes: Includes all households with positive income. Agricultural definition was described in data sources section. The population sizes from Table 2.2 use World Bank data, while these estimates come directly from our labor surveys. Proportions are not identical, but the differences agricultural populations are very similar.
Sources: Bolivia - EH (2014); Chile - CASEN (2006); Colombia - GEIH (2014); Ecuador - ENEMDU (2007); El Salvador - EHPM (2008); México - ENIGH (2016); Paraguay - EPH (2017); Perú - ENAHO (2012); Uruguay- ECH (2011).

Comparing rural and agricultural decompositions, we first look at the $MLD_{Between}$ column which calculates the between-component for both decompositions. This number gives us an idea of the magnitude that the difference in average incomes of both groups contributes to total inequality in each decomposition. We see that the order of magnitude of both decompositions' between-components is largely equivalent.

Furthermore, we extend the decompositions using ELMO ratios (Elbers et al. (2008)). These ratios show the ratio of the observed between-component and the "maximum between-component" possible from the data. If the ratio is very small, it means that the partition has weak explanatory power to explain differences in averages of groups, while if the ratio is large, it means that that specific partition identifies a clear characteristic along which there are large differences in group averages.[10] The ELMO ratios strenghthen our previous findings as we observations that similarity in between-components of rural and agricultural decompositions are robust to the different sizes of agricultural and rural sectors.

For the within-components ($MLD_{Rural/Agricul}$ and $MLD_{Urban/Non-Agricul}$), magnitudes are again largely similar. This would seem to imply that agricultural inequality follows roughly the same dynamics as rural inequality for these countries, except for Bolivia, Chile, and Paraguay's ($MLD_{Rural/Agricul}$). Chile exhibits a higher MLD_{Rural}, while Bolivia and Paraguay a higher $MLD_{Agricul}$. This implies that Chile's agricultural sector displays less inequality than its rural sector and the opposite effect for Bolivia and Paraguay (higher agricultural inequality than rural inequality).

However, the large congruence in all other components of the decompositions implies that agricultural and rural inequality estimates largely track in our data. This result provides support to the claim that agricultural inequality is an important factor determining non-urban inequality, and furthermore, that agricultural-non-agricultural decompositions tell largely the same story as rural-urban decompositions.

Focusing now on the agricultural-non-agricultural decomposition, we analyze the contribution of agriculture in total income inequality in our sample of countries in Table 2.6. Between-component ranges from less than 1 percent percent (Uruguay) to 13 percent (Perú). In general it hovers around 10 percent. Theil between-components are naturally small, so the fact they are this big already (corroborated by ELMO ratios being usually >15%) imply that the agricultural-non-agricultural between-component is present and observable in all our countries except Uruguay. Within-component ranges from eleven percent (Uruguay) to 48 percent (Bolivia).[11] Alltogether, between and within-components related to agricultural sector account for a range between 11 percent (Uruguay), Chile (16%), Colombia (27%), México (27%), Ecuador (37%), El Salvador (41%), Paraguay (48%), Perú (54%), and 58 percent (Bolivia). Average across sample is: 35 percent. This indicates that a rough estimate of role of agriculture in total inequality in LAC is 35 percent, with evident heterogeneity across countries.

This decomposition exercise clearly delineates the two mechanisms through which the agricultural sector contributes to country-wide income inequality, which we discuss in detail

[10] These ratios give normalized and comparable estimates of how much a particular between-partition is explaining, beyond the specific size of the between-component.

[11] Within-component share of total inequality = $\frac{MLD_{Agricul} * Pop_{Agricul}}{MLD_{Total}}$

in the next section: (i) Gap in incomes *between* the agricultural and non-agricultural sector, and (ii) Unequal distribution of incomes *within* the agricultural sector.

We check if the within-agricultural income inequality is not being driven by regional inequalities in Table 2.7. Here, we perform the same MLD decomposition as before, but this time only on agricultural households, using regions as the groups across which we decompose income inequality. The analysis shows us that the maximum between-component is 18 percent in Bolivia, concurrently with a maximum ELMO ratio of 21 percent, a significant size. However, the rest of the countries show the opposite magnitude, with Colombia and México barely cracking and all other countries below 10 percent for their ELMO ratios. These results indicate that differences across regions are not driving inequality occurring in the agricultural sector other than in Bolivia. Thus, there are other factors which we are not identifying in the decomposition which are driving within-agricultural inequality and which are likely present in all regions. In the next section, we provide a conceptual framework that sheds light on these potential factors.

Table 2.7. Agricultural Sector Regional Theil Decomposition

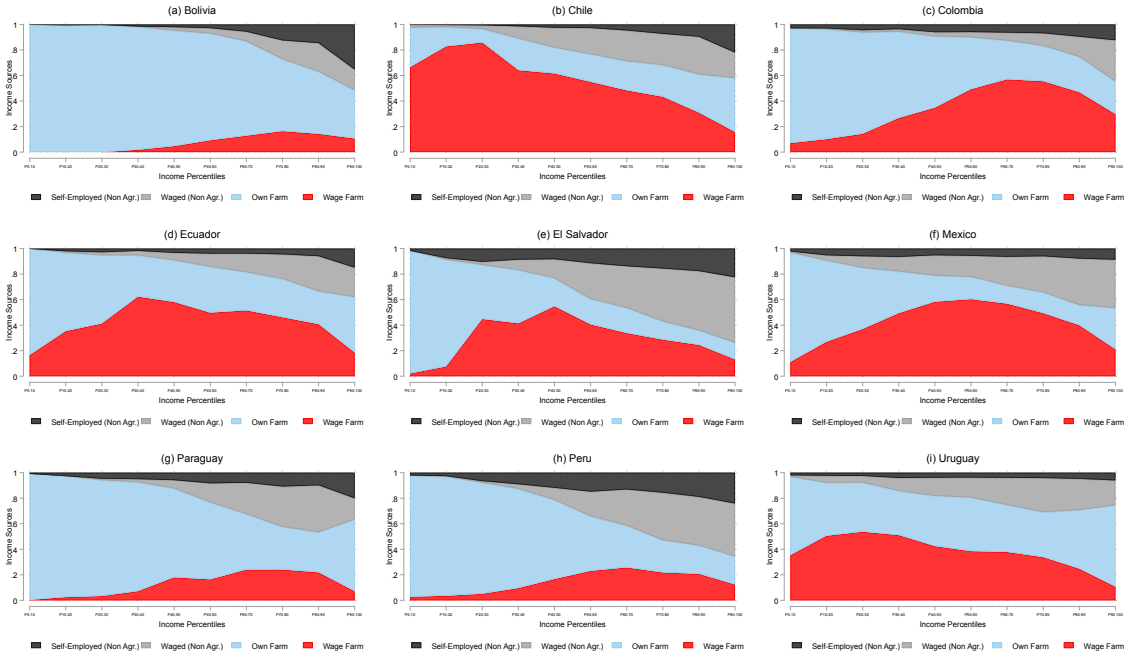
Country	$\text{MLD}_{Agricull}$	$\text{MLD}_{Between-Regions}$	ELMO	$\text{MLD}_{Within-Regions}$	# of Regions
Bolivia	0.87	0.16	0.21	0.71	9
Chile	0.38	0.02	0.06	0.36	15
Colombia	0.42	0.05	0.11	0.37	24
Ecuador	0.53	0.03	0.07	0.50	16
El Salvador	0.50	0.01	0.03	0.49	14
México	0.72	0.08	0.12	0.64	32
Paraguay	0.81	0.06	0.08	0.75	15
Perú	0.78	0.07	0.08	0.72	25
Uruguay	0.36	0.02	0.07	0.34	19

Notes: Includes all agricultural households with positive income. Agricultural definition was described in data sources section.
Sources: Bolivia - EH (2014); Chile - CASEN (2006); Colombia - GEIH (2014); Ecuador - ENEMDU (2007); El Salvador - EHPM (2008); México - ENIGH (2016); Paraguay - EPH (2017); Perú - ENAHO (2012); Uruguay- ECH (2011).

In Figure 2.3 we rank all agricultural households according to their total household agricultural income. For each decile, we calculate the share of income coming from four sources: (i) own-farm income, (ii) agricultural wages, (iii) non-agricultural self-employed income, and (iv) non-agricutural wages. We then plot the percentage of total income that each source represents at each decile. This approach provides information on the potential sources of income heterogeneity across households. We perform the same analysis at the individual level in Figure B.1 as a robustness check.

Several patterns arise across our sample of countries. First, the poorest households in all countries (except Chile and Uruguay) rely on own farm income as their main source of income and, as households become richer, they rely less on this source of income. This first result

Figure 2.3. Household Level Decomposition of Income Sources



Notes: Includes only agricultural households, whether they are rural or urban. Households are ranked in deciles across income distribution. Total agricultural income of all households in each decile are then plotted according to four different income sources.
Sources: Bolivia - EH (2014); Chile - CASEN (2006); Colombia - GEIH (2014); Ecuador - ENEMDU (2007); El Salvador - EHPM (2008); México - ENIGH (2016); Paraguay - EPH (2017); Perú - ENAHO (2012); Uruguay- ECH (2011).

implies the existence of many low-income farm households across our sample of countries.[12] Paraguay buckles the trend in the top 10 percent as own farm income dominates the very top of the distribution, mimicking Chilean and Uruguayan dynamics. This observation follows quite directly from the presence of extremely large farms in all three countries, which are able to accumulate large incomes.

Second, we see varying degrees of importance of waged agricultural work. In Chile and Uruguay, this source of income is important for the poorest households and decreases as households become richer. In Colombia, Ecuador, El Salvador, and México, this source is minimal for poorest households, increasing until around the middle of the distribution representing a very significant part of total income, and decreasing at the top of the distribution. Bolivia, Paraguay, and Perú follow a similar trend, but agricultural wages are a significant source of total income only for the top of the distribution.

Lastly, non-agricultural income is almost non-existent for the lowest parts of the distributions across all countries. As a household moves higher along the distribution, non-agricultural income (essentially) monotonically increases and becomes a significant source of income across the sample. The only country whose highest decile does not rely 30 percent or more on non-agricultural income is Uruguay. This implies that poor households do not have

[12]The Chilean and Uruguayan anomaly can partly explain their smallest agricultural-non-agricultural gaps, as these might imply the low prevalence of small, low-income farms. This will be explored more in section 2.5.

the capability to diversify income sources away from agriculture, while rich households do precisely this, and in countries like El Salvador and Peru, rely on non-agricultural income for more than half of their income.

In summary, this section has provided us with several stylized facts for our sample of countries: (i) agricultural income inequality contributes significantly to rural income inequality and explains an average 35 percent of total income inequality, with large heterogeneity across countries; (ii) the mechanisms through which agricultural income inequality affects total income inequality are the *between* agricultural-non-agricultural income gap and *within* agricultural sector inequality, and; (iii) decomposition of sources of household income points to the existence of many low-income farm households and a significant role of agricultural wages in the middle of the distribution, and high reliance on non-agricultural income for the top of the distribution. Rationalizing these empirical observations requires a structured framework, which we move to now.

2.4 Land size, agricultural productivity, and income inequality

In the previous section, we showed that both income inequality between households in agriculture and households in other sectors, and income inequality within agricultural households contribute to overall inequality in the countries of analysis. In this section, we discuss the potential mechanisms driving these income differences. We focus on agricultural productivity as a key determinant of rural income and argue that low and heterogeneous agricultural productivity within countries can explain these income differences.

The relationship between farm size and income is mediated by numerous factors. In Appendix B.2, we develop a simple theoretical framework to directly track the effect of distribution of farmer productivity on distribution of farm size and consequently, the effect on land yields. The model gives us three main results relating these variables, and hence, gives us intuition regarding the determinants of agricultural income inequality:

Result 1. In the efficient land allocation, the distribution of farmers' productivity determines the farm size distribution and more productive farmers operate larger farms. Because farmers' productivity z_i is the only source of variation across individuals in the efficient farm size, its distribution determines the farm size distribution. Moreover, optimal farm size ℓ_i^* increases with farmer's productivity z_i. Thus, farmers with higher values of z_i demand more land and operate larger farms in the efficient allocation.

Result 2. If labor is fixed across farms, in the efficient allocation, land yields are constant across farms and output per worker increases with farm size. In the efficient allocation, land yields $\frac{y^*}{\ell^*}$ depend only on aggregate farm productivity $\sum_{j=1}^{N} z_j$ and total land endowment L, and do not vary across farms. Moreover, because labor is fixed across farms

and equal to 1, output per worker is equal to total farm output y_i^* which increases with farm size ℓ^*.

Result 3. The distribution of farmers' productivity z_i determines the income distribution and equalizing farm size does not eliminate income inequality. In this simple framework, farm output determines farmers' income. Individual productivity z_i is the only source of variation of output across farms and, therefore, the dispersion in the distribution of z_i determines income inequality. Moreover, if land is evenly distributed across farmers, farm output is $\hat{y}_i = z_i^{1-\gamma} \left(\frac{L}{N}\right)^{\gamma}$, which still depends on individual productivity z_i and, therefore, the dispersion of individual income is still determined by the dispersion on individual productivity z_i.

In the remaining of this section, we present a literature review of the mechanisms highlighted by the model

The reallocation of labor across sectors

Urban-rural income gaps have been widely documented in developing countries. These gaps reflect productivity differences between workers in agriculture and other sectors, which account for most of the differences in aggregate productivity between high and low-income countries (Caselli, 2005; Gollin et al., 2014; Restuccia et al., 2008). Previous studies attribute the persistence of these differences in labor productivity across sectors to frictions that hinder the reallocation of workers across sectors and the technological progress in agriculture (Adamopoulos et al., 2022; Adamopoulos and Restuccia, 2014; Bryan et al., 2014; Lagakos and Waugh, 2013; Restuccia et al., 2008).

Some of these studies introduce models with heterogeneous agents, as in the framework presented above, and explain the persistence of low agricultural productivity with factors shaping the self-selection of low productivity agents in agriculture. These factors include a relatively low skill intensity in agriculture (Young, 2013), subsistence constraints (Lagakos and Waugh, 2013), and frictions in land and capital markets (Adamopoulos et al., 2022; Adamopoulos and Restuccia, 2014; Chen et al., 2022). These frictions to the reallocation of workers from agriculture to other sectors might induce a large heterogeneity in individual farm productivity, which as illustrated in our theoretical framework, might result both in low and heterogeneous agricultural income.

Farm size, agricultural productivity, and income

A large body of work has documented a negative relationship between land yields (the value of agricultural output per unit of land) and farm size in developing countries (see, for example, Sen (1962), Berry et al. (1979), Heltberg (1998), Feder (1985), and Lamb (2003)). This has given rise to the so-called "inverse relationship" between farm size and land productivity. Recent literature has shown that this negative relationship persists even after taking into account potential biases induced by differences in land quality across farms and measurement error (Barrett et al., 2010; Julien et al., 2019). This empirical regularity has motivated the idea

of small farms as a vehicle to reduce poverty and increase welfare in rural areas of developing countries. However, the welfare implications of the negative relationship between land yields and farm size are not straightforward and depend on the underlying causes of this relationship.

As illustrated in our theoretical framework, when analyzed within regions, a negative relationship between farm size and land yields suggests deviations from the efficient allocation of land that result in lower aggregate income in agriculture. Some studies attribute these deviations to policy-related distortions in the land market that prevent the most productive farmers from operating at their optimal scale and result in the misallocation of land (Adamopoulos and Restuccia, 2014, 2020; Chen et al., 2022). Other studies argue that the inverse relationship between land productivity and farm size can result from frictions in the labor market (Feder, 1985; Foster and Rosenzweig, 2022; Lamb, 2003). These frictions increase the costs of hired labor relative to family labor (Barrett et al., 2010; Foster and Rosenzweig, 2022). Lower labor costs in smaller farms induce a higher labor intensity, which generates higher yields per unit of land and lower output per worker in small farms (Fan and Chan-Kang, 2005; Foster and Rosenzweig, 2022; Gollin, 2019). In line with this, some studies have documented a negative relationship between farm size and labor intensity and a strong positive relationship between farm size and the value of output per worker in developing countries (Gollin, 2019). To the extent that labor productivity determines the income of agricultural workers, the positive relationship between output per worker and farm size might imply a positive relationship between farm size and household income for landless agricultural households (Fan and Chan-Kang, 2005; Foster and Rosenzweig, 2022).

Once differences in labor use are considered, it is unclear whether small farms have higher overall productivity, and are thus able to generate higher income for farmers (Rada and Fuglie, 2019). Moreover, mechanization and technological progress in agriculture allow farms larger than a particular threshold to offset the increases in labor costs and create economies of scale that might cause a positive relationship between farm size and productivity (Foster and Rosenzweig, 2022). This positive relationship has been documented in developed countries, where average farm size is substantially larger than in developing countries (Foster and Rosenzweig, 2022; Key, 2019; Rada and Fuglie, 2019; Sheng and Chancellor, 2019). Foster and Rosenzweig (2022) propose a theoretical model that combines frictions in the labor market and economies of scale from mechanization to explain the differences in the relationship between farm size and productivity across countries. According to their theoretical framework, in low-income countries where farms are small, the frictions in the labor market drive the negative relationship between farm size and productivity. As farm size increases, economies of scale emerge, and the relationship flips. Results from their counterfactual estimations with Indian data suggest that increases in farm size induce substantial increases in land yields and output per worker and decreases in the number of farms and the share of agricultural workers. They show that the decrease in agricultural workers is less than proportional to the decline

in the number of farms, which suggests that in the current land distribution, there is both an excess of labor in agriculture and an underutilization of the existing labor.

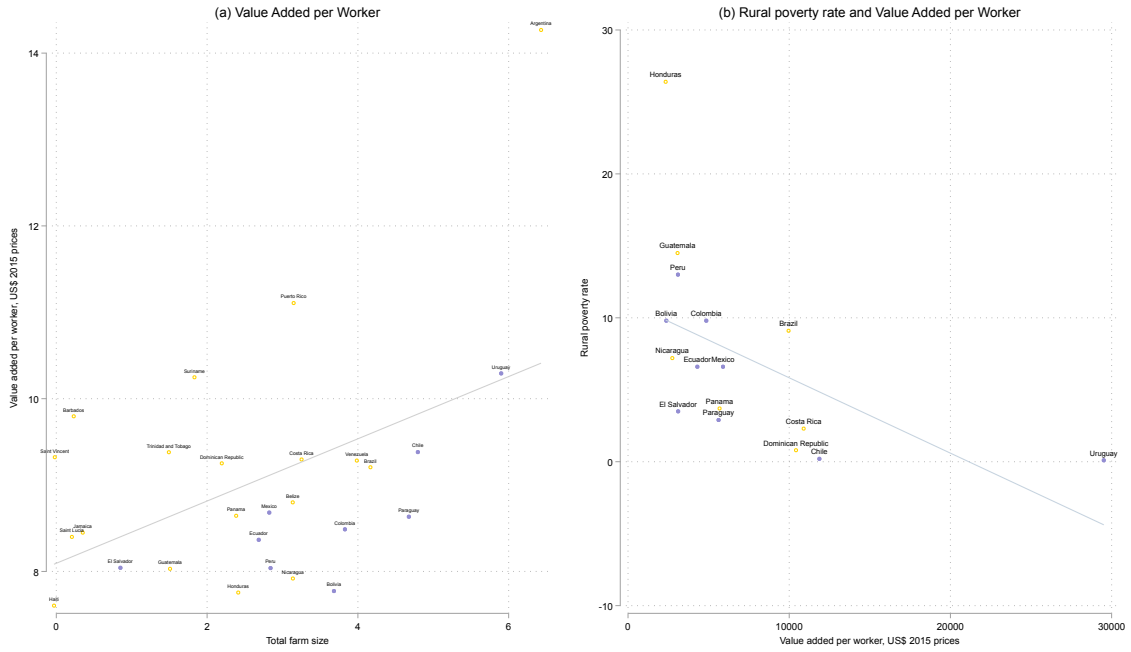
Overall, the non-null relationship between land yields and farm size, along with the persistence of small farms, despite the potential gains from land consolidation, might be associated with distortions in land markets (Adamopoulos and Restuccia, 2014, 2020; Chen et al., 2022). Moreover, eliminating these distortions can reduce income inequality. For example, Chen et al. (2022) show that eliminating the restrictions to land rentals in Ethiopia can increase income of rural households, with larger effects in low income households. Restrictions to land rentals in LAC can take various forms. In some countries there are explicit restrictions, in others, the lack of formal titles and and other institutions to resolve land disputes, hinders the development of land rental markets, creating a close link between land property and land use (Alston and Mueller, 2010; Assunção and Chiavari, 2014).

In sum, agricultural productivity determines the income of rural households through agricultural wages and farm profits and by inducing the reallocation of labor in non-agricultural activities. Lower productivity in agriculture, relative to other sectors, can explain income inequality between rural and urban areas. Moreover, the dispersion in agricultural productivity within regions can explain income inequality within rural households, which can be exacerbated by distortions in land and labor markets. In the following sections, we explore the potential role of agricultural productivity and farm size in shaping the patterns we observe in income inequality.

2.5 Rural income, agricultural labor, and farm size

We now explore the relationship between farm size, agricultural labor, and household income in our data. To motivate our analysis, Figure 2.4 shows that there is positive correlation between average farm size and agricultural value added per worker across LAC countries. Also, countries with a higher agricultural value added per worker have lower rural poverty rates. The figure also shows substantial variation across LAC in average farm size, agricultural productivity, and rural poverty. Our sampled countries cover a wide range of this variation. We now focus on these countries and present patterns underlying these aggregate relationships.

Figure 2.4. Aggregate Agricultural Productivity, Farm Size, and Rural Poverty



Notes: Panel (a) reports the relationship between log agricultural value added per worker and log average farm size. Panel (b) reports the relationship between rural poverty rate at $2.15/day (2017 PPP) and agricultural value added per worker.
Sources: Agricultural value added and average farm size (FAO), poverty (World Bank Poverty and Inequality platform).

2.5.1 Two facts about farm size and agricultural labor

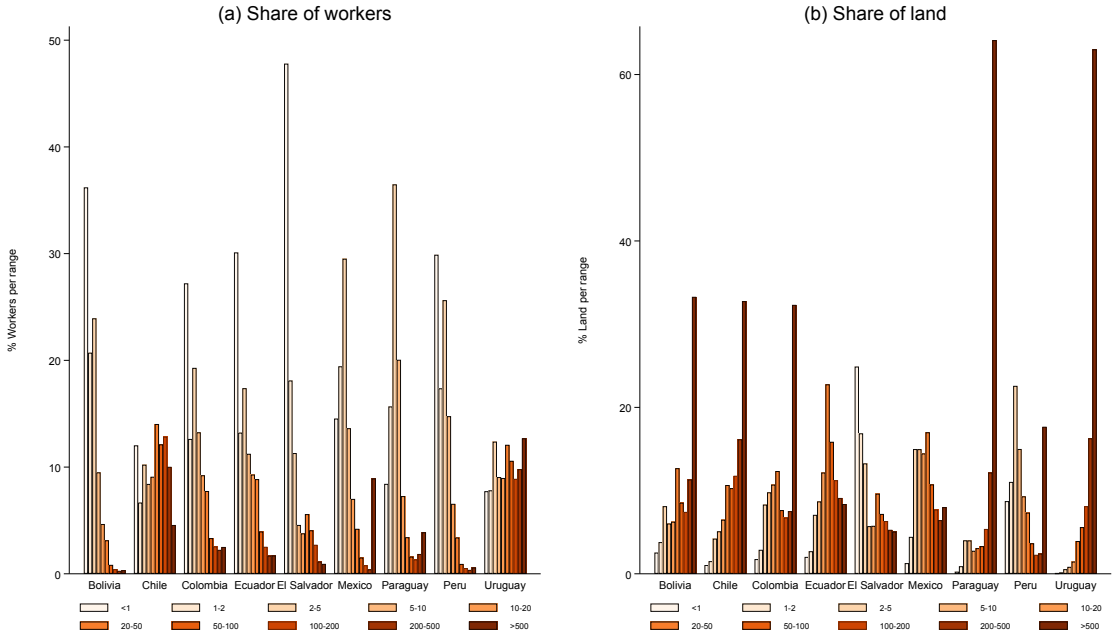
We document two facts about farm size and agricultural labor that we observe in our sample of countries and, as discussed in the previous section, can have implications on the income of agricultural households. The results we present below use cultivated area as a measure of farm size and we exclude from the sample farms that do not report cultivated area. In the appendix, we present the results including all farms and using total farm size as a measure of farm size.

Fact 1. Small farms employ large shares of workers, but only a small proportion of land.

Panel a of Figure 2.5, shows that the share of workers in farms with less than one hectare of cultivated area ranges from eight percent in Uruguay and Paraguay to 47 percent in El Salvador. Also, between 28 and 81 percent of farm workers are employed in farms with less than five hectares. Despite this, Panel b, shows that the land in these small farms ranges between 0.75 percent in Uruguay and 55 percent in El Salvador.[13] This distribution of land and labor across farms implies that labor intensity falls with farm size. As shown in Figure 2.6, farms with less than one hectare employ on average 7.32 workers per hectare; while farms between 20 and 50 hectares employ on average 0.61 workers per hectare.

[13]When we compute farm size as total landholdings instead of cultivated area, the differences are more striking. In particular, the share of workers in farms with less than 5 hectares ranges between 4.4% and 72%, and the shares of land in these farms ranges between 0.05% and 34.34% (See Appendix Figure B.2).

Figure 2.5. Share of land and labor by farm size



Notes: The sample includes all farms that report at least one worker. Farm workers include family and hired workers (permanent and seasonal). Farm size measured as cultivated area (hectares).
Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012); Uruguay - CGAXIX (2011).

Fact 2. Within regions, output per worker increases with farm size and the relationship between land yields and farm size varies across countries and farm size. Figures 2.7 and 2.8 present the relationship between farm size, output per worker and land yields, after controlling for region fixed effects.[14] Within regions, larger farms have higher output per worker. In contrast, the relationship between farm size and land yields varies across countries and farm size. In general, for the smallest farms in each country we observe a negative (Bolivia, Chile, Ecuador, México, Perú) or null (El Salvador and Colombia) relationship. For the largest farms, we observe a positive (Chile, Ecuador, El Salvador, Paraguay, Perú) and a null (Bolivia and México) relationship.[15] Appendix Figures B.3 and B.4 present the results using total land holdings as a measure of farm size.

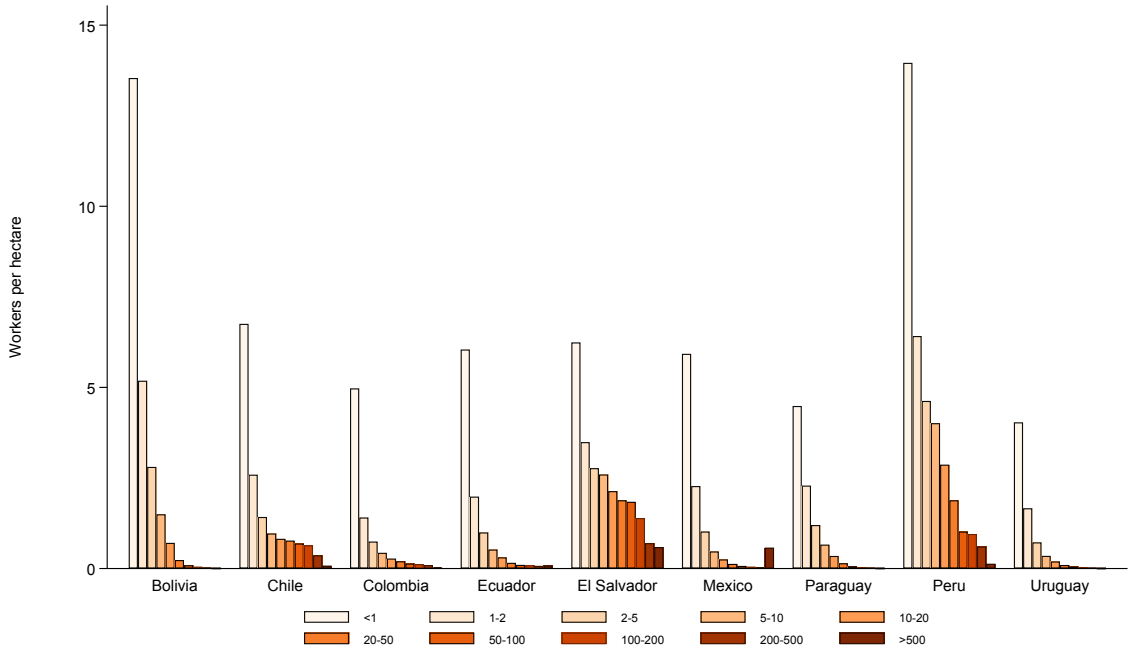
As discussed in our conceptual framework, the heterogeneity in land yields across the farm size distribution suggest the existence of distortions in land and labor markets. The negative

[14] Regions are *municipios* (Colombia, El Salvador, México, and Paraguay), *cantones* (Ecuador), and *provincias* (Bolivia, Chile and Perú).

[15] Colombia is the exception with a negative relationship between farm size and land yields for the largest farms. In Chile, we are only able to compute the value of output from cereals, leguminous vegetables, root crops, and industrial crops (see table B.1).

Appendix Figures B.5 and B.6 present similar patterns when we compare averages across regions. In the case of Chile and Ecuador, these regional aggregates mask the non-linearity in the relationship between farm size and land yields and display a null relationship between average farm size and average land yields across regions. This could be explained by heterogeneity within regions in farm size and land yields.

Figure 2.6. Labor intensity by farm size



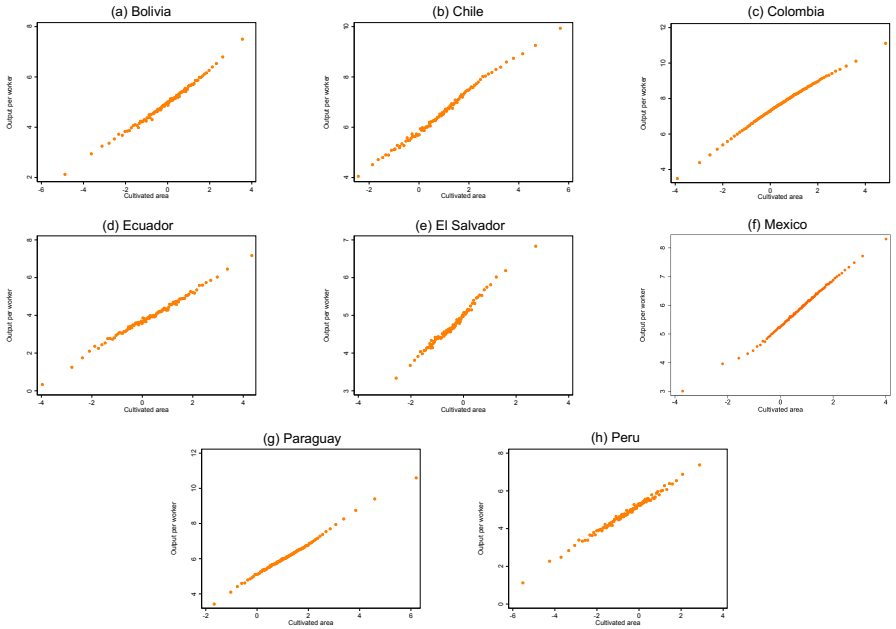
Notes: The sample includes all farms that report at least one worker. Farm workers include family and hired workers (permanent and seasonal). Farm size measured as cultivated area (hectares).
Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012); Uruguay - CGAXIX (2011).

relationship between land yields and farm size that we find for small farms in most of the countries is consistent with the empirical regularity in the "inverse relationship" literature. Similarly, the positive relationship that we find for large farms in some countries is consistent with the evidence of Foster and Rosenzweig (2022) for India and suggest the existence of economies of scale from mechanization along with market distortions that prevent the most productive farmers to accumulate land.

The null relationship for small farms in El Salvador might suggest that due to the small scale, transaction costs associated with hired labor are not significant enough to induce a negative relationship. This is also consistent with the fact that international migration has put pressure on agricultural labor supply (AENOR, 2022). Moreover, the few large farms are likely to be very different from the rest in terms of production technologies, and both exhibit economies of scale and face frictions that hinder the scale expansion.

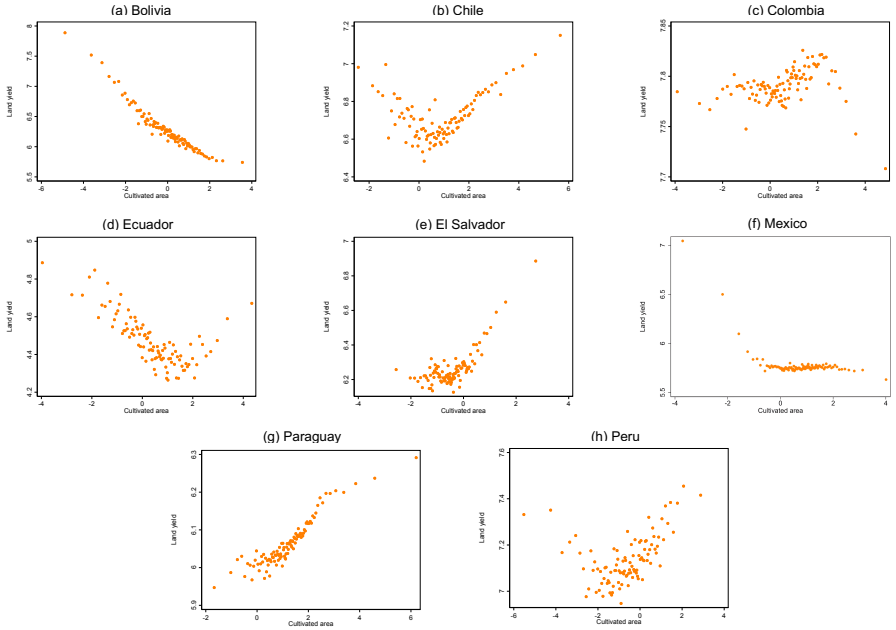
These relationships between farm size, on the one hand, and land and labor yields, on the other, shape the relationship between rural income and farm size. In the next section, we combine information from household surveys with information on farm size to explore this relationship.

Figure 2.7. Output per worker and farm size



Notes: Binned scatterplots of the relationship between log of output per worker and log of cultivated area controlling for fixed effects of regions within countries.
Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - ENA (2019).

Figure 2.8. Land yields and farm size



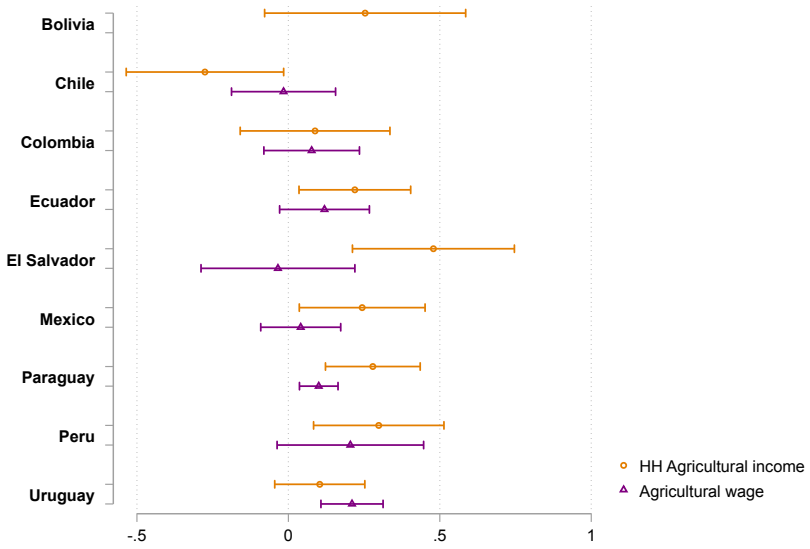
Notes: Binned scatterplots of the relationship between log of land yields and log of cultivated area controlling for fixed effects of regions within countries.
Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - ENA (2019).

2.5.2 Heterogeneity across regions and average agricultural income

We estimate a linear regression per country of the average household income from agriculture and average wages on average farm size across regions. The unit of observation in each regression is the region, which corresponds to the smallest geographical unit for which household surveys are representative: *departamentos* (Bolivia, Colombia, Salvador, Paraguay, Perú and Uruguay), *provincias* (Ecuador), *regiones* (Chile), *federal entity* (México). We control for average soil quality and rainfall in each region, and therefore, our results are not likely to be driven by differences in agro-climatic conditions across regions.

Figure 2.9 presents the coefficient estimates from these linear regressions. In all countries except for Chile, there is a positive relationship between regional averages of agricultural income and average farm size. This positive relationship is statistically significant for Ecuador, El Salvador, México, Paraguay, and Perú. In Paraguay, Perú, and Uruguay, we also find a positive and statistically significant relationship between agricultural wages and farm size.[16] Figure 2.10 shows that in Bolivia, Paraguay, and Perú there is also a positive and statistically significant relationship between the 25th percentiles of income and farm size. This implies that not only the average household, but the poorest households have higher incomes in areas where the smallest farms are larger.

Figure 2.9. Relationship between household agricultural income and farm size

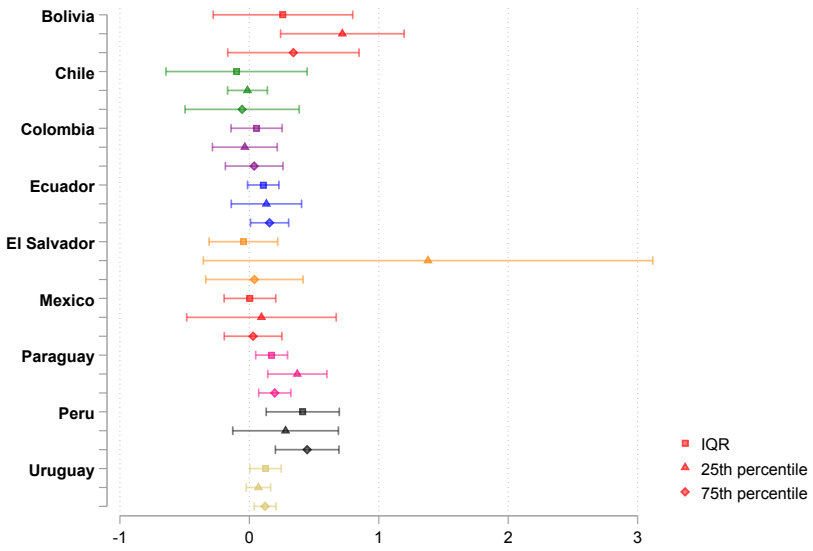


Notes: Coefficient estimates from independent linear regressions with log of average farm size (measured as cultivated area) as independent variable and log of average agricultural income and log of average agricultural wages across regions as dependent variables. Regional averages of soil quality and rainfall are included as control. Bolivia is not included, because the sample that receives agricultural wages is not representative.
Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay- ECH (2011) and CGAXIX (2011).

[16] Appendix tables B.4 and B.5 show the point estimates, and Appendix Figure B.7 presents the relationship with total land area as measure of farm size.

Although the regional aggregates are affected by price differences across regions, the fact that we find a strong positive relationship between farm size and output per worker within regions, suggests that labor productivity differences across farms of different sizes can drive part of the positive relationships that we find between regional averages of income and farm size. The positive correlation between farm size and household income suggests that the prevalence of labor in low output, labor intensive farming can be associated with overall low income in rural areas and might explain part of the income inequalities between households in agriculture and non-agriculture. As discussed in the previous section, the persistence of this gap might result from distortions in land, labor, and capital markets, as well as other factors shaping the selection of agents across sectors (Adamopoulos et al., 2022; Adamopoulos and Restuccia, 2014; Bryan et al., 2014; Lagakos and Waugh, 2013; Restuccia et al., 2008; Young, 2013).

Figure 2.10. Relationship between interquartile range, 25th, and 75th percentile of agricultural income and farm size



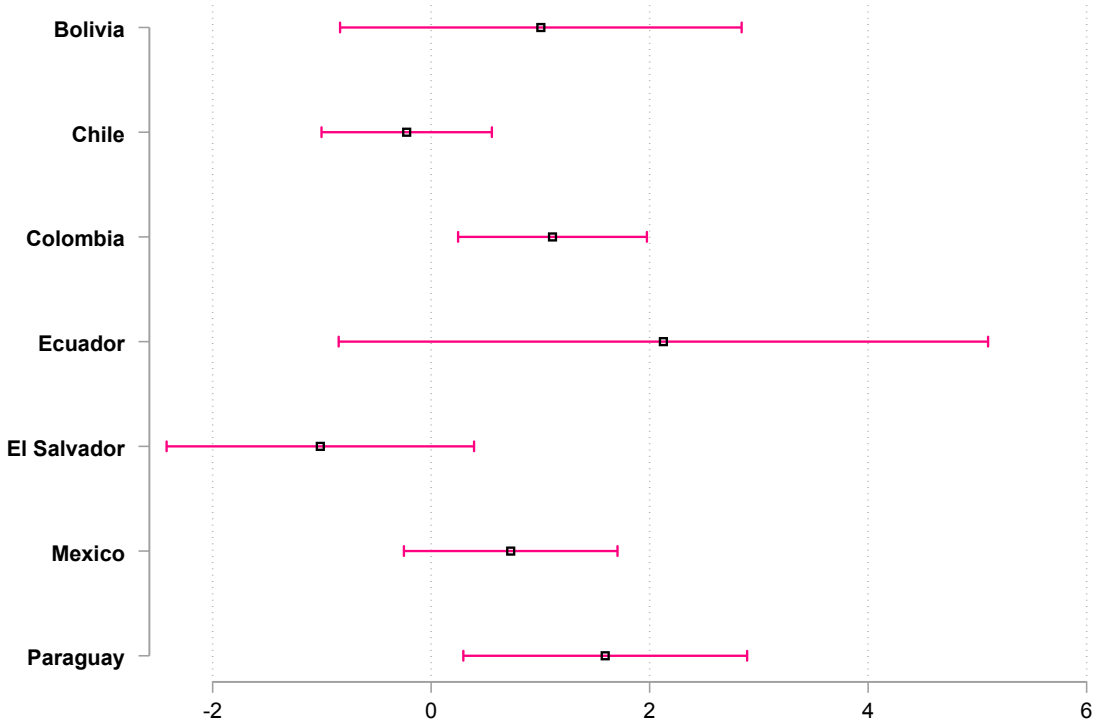
Notes: Coefficient estimates from independent linear regressions with log of IQR, the 25th and 75th percentiles of cultivated area (independent variables) and agricultural income (dependent variables) across regions. Regional averages of soil quality and rainfall included as controls. Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay- ECH (2011) and CGAXIX (2011).

2.5.3 Heterogeneity within regions and the within agricultural income inequality

Heterogeneity in agricultural labor productivity within regions can explain income inequality across agricultural households. Figure 2.11 shows that in all countries but Chile and El Salvador, there is a positive correlation between the dispersion of output per worker and the dispersion in agricultural income within regions (measured as the interquartile range). This

positive relationship is statistically significant in Colombia and Paraguay.

Figure 2.11. Relationship between interquartile of range of output per worker and interquartile of agricultural income



Notes: Coefficient estimates from independent linear regressions with the log of interquartile range of output per worker as dependent variable and the log of interquartile range of agricultural income as independent variable across regions. Regional averages of soil quality and rainfall included as controls.
Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008).

When analysing the relationship between the dispersion in farm size and income inequality we find mixed results. Figure 2.10 and Appendix Table B.6 show a positive and statistically significant relationship between the interquartile range of the distribution of agricultural income and farm size in Ecuador, Paraguay, Perú, and Uruguay, while in the rest of the countries the relationship is not statistically significant. Also, Figure B.8 documents that in most countries, there is a negative relationship between the land Gini and the income Gini. Only in Paraguay and Perú, this relationship is positive, but it is not statistically significant.

According to our theoretical framework, in the absence of agricultural labor markets, the dispersion of agricultural productivity determines both the dispersion in farm size and agricultural income inequality. The lack of correlation between these variables, along with the heterogeneity that we observe in land yields across farms might suggest that the existence of distortions that deviate farm size from its efficient allocation.

2.6 Conclusions

The findings in this chapter highlight that the dynamics in the agricultural sector explain one third of total income inequality in the nine countries we study. Income inequality within the agricultural sector, driven by large differences in agricultural productivity, largely explain this effect. Indeed, the poorest agricultural households in the nine countries are mostly small farmers, with little participation in agricultural and non-agricultural labor markets. Earnings of wealthiest households, on the other hand, are more diversified and depend more on agricultural wages.

Our results suggest that the large prevalence of small farms in most of these countries, concomitantly with frictions in labor and land markets, are associated with low agricultural incomes for small producers. Indeed, high labor intensity in small farms and low output per worker, driven by an overuse of family workers, seem to partially explain the low agricultural income of these households and, consequently, income inequality within in the agricultural sector.

Tackling rural poverty and income inequality thus requires understanding the complex relation between land size, agricultural productivity and agricultural income. The widespread idea of high productivity small farms misses an important piece of the puzzle. Output per worker, which largely determines agricultural income, is much lower in these farms. Moreover, this idea could lead to the aim of equalizing farm size within regions, which could lead to lower and more disperse income.

References

Acemoglu, D., M. A. Bautista, P. Querubín, and J. A. Robinson (2007). Economic and political inequality in development: The case of cundinamarca, colombia. Working Paper 13208, National Bureau of Economic Research.

Acemoglu, D. and J. Robinson (2012). *Why Nations Fail. The Origin of Power, Prosperity and Poverty*. Penguin Random House.

Adamopoulos, T. (2008). Land inequality and the transition to modern growth. *Review of Economic Dynamics 11*(2), 257–282.

Adamopoulos, T., L. Brandt, J. Leight, and D. Restuccia (2022). Misallocation, selection, and productivity: A quantitative analysis with panel data from china. *Econometrica 90*(3), 1261–1282.

Adamopoulos, T. and D. Restuccia (2014). The size distribution of farms and international productivity differences. *American Economic Review 104*(6), 1667–97.

Adamopoulos, T. and D. Restuccia (2020). Land reform and productivity: A quantitative analysis with micro data. *American Economic Journal: Macroeconomics 12*(3), 1–39.

AENOR (2022). Estudio cualitativo para el análisis del vínculo entre producción agrícola, migración y cambio climático en el salvador. Technical report, AENOR.

Alston, L. J. and B. Mueller (2010). Property rights, land conflict and tenancy in brazil. Technical report, National Bureau of Economic Research.

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

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). Firms and the decline in earnings inequality in brazil. *American Economic Journal: Macroeconomics 10*(1), 149–89.

Assunção, J. and J. Chiavari (2014). Land rental markets in brazil: A missed opportunity. *Climate Policy Initiative: San Francisco, CA, USA*.

Baland, J.-M. and J. A. Robinson (2008). Land and power: Theory and evidence from chile. *American Economic Review 98*(5), 1737–65.

Barrett, C. B., M. F. Bellemare, and J. Y. Hou (2010). Reconsidering conventional explanations of the inverse productivity–size relationship. *World Development 38*(1), 88–97.

Bauluz, L., Y. Govind, and F. Novokmet (2020). Global land inequality. *World Inequality Lab Working Paper 10*.

Berry, R. A., W. R. Cline, et al. (1979). *Agrarian structure and productivity in developing countries: a study prepared for the International Labour Office within the framework of the World Employment Programme*. Johns Hopkins Univ. Press.

Birdsall, N. and J. L. Londoño (1997). Asset inequality matters: an assessment of the world bank's approach to poverty reduction. *The American Economic Review 87*(2), 32–37.

Blanchet, T., J. Fournier, and T. Piketty (2022). Generalized pareto curves: Theory and applications. *Review of Income and Wealth 68*(1), 263–288.

Bouillon, C. P., A. Legovini, and N. Lustig (2003). Rising inequality in mexico: Household characteristics and regional effects. *The Journal of Development Studies 39*(4), 112–133.

Bourguignon, F., F. H. G. Ferreira, and M. Menéndez (2007). Inequality of opportunity in brazil. *Review of Income and Wealth 53*(4), 585–618.

Brockett, C. D. (1992). Measuring political violence and land inequality in central america. *American political science Review 86*(1), 169–176.

Brockett, C. D. (2019). *Land, power, and poverty: agrarian transformation and political conflict in Central America*. Routledge.

Bryan, G., S. Chowdhury, and A. M. Mobarak (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica 82*(5), 1671–1748.

Bucheli, M., N. Lustig, M. Rossi, and F. Amábile (2014). Social spending, taxes, and income redistribution in uruguay. *Public Finance Review 42*(3), 413–433.

Buonanno, P. and J. F. Vargas (2019). Inequality, crime, and the long run legacy of slavery. *Journal of Economic Behavior & Organization 159*, 539–552.

Carranza, R., M. De Rosa, and I. Flores (2023). Wealth inequality in latin america.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth 1*, 679–741.

Chauvin, J. P. and J. Messina (2020). Disparidades regionales y segregación urbana. In M. Busso and J. Messina (Eds.), *La crisis de la desigualdad: América Latina y el Caribe en la encrucijada*. Interamerican Development Bank.

Chen, C., D. Restuccia, and R. Santaeulalia-Llopis (2017). Land misallocation and productivity. Technical report, National Bureau of Economic Research.

Chen, C., D. Restuccia, and R. Santaeulàlia-Llopis (2022). The effects of land markets on resource allocation and agricultural productivity. *Review of Economic Dynamics 45*, 41–54.

Ciamarra, U. P. (2004). Access to land through rental markets: a (counter-) evolution in the world bank's land policy. *Land Reform, Land Settlement and Cooperatives. United Nations Food and Agriculture Organization. Rome*.

Cipollina, M., N. Cuffaro, and G. D'Agostino (2018). Land inequality and economic growth: a meta-analysis. *Sustainability 10*(12), 4655.

Cruces, G., C. García Domench, and L. Gasparini (2014). Inequality in education: Evidence for latin america. In G. A. Cornia (Ed.), *Falling Inequality in Latin America: Policy Changes and Lessons*. Oxford University Press.

de la Torre, A., J. Messina, and J. Silva (2017). *The Inequality Story in Latin America and the Caribbean: Searching for an Explanation*. Springer International Publishing.

Deininger, K. and P. Olinto (1999). Asset distribution, inequality, and growth. *Inequality, and Growth (November 1999)*.

Deininger, K. and L. Squire (1998). New ways of looking at old issues: inequality and growth. *Journal of development economics 57*(2), 259–287.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of development economics 84*(2), 755–776.

Elbers, C., P. Lanjouw, J. A. Mistiaen, and B. Özler (2008). Reinterpreting between-group inequality. *The Journal of Economic Inequality 6*(3), 231–245.

Engerman, S. L., S. H. Haber, and K. L. Sokoloff (2000). 11. inequality, institutions and differential paths of growth among new world economies. *Institutions, contracts and organizations*, 108.

Eslava, F. and F. Valencia (2023). Origins of latin american inequality.

Fan, S. and C. Chan-Kang (2005). Is small beautiful? farm size, productivity, and poverty in asian agriculture. *Agricultural economics 32*, 135–146.

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

Ferreira, F. H., J. Messina, J. Rigolini, L.-F. López-Calva, M. A. Lugo, and R. Vakis (2013). *Economic Mobility and the Rise of the Latin American Middle Class*. The World Bank.

Ferreira, F. H. G., S. P. Firpo, and J. Messina (2018). Ageing poorly?: Accounting for the decline in earnings inequality in brazil, 1995-2012. IDB Publications (Working Papers) 8220, Inter-American Development Bank.

Ferreira, F. H. G. and J. Gignoux (2014). The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review 28*(2), 210–246.

Ferreira, F. H. G., P. G. Leite, and J. A. Litchfield (2008). The rise and fall of brazilian inequality: 1981–2004. *Macroeconomic Dynamics 12*(S2), 199–230.

Foster, A. D. and M. R. Rosenzweig (2022). Are there too many farms in the world? labor market transaction costs, machine capacities, and optimal farm size. *Journal of Political Economy 130*(3), 636–680.

Frankema, E. (2010). The colonial roots of land inequality: geography, factor endowments, or institutions? *The Economic History Review 63*(2), 418–451.

Frankema, E. H. (2005). The colonial origins of inequality: Exploring the causes and consequences of land distribution. Technical report, IAI Discussion Papers.

Gáfaro, M. M., A. M. Ibáñez, D. Zarruk, et al. (2012). Equidad y eficiencia rural en colombia: una discusión de políticas para el acceso a la tierra.

Galán, J. S. (2018). Tied to the land? intergenerational mobility and agrarian reform in colombia.

Gasparini, L., M. Horenstein, E. Molina, and S. Olivieri (2008). Income polarization in latin america: Patterns and links with institutions and conflict. *Oxford Development Studies 36*(4), 461–484.

Gollin, D. (2019). Farm size and productivity: Lessons from recent literature. *IFAD Research Series* (34), 1–35.

Gollin, D., D. Lagakos, and M. E. Waugh (2014). The agricultural productivity gap. *The Quarterly Journal of Economics 129*(2), 939–993.

Gomez-Lobo, A. and D. Oviedo (2023). Spatial inequalities in latin america: mapping aggregate to micro-level disparities.

Grajales, R. V., O. S. Arana, and E. E. M. Campa (2018). Still looking for the land of opportunity: Regional differences in social mobility in mexico. *Sobre México. Revista de Economía 4*(1), 54–69.

Heltberg, R. (1998). Rural market imperfections and the farm size productivity relationship: Evidence from pakistan. *World Development 26*(10), 1807–1826.

Julien, J. C., B. E. Bravo Ureta, and N. E. Rada (2019). Assessing farm performance by size in malawi, tanzania, and uganda. *Food Policy 84*, 153–164.

Key, N. (2019). Farm size and productivity growth in the united states corn belt. *Food Policy 84*, 186–195.

Lagakos, D. (2020). Urban-rural gaps in the developing world: Does internal migration offer opportunities? *Journal of Economic perspectives 34*(3), 174–192.

Lagakos, D. and M. E. Waugh (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review 103*(2), 948–80.

Lamb, R. L. (2003). Inverse productivity: Land quality, labor markets, and measurement error. *Journal of Development Economics 71*(1), 71–95.

Levy, S. and N. Schady (2013). Latin america's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives 27*(2), 193–218.

Lustig, N. (2016). Inequality and fiscal redistribution in middle income countries: Brazil, chile, colombia, indonesia, mexico, peru and south africa. *Journal of Globalization and Development 7*(1), 17–60.

Lustig, N., L. F. Lopez-Calva, and E. Ortiz-Juarez (2013). Declining inequality in latin america in the 2000s: The cases of argentina, brazil, and mexico. *World Development 44*, 129–141.

Lustig, N., L.-F. Lopez-Calva, E. Ortiz-Juarez, and C. Monga (2016). *Deconstructing the Decline in Inequality in Latin America*. Palgrave Macmillan UK.

Lustig, N., C. Pessino, and J. Scott (2014). The impact of taxes and social spending on inequality and poverty in argentina, bolivia, brazil, mexico, peru, and uruguay: Introduction to the special issue. *Public Finance Review 42*(3), 287–303.

Maurizio, R. and G. Vázquez (2016). Distribution effects of the minimum wage in four latin american countries: Argentina, brazil, chile and uruguay. *International Labour Review 155*(1), 97–131.

Messina, J. and J. Silva (2019). Twenty years of wage inequality in latin america. *The World Bank Economic Review 35*(1), 117–147.

Montero, E. (2022). Cooperative property rights and development: Evidence from land reform in el salvador. *Journal of Political Economy 130*(1), 48–93.

Muñoz-Sabater, J. (2019). Era5-land monthly averaged data from 1981 to present.

Muñoz-Sabater, J. (2021). Era5-land monthly averaged data from 1950 to 1980.

Nayo, A., D. Nkoro, C. Turtoi, and O. Cara (2019). Main results and metadata by country (2006–2015). *FAO Statistical Development Series (FAO) eng no. 17*.

Neidhöfer, G., J. Serrano, and L. Gasparini (2018). Educational inequality and intergenerational mobility in latin america: A new database. *Journal of Development Economics 134*, 329–349.

North, D. C., W. Summerhill, and B. Weingast (2000). Order, disorder and economic change: Latin america vs. north america. *Governing for prosperity 19*.

Rada, N. E. and K. O. Fuglie (2019). New perspectives on farm size and productivity. *Food Policy 84*, 147–152.

Restuccia, D., D. T. Yang, and X. Zhu (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics 55*(2), 234–250.

Rodríguez-Castelán, C., L.-F. López-Calva, N. Lustig, and D. Valderrama (2016). Understanding the dynamics of labor income inequality in latin america. Technical Report 7795, World Bank.

Sen, A. K. (1962). An aspect of indian agriculture. *Economic Weekly 14*(4-6), 243–246.

Sheng, Y. and W. Chancellor (2019). Exploring the relationship between farm size and productivity: Evidence from the australian grains industry. *Food Policy 84*, 196–204.

Sokoloff, K. L. and S. L. Engerman (2000). Institutions, factor endowments, and paths of development in the new world. *Journal of Economic perspectives 14*(3), 217–232.

Young, A. (2013). Inequality, the urban-rural gap, and migration. *The Quarterly Journal of Economics 128*(4), 1727–1785.

Chapter 3

Land Inequality in the Developing World[1]

3.1 Introduction

Agricultural land remains central to the livelihoods of the world's poor, with three out of four of the poorest billion individuals working in agriculture (FAO, 2016). Despite decades of economic growth and structural transformation, the absolute number of agricultural workers worldwide has barely declined, and hundreds of millions of households continue to rely on land and land-based activities for income and subsistence (FAO, 2023). For these households, land is not only a factor of production but also a store of wealth, a source of collateral, and a determinant of occupational choice and economic security (Binswanger and Deininger, 1997; Deininger and Feder, 2001). The distribution of land has therefore long been viewed as a key determinant of development, shaping institutions, human capital accumulation, political power, and the persistence of inequality (Acemoglu et al., 2002; Engerman and Sokoloff, 1997, 2002; Galor et al., 2009). Yet, despite its central role, surprisingly little is known about who owns land today, how valuable it is, and how unequally it is distributed across and within developing countries.

A central reason for this gap is that existing cross-country evidence on land inequality relies almost exclusively on agricultural census data, which measure the distribution of operational holdings by land area (Deininger and Squire, 1998; Frankema, 2010; Lowder et al., 2021). This approach suffers from three limitations. First, censuses measure land under operation rather than land ownership, the relevant concept for understanding wealth, access to credit, and economic power (Binswanger et al., 1995; Deininger and Feder, 2001). Second, census-based measures focus on land area, implicitly treating land as a homogeneous good, even though land

[1]This chapter is based on joint work with Luis Bauluz, Yajna Govind, and Filip Novokmet.

values vary widely due to differences in soil quality, irrigation, infrastructure, and proximity to markets. Third, agricultural censuses exclude the landless population by construction, omitting a substantial share of agricultural households in many developing countries (World Bank, 2008). Together, these limitations imply that existing measures provide an incomplete, and potentially misleading, picture of land inequality.

Beyond measurement, we also lack basic evidence on the role of land in shaping household economic outcomes. Much of the growth literature treats land inequality as an initial condition, examining its long-run relationship with growth or institutional development (Deininger and Squire, 1998; Easterly, 2007; Frankema, 2010). Far less is known about how landownership and land value relate contemporaneously to household income, access to credit, and overall wealth within developing countries today.

This paper addresses these gaps by constructing a new, harmonized household-level database on land ownership, land value, income, credit, and wealth across seventeen developing countries in Africa, Asia, and Latin America.[2] Together, these countries account for around two-thirds of the developing world by population or GDP. Using nationally representative household surveys with detailed land modules, primarily LSMS and national surveys, we shift the unit of analysis from farms to households and from land area to land value, while explicitly incorporating landless agricultural households. This framework allows us to study the microeconomic role of land for household outcomes, the distribution of land wealth, and the incidence of land-based poverty across countries.

We first use this database to examine the relationship between land and household economic outcomes. We show that landownership is strongly associated with access to credit among agricultural households, but is not, by itself, systematically associated with higher income. Conditional on owning land, land value explains substantially more variation in both income and borrowing than land area. Households with more valuable land earn significantly higher incomes and have greater access to credit markets. These results highlight land value as the economically relevant measure of land endowments in rural economies, consistent with standard economic theory in which assets matter through their value, as collateral and as claims on future income streams, rather than through their physical quantity.

We then turn to the measurement of land inequality. We introduce a ladder of increasingly comprehensive inequality concepts, moving from land area among landowners, to land value among landowners, and finally to land value including the landless population. This approach yields three main findings. First, land-area inequality measured in household surveys is remarkably similar to land-area inequality measured in agricultural censuses. This close correspondence is reassuring for both data sources: despite conceptual and practical differences in how land is recorded, the two approaches yield broadly consistent measures of

[2]While land and income data are available for most countries, information on loans and wealth is only available for eleven and four countries, respectively.

land-area inequality, suggesting that they capture the same underlying distribution. Second, once land value and the landless are incorporated, land inequality is at least as concentrated as overall wealth. Under this benchmark measure, the Gini coefficient of land wealth is 0.77, similar in magnitude to reported wealth inequality in developing countries (see, e.g., Chancel et al. 2021; UBS 2024). Land is therefore not a special, more equitably distributed asset, but a central component of wealth concentration.

Third, accounting for land value also alters cross-country patterns of inequality. Differences in land-area inequality across countries are large (Deininger and Squire, 1998; Frankema, 2010), but they shrink considerably when inequality is measured in terms of land value rather than area: the coefficient of variation falls from around 0.21 for land area to around 0.14 for land value. This greater similarity across countries is related to a new stylized fact that we document: a robust inverse relationship between the size of landholdings and the value per hectare, which is only partly explained by the well-known inverse productivity–size relationship (e.g., Berry and Cline, 1979; Sen, 1962; Vollrath, 2007). Overall, countries that appear highly unequal in land area tend to look less unequal in land value, while the opposite holds for countries with more egalitarian land-area distributions.

Finally, we use our land value data to study land-based asset poverty among agricultural households. Motivated by theories of poverty traps and evidence emphasizing minimum asset thresholds (Balboni et al., 2022; Banerjee and Newman, 1993; Galor and Zeira, 1993), we complement consumption-based poverty measures with a measure of land-asset poverty. We assess the share of agricultural households whose landholdings fall below a benchmark threshold of US$1,000 in PPP terms, comparable to the scale studied in the most recent asset-transfers literature (Banerjee et al., 2021; Barker et al., 2024; Blattman et al., 2020). We document substantial cross-country heterogeneity: fewer than 15 percent of agricultural households fall below this threshold in China and Vietnam, compared with more than 60 percent in Malawi and Ethiopia.

Expressing the implied transfers as a share of aggregate land wealth provides orders of magnitude for the extent of land deprivation. Importantly, the total land value required to close these gaps is often modest relative to existing land wealth: in many countries, raising all agricultural households above the threshold would require less than 10 percent of aggregate land value, and in several cases well under 1 percent. These magnitudes add nuance to long-standing debate over land concentration and agricultural productivity. A large literature emphasizes that misallocation and fragmentation can depress agricultural productivity and that consolidation may raise output (e.g., Restuccia et al. 2008; Adamopoulos and Restuccia 2014). Our accounting exercise does not evaluate the efficiency effects of redistribution. However, the fact that lifting all agricultural households above a minimum land-asset threshold often requires only a small fraction of aggregate land wealth reframes the equity–efficiency tradeoff in quantitative terms: improving the asset position of the poorest households need not

mechanically entail large aggregate reallocations of land wealth.

Taken together, our results show that land inequality should be understood and measured as a form of wealth inequality, and that land-based deprivation is often large in incidence but modest in aggregate magnitude. Land area is an inadequate metric, and even adjustments for productivity differences are insufficient, because land value captures not only current productive capacity but also location, infrastructure, and expected future income streams. By combining household-level evidence on land, income, credit, and wealth with new measures of land inequality and land-based asset poverty, this paper offers a unified framework for assessing the economic role of land in developing countries, with direct relevance for research and policy debates on agriculture, inequality, and poverty.

Literature Review

This paper contributes to three main strands of the literature. First, it is closely related in purpose to Banerjee and Duflo (2007), who document the economic lives of the poor using detailed household surveys across developing countries. While their work emphasizes consumption, occupations, and small asset holdings, we focus on the distribution of the single most important productive asset in rural economies, land, and investigate its role in shaping the economic outcomes of both poor and non-poor agricultural households. By comparing landless and landowning households, we also uncover important heterogeneities within and across the two groups.

Second, the paper contributes to the measurement of land inequality. A large literature studies farm size, productivity, and structural transformation, often emphasizing the relationship between operational scale and agricultural productivity (e.g., Adamopoulos and Restuccia, 2014; Berry and Cline, 1979; Lowder et al., 2016; Vollrath, 2007). A related strand examines the long-run relationship between land inequality, income inequality, and economic growth using agricultural census data (Deininger and Squire, 1998; Easterly, 2007; Frankema, 2010). We build on this work by focusing on household-level data, on value instead of area, on owners instead of operators, and by showing that accounting for land value and landlessness substantially alters both the level and the cross-country ranking of land inequality.

Third, our analysis relates to the growing literature on asset endowments and poverty traps. A series of randomized controlled trials evaluates the effects of one-time transfers of productive assets on poor households (Balboni et al., 2022; Banerjee et al., 2021; Barker et al., 2024; Blattman et al., 2020), emphasizing minimum asset thresholds for escaping persistent poverty (Banerjee and Newman, 1993; Galor and Zeira, 1993). Our contribution is to provide household-level evidence on the incidence of very low land endowments and on the implied magnitudes needed to raise agricultural households above specified asset thresholds. We view our measure of land-based asset poverty as complementary to the widely used consumption-based poverty measures developed by the World Bank and used by the UN to monitor poverty worldwide. While consumption captures realized living standards,

asset poverty reflects the lack of productive endowments that may limit households' ability to generate sustained income growth.

The remainder of the paper proceeds as follows. Section 3.2 describes the data and concepts. Section 3.3 examines the relationship between landownership, land value, and household economic outcomes. Section 3.4 presents our new measures of land inequality and documents cross-country patterns. Section 3.5 studies land-based asset poverty and discusses its implications. Section 3.6 concludes.

3.2 Concepts and Data

This section introduces the conceptual framework and data used to build a harmonized micro-level database of agricultural households across developing countries. The database serves two related purposes. First, it allows us to study the role of land ownership in shaping household economic welfare, including income and access to credit, as analyzed in Section 3.3. Second, it provides the basis for constructing comprehensive measures of land inequality that account for land values and explicitly incorporate landless agricultural households, which are the focus of Section 3.4.

3.2.1 Data Sources and Database Construction

Our primary data source consists of nationally representative household surveys with detailed modules on land ownership, agricultural activity, income, and borrowing. The final sample, reported in Table 3.1, covers seventeen developing countries across Africa, Asia, and Latin America and includes the largest developing economies, accounting for roughly two-thirds of the developing world by population or GDP.

Given that only a few countries have more than one survey wave with land value data, our analysis is cross-sectional. For each country, we use the most recent survey wave available that has data on both land area and land value. This focus on the cross-section is also justified by the fact that land inequality is highly persistent over time. Consistent with the existing literature documenting persistence in land inequality over long horizons (e.g., Deininger and Squire 1998), we also find strong stability across survey waves in the subset of countries for which more than one wave is available. We also exploit the Demographic and Health Surveys (DHS), which contain standardized questions on land ownership and plot size across many countries. The Gini coefficient for land-area inequality across surveys and waves is shown in Appendix Figure C.2.

We harmonize land, income, loan, and wealth variables across surveys using consistent definitions and aggregation rules. Table 3.1 reports the list of survey waves for each country included in the analysis, and an indication of the availability of the main variables in each survey. All estimates are weighted using survey sampling weights to ensure national representativeness.

Table 3.1. Database Summary: Survey Sources and Key Characteristics

Country	Survey type	Year	Obs.	Agric. HH share (%)	Share landless (%)	Variable availability				
						Area	Value	Loans	Wealth	Income
South Asia										
India	National	2012	110,800	48.9	43.3	✓	✓		✓	
Bangladesh	National	2011	6,503	68.9	32.4	✓	✓	✓		✓
Pakistan	National	2010	16,345	38.7	32.3	✓	✓	✓		✓
Indonesia	National	2014	15,921	35.9	27.9	✓	✓	✓		✓
China–Vietnam										
China	National	2012	13,218	63.4	5.6	✓	✓			✓
Vietnam	National	2014	9,399	64.2	10.9	✓	✓	✓		✓
Latin America										
Ecuador	National	2014	28,970	28.2	35.7	✓	✓	✓		✓
Guatemala	LSMS	2000	7,276	47.5	44.6	✓	✓	✓		✓
Brazil	LSMS	1997	4,940	17.9	43.9	✓	✓	✓		✓
Panama	LSMS	2008	7,045	33.7	27.7	✓	✓			✓
Nicaragua	National	2014	6,851	29.6	39.7	✓	✓			✓
Sub-Saharan Africa										
Ethiopia	LSMS	2018	6,894	63.1	35.3	✓	✓	✓	✓	✓
Gambia	National	2015	13,282	36.5	32.7	✓	✓	✓		✓
Malawi	LSMS	2016	12,447	78.5	25.4	✓	✓	✓		✓
Subsample 2016	LSMS	2016	2,508	81.3	19.2	✓	✓		✓	
Niger	LSMS	2014	3,617	82.1	15.4	✓	✓			✓
Nigeria	LSMS	2015	4,697	54.3	22.1	✓	✓			✓
Tanzania	LSMS	2014	3,352	70.0	20.6	✓	✓	✓		✓
Subsample 2019	LSMS	2019	1,184	93.6	25.5	✓	✓		✓	
Total / Average			**275,249**	**48.0**	**34.5**					

Notes: Agric. HH share is the weighted share of agricultural households among all households. *Share landless* is the weighted share of landless households among agricultural households. Checkmarks indicate availability of the corresponding variable. No data cleaning is applied: no winsorization, no dropping of zero or missing values. Malawi subsample refers to the IHPS panel; Tanzania subsample refers to the NPS-SDD panel. Both are used for wealth analysis.

3.2.2 Definitions and Measurement

Our unit of analysis is the *agricultural household*, defined to include both landowning and landless households. Any household that owns agricultural land is classified as a landowner. A household is classified as landless if it does not own agricultural land but reports at least one member employed in agriculture. This definition captures the full population whose livelihoods are directly linked to agriculture and provides an appropriate denominator for analyzing inequality in agrarian economies. In robustness checks, we consider more restrictive definitions of landlessness that only include households whose head reports agriculture as the primary activity or, when this information is unavailable, households in which more than half of the members work in agriculture.

The main challenge in measuring land ownership inequality in a cross-country analysis is the definition of ownership in environments with heterogeneous property-rights institutions. For comparability, we rely on two main questions to determine land tenure: i) whether the household declares owning land when the question is directly asked in the survey, and in the absence of this question, ii) whether a household reports having acquired a plot of land through transfer of rights, either purchase or inheritance. Following the development literature,

our classification proxies for a "bundle of rights" approach to land ownership (Schlager and Ostrom 1992; Deininger 2003). Our classification would include the rights to use the land for production, to control production decisions, and to transfer the land within the household. It potentially captures both formally titled land and long-term, socially or legally recognized use rights that are de facto exclusive and heritable, as in China and Vietnam, as well as customary tenure systems prevalent in parts of Sub-Saharan Africa. By contrast, communal land that cannot be exclusively controlled or transferred by the household is excluded from our baseline ownership definition[3]. Appendix Table C.9 summarizes the ownership questions and definitions used in each country.

Consistent with this definition, our analysis focuses on land privately owned by households and excludes land owned by corporations or held under purely communal regimes. This restriction mirrors standard practice in the measurement of household wealth (Alvaredo et al. 2021), which includes only assets over which households hold ownership rights and excludes assets that are merely used without control or transfer rights (for example, public or employer-provided housing). The focus on household-owned land is also justified by the central role of households in global agriculture. Family farms are estimated to operate roughly 75% of the world's agricultural land (Lowder et al. 2016).

For landowning households, we measure agricultural land along two dimensions. Land area is the sum of the areas of all agricultural plots owned by the household. Land value is measured using households' self-reported assessments of the market value of each plot or of total landholdings, depending on survey design. Measuring land in value terms is essential because land is heterogeneous: differences in soil quality, irrigation, location, and market access are reflected in value but not in area. Self-reported land values are standard in household surveys and are analogous to the measurement of other assets, such as housing, in wealth surveys (Davies et al. 2011; Davies et al. 2017).

We also measure total household income. Total income is the sum of agricultural income (own-farm net income, farm wages, and land rents), non-agricultural income (business profits and wages), and non-labour income (pensions, remittances, transfers, and other unearned income). Components are aggregated from the detailed modules of each household survey following a harmonized protocol described in the Appendix (Section C.5.2). Household credit is measured using self-reported outstanding loan balances.

3.2.3 Inequality Measurement Concepts

Cross-country studies of land inequality have traditionally relied on agricultural censuses, which report the distribution of *operational holdings* by size (e.g., Deininger and Squire 1998; Vollrath 2007; Frankema 2010). An operational holding is defined by the FAO as an economic

[3]For example, this definition excludes land "granted by village leaders", a common land tenure set-up in Sub-Saharan countries.

unit of agricultural production under a single management, irrespective of legal ownership. This approach has been foundational and offers two important advantages: it generally covers agricultural land across ownership types, irrespective of the legal owner (with the exception of communal land, which is not considered an operational holding), and it measures land area directly and with high accuracy.[4] However, it provides an incomplete measure of inequality in land as an economic asset.

First, census data measure land *operated* rather than land *owned*. Because ownership and operation need not coincide, especially where land is leased or managed by tenants, census-based measures conflate the two and may understate the concentration of land ownership when a single owner controls multiple operational holdings.

Second, census-based measures focus exclusively on land area, implicitly treating land as a homogeneous good. This overlooks large differences in land value arising from soil quality, irrigation, location, and market access, and therefore understates inequality in land wealth even among landowners.

Third, agricultural censuses exclude landless households by construction. In many developing countries, a substantial share of the agricultural population depends on agriculture for its livelihood without owning land. Omitting these households mechanically biases inequality measures downward and fails to account for cross-country heterogeneity in the prevalence of landlessness.

Our approach addresses these limitations by shifting from censuses to household surveys, measuring land both by area and value, and explicitly incorporating landless agricultural households into the inequality distribution. It provides, for the first time, a consistent basis for re-evaluating patterns of land inequality across the developing world using three distinct inequality measures: land area, land value, and land value among all agricultural households, including the landless.

3.2.4 Data Validation

The shift from censuses to surveys raises a natural question: do household surveys provide reliable measures of land inequality? We address this concern by benchmarking our baseline survey-based estimates against two independent sources: agricultural censuses and the Demographic and Health Surveys (DHS), which include standardized land modules across many countries but do not collect land value data.

Land-area inequality from household surveys aligns very closely with census-based measures ($R^2 = 0.94$ across Gini coefficients). This high correlation has two implications.

[4] Census data typically cover not only household farms but also government land and land operated by private corporations, although sector coverage varies across countries and over time (Lowder et al. 2016). For example, in many African countries, agricultural censuses cover primarily land operated by the household sector, excluding corporate land.

First, it suggests that the distribution of operational holdings provides a reasonable first-order proxy for the distribution of land area among landowners. This lends support to the large literature that has relied on census data to study land-area inequality. Second, it provides reassurance that household surveys yield a reliable picture of the distribution of land held by households. DHS-based measures of land-area inequality closely match our benchmark survey estimates across countries, providing further reassurance that our results are not driven by the particular choice of surveys in our sample. See Appendix C.1.1 for supporting figures.

However, the correspondence weakens substantially when we relate census-based area inequality to survey-based value inequality ($R^2 = 0.56$), indicating that land value introduces fundamentally new information beyond what area-based measures capture. Even within the same household surveys, the correlation between land-area and land-value inequality is only moderate ($R^2 = 0.63$). We explore the implications of these differences in the sections that follow. Section 3.3 examines whether land area or land value better predicts household economic outcomes, while Section 3.4 reconsiders land inequality using land value and incorporating landless households.

3.3 Agricultural households: land, income, and credit

This section examines how land ownership shapes households' economic outcomes in developing countries. We document that *land value*, not land area, is the key economic dimension of land endowment. To do so, we harmonize household surveys across developing countries, in the spirit of Banerjee and Duflo (2007), covering both agricultural and non-agricultural households and focusing on land as the central productive asset in rural economies. We estimate pooled OLS regressions with country-survey fixed effects, exploiting within-country variation across households. Our main specification is as follows:

$$Y_{ij} = \alpha + \beta_1 LogLandArea_{ij} + \beta_2 LogLandValue_{ij} + \lambda_j + \epsilon_{ij} \tag{3.1}$$

where Y_{ij} are three sets of outcomes of household i in country j: household income (in logs), wealth (using four surveys with wealth modules), and credit (both the extensive and the intensive margins). Column (1) of Table 3.2 compares agricultural to non-agricultural households using the full population. Column (2) restricts the analysis to agricultural households. The remaining columns further restrict our sample to agricultural landowners, and regress Y_{ij} on log land area and log land value, controlling for survey-year fixed effects. The standard errors are clustered at the survey-strata level. We test several ways of accounting for zeros in Supplementary Appendix C.2.2 and report country-specific results in Supplementary Appendix C.2.3.

Table 3.2 reports the regression results. For context, column (1) documents that agricultural

households earn substantially less income than non-agricultural households (a coefficient of −0.820 implying a 56 percent lower income on average), consistent with lower labor productivity in agriculture, particularly in developing countries (Restuccia et al. 2008).

Table 3.2. Income, Wealth, and Loans Regressions

	Income				Wealth	Loans	
	ln(Income)				asinh(Wealth)	Any loans	ln(Loans)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Agric. HH	-0.820***	–	–	–	–	–	–
	(0.067)						
Landowner	–	0.057	–	–	–	–	–
		(0.067)					
Log Land Area	–	–	0.153***	-0.059**	-0.014	0.002	0.079***
			(0.028)	(0.026)	(0.030)	(0.004)	(0.028)
Log Land Value	–	–	–	0.445***	0.622***	-0.001	0.221***
				(0.034)	(0.050)	(0.004)	(0.030)
Survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	471	447	441	441	175	349	295
Surveys	16	16	16	16	4	11	11
Sample	Full pop	Agric	Owners	Owners	Owners	Owners	Borrowers
Observations	130,541	67,409	50,031	50,031	34,404	39,267	8,299
Adj. R^2	0.427	0.380	0.384	0.479	0.595	0.116	0.673
Within R^2	0.0481	0.0001	0.0137	0.1661	0.3044	0.0000	0.0701
Mean dep. var.	6.999	6.238	6.089	6.089	11.227	0.211	5.941

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions include survey fixed effects and probability weights. Standard errors (in parentheses) are clustered at the survey–strata level. Columns (1)–(4) pool 16 income surveys; columns (6)–(7) use the income surveys with reported loan data; column (5) uses the 4 wealth surveys (Ethiopia2018, Malawi IHPS 2016, Tanzania2019, India2012). Income is household income excluding transfers, winsorized at the 99th percentile, equivalized per capita.

Within agriculture, a striking pattern emerges: land ownership alone is not associated with a significantly higher income (column 2). Quantile regressions reported in Supplementary Appendix C.2.1 reveal that the relationship between land ownership and income is highly non-monotonic: land ownership is negatively and significantly associated with income at the bottom of the distribution, close to zero at the median, and positively associated at the top. Hence, landowning households are more heterogeneous in income terms than the landless, comprising a disproportionate share of both the poorest and the richest agricultural households.[5] Taken together, this heterogeneity underscores an important insight: landlessness is not synonymous to poverty or vulnerability.[6]

Turning to the main results, on the intensive margin (among landowners), columns (3) and (4) report interesting patterns on the relationship between income, land area and land value. There is a positive relationship between land area and income with an elasticity of

[5]The fact that many poor households own land, often in very small plots, is emphasized in micro studies of poverty (Banerjee and Duflo 2007; Rapsomanikis 2015).

[6]Supplementary Appendix C.2.8 shows that landless households are more diversified, engaging more frequently in multiple occupations and off-farm activities; this result is robust to alternative definitions of landlessness.

0.153 (column 3). However, once land value is included in the model, alongside land area, the pattern is quite different. Land value is strongly positively associated with income (elasticity of 0.445), while the land area coefficient reverses sign and becomes negative. Hence, conditional on land value, larger land holdings are associated with lower income, suggesting that extensive but low-value land does not enhance income.

This result highlights that the relevant attribute of land is its value, which likely reflects productivity, location, soil quality, irrigation, and market access. Land value is a substantially more powerful predictor of income than land size, and its relative importance holds despite land values being self-reported and potentially noisier than land area.

This finding generalizes beyond income. Column (5) shows that land value is also strongly associated with household wealth (elasticity of 0.622), notably larger than the income coefficient of 0.445, while land area contributes little additional explanatory power, mirroring the income results.[7] Here again, land value emerges as the economically salient dimension of land endowments, capturing both current productive assets and the capacity to generate future wealth as discussed below.

Access to credit (columns 6 and 7) provide additional evidence of the economic role of land, as it is a key constraint on investment and income growth in rural economies, particularly where land serves as collateral (e.g., Besley 1995; Carter and Olinto 2003; Deininger and Feder 2009). Column (6) shows that neither land area nor land value significantly predicts whether a household borrows. However, conditional on borrowing (Column 7), both land area and land value significantly predict loan size, with land value having a substantially larger effect than land area (elasticity of 0.221 *vs.* 0.079 respectively). The coefficient for land value is 2.8 times larger, suggesting that land value weighs more heavily than quantity when determining collateral-based borrowing capacity. This result validates the mechanisms suggested by the wealth findings: land value predicts wealth partly through enabling collateral-based credit for productive investment.

Three main findings emerge. First, agricultural households are substantially poorer than non-agricultural households on average. Second, land ownership heterogeneity within agriculture is quite substantial: many poor households own land, but of low value. Third, and most importantly, land *value*, not land area, is the dimension of land that consistently predicts income, wealth, and access to credit. These results have direct implications for measuring land inequality: if land value is the dimension that maps into income generation and borrowing capacity, then inequality in land value is the relevant object for understanding economic inequality in rural economies.

The reported results are robust to multiple specification checks. Appendix C.2.2 shows that results hold under alternative income transformations (log(1+y), inverse hyperbolic sine) and

[7] Note that column (5) is estimated on four surveys that have wealth modules (Ethiopia 2018, Malawi IHPS 2016, Tanzania 2019, India 2012).

consumption measures, are not driven by cross-country composition effects (Appendix C.2.3), and remain robust to including non-titled land, including transfers in income and varying winsorization thresholds (Appendix C.2.4). Importantly, high-value landowners are more likely to have missing income reports, implying our estimates of the land value coefficient are, if anything, conservative (Appendix C.2.5). Results are similarly robust across alternative poverty-line indicators (Appendix C.2.6) and wealth definitions (Appendix C.2.7).

3.4 Land Wealth Inequality

The micro evidence in Section 3.3 shows that land value, rather than land area, is the dimension of land most closely associated with household income and access to credit. We now turn to the distribution of land itself. Our objective is twofold: first, to document new stylized facts on land inequality across developing countries using harmonized micro data; and second, to explain why cross-country differences in land inequality depend crucially on how land is measured.

3.4.1 From Land Area to Land Wealth

We construct three increasingly comprehensive measures of land inequality: (i) land area among landowners, (ii) land value among landowners, and (iii) land value including landless agricultural households. These measures form a ladder from a purely physical concept of land to a comprehensive measure of land wealth and access.

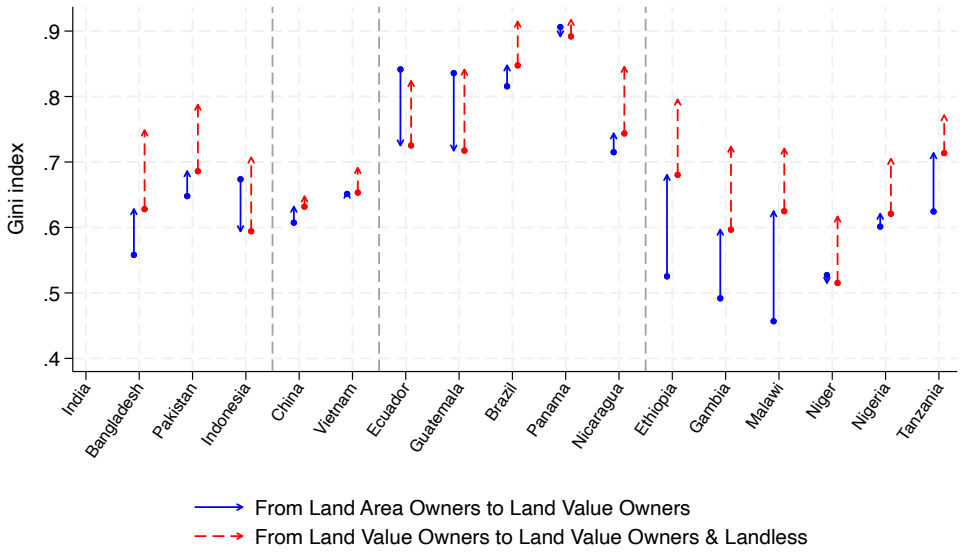
Figure 3.1, Panel A reports country-level Gini indices under each concept. The blue segments show the change in inequality when moving from land area to land value among owners, while the red segments show the additional effect of including landless households. Two patterns emerge.

First, inequality in land wealth is relatively high. Averaging across countries (Panel B), the Gini index is 0.65 for land area among owners, 0.68 for land value among owners, and 0.77 when land value is measured including the landless. Thus, once the landless are incorporated, land wealth inequality reaches levels comparable in magnitude to overall wealth inequality documented in developing countries in global wealth reports.[8] By comparison, income inequality in the same surveys averages 0.58. Agricultural land is therefore not a relatively egalitarian asset class; rather, it is at least as concentrated as other forms of wealth.

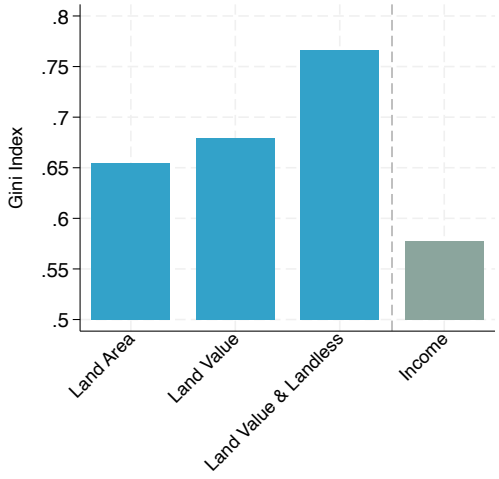
Second, the transition from area to value varies substantially across countries. In several South Asian and Sub-Saharan African countries, including Bangladesh, Pakistan, India, Ethiopia, and Malawi, the Gini index is higher for land value than for land area, indicating that price dispersion amplifies inequality. By contrast, in Latin American countries such as

[8] See, for example, the World Inequality Report 2022 (Chancel et al. 2021) and the UBS Global Wealth Report 2024 (UBS 2024).

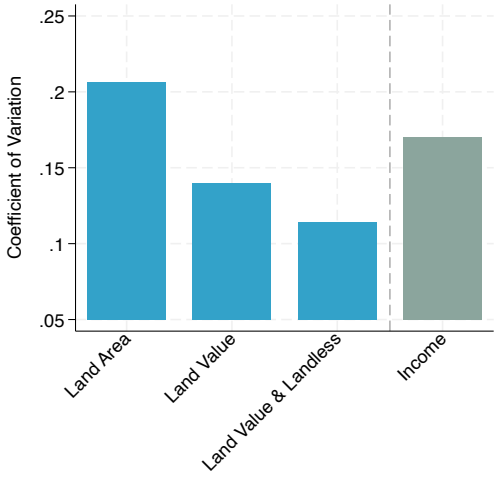
Figure 3.1. Land Inequality Across Measures



(a) Gini index across measurement concepts by country



(b) Cross-country mean Gini by concept



(c) Cross-country Coefficient of Variation by concept

Notes: Panel (a) reports country-level Gini coefficients for land inequality under a sequence of measurement concepts. The blue segment shows the change in the Gini when moving from land *area* among landowners to land *value* among landowners. The red segment shows the change when moving from land value among landowners to land value including landless households (i.e., assigning zero land value to the landless). Panels (b) and (c) summarize cross-country dispersion in Gini coefficients by concept: panel (b) plots the unweighted cross-country mean (each country receives equal weight) and panel (c) plots the coefficient of variation.

Ecuador, Panama, and Brazil, as well as in China, Vietnam or Indonesia, the shift from area to value is small or even negative. In these settings, large land-area holdings tend to command lower prices per hectare, consistent with extensive holdings being located on lower-value or more remote land.

The inclusion of landless households mechanically raises measured inequality in all

countries, but the magnitude of the increase varies considerably. The effect is largest in countries with high landlessness rates, for example in Brazil, Pakistan, and Ethiopia, where roughly 30 to 40 percent of agricultural households own no land (Appendix Figure C.3), and smallest in China and Vietnam, where landlessness is rare. This heterogeneity underscores that analyses restricted to landowners overlook a fundamental source of inequality: the binary divide between households that own land and those that do not.

Beyond levels, the dispersion of inequality across countries also changes markedly depending on the concept used. Panel C reports the cross-country coefficient of variation (CoV) of the Gini index. For land area among owners, the CoV equals 0.21, indicating substantial heterogeneity. When inequality is measured by land value among owners, the CoV declines to 0.14. Including landless households reduces it further to 0.12, a total decline of more than 40 percent relative to area-based measures. The relatively small additional reduction when incorporating the landless suggests that variation in landlessness is not strongly correlated with initial land inequality. By comparison, the CoV of income inequality across the same countries is 0.17, lying between the dispersion of land-area inequality and that of land-value inequality among owners. Countries are therefore more similar in land wealth inequality than in land-area inequality, and in fact more similar than in income inequality.

These results nuance the usual stories that focus on strong differences across regions in how concentrated land ownership is. When measured by area, countries dominated by large, extensive estates often appear markedly more unequal than those characterized by smallholder agriculture. Once land is measured by value and landless households are included, these differences narrow substantially. The relevant object for understanding economic inequality is not the physical distribution of hectares, but the distribution of land wealth.

3.4.2 The Inverse Value-Size Relationship

Why do cross-country differences shrink when land is measured by value rather than area? The answer lies in a systematic relationship between landholding size and land prices.

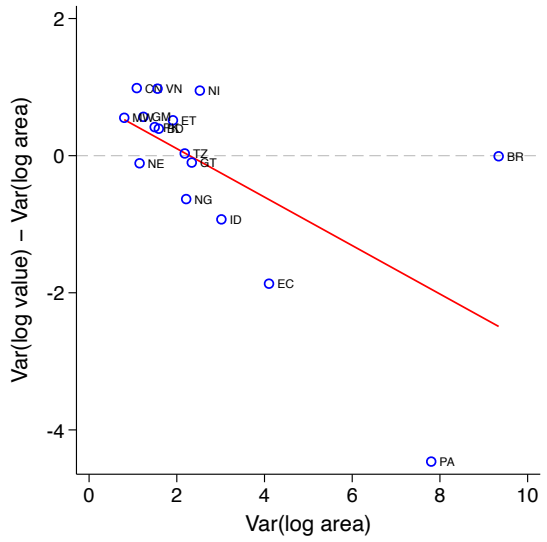
Figure 3.2, Panel A plots the change in measured inequality, defined as $\mathrm{Var}(\log V) - \mathrm{Var}(\log A)$, against initial land-area inequality. Countries with highly unequal land-area distributions experience the largest declines in inequality when switching to land value, while countries with relatively equal land-area distributions often experience small increases. The relationship is strongly negative, indicating that the countries appearing most unequal by area tend to appear less unequal once land is measured by value.

To interpret this pattern, we rely on the identity

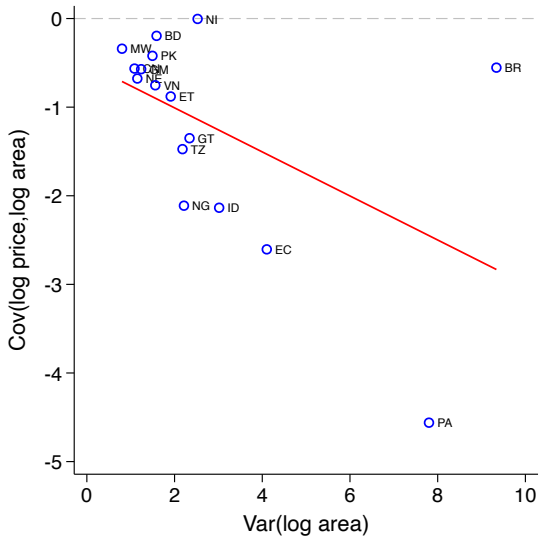
$$\mathrm{Var}(\log V) - \mathrm{Var}(\log A) = \mathrm{Var}(\log P) + 2\,\mathrm{Cov}(\log A, \log P),$$

where $V = A \times P$ denotes land value, A land area, and P the price per hectare. The change in

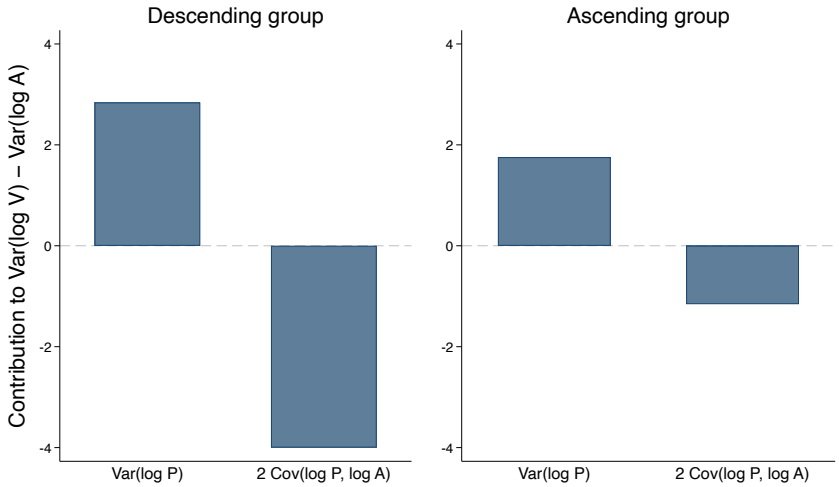
Figure 3.2. The inverse value–size relationship and land inequality



(a) Area Inequality and Changes Area to Value



(b) The covariance between land size and prices per hectare



(c) Decomposing changes in inequality

Notes: Panel (a) plots the change in measured land inequality when moving from area-based to value-based measures, defined as $\mathrm{Var}(\log V) - \mathrm{Var}(\log A)$, against land-area inequality $\mathrm{Var}(\log A)$, where V denotes land value and A land area. Panel (b) documents the inverse value–size relationship by plotting the covariance between land prices and land size $\mathrm{Cov}(\log P, \log A)$ against $\mathrm{Var}(\log A)$, where P is the price per hectare. Panel (c) groups countries according to whether $\mathrm{Var}(\log V) - \mathrm{Var}(\log A)$ is negative ("descending") or positive ("ascending") and decomposes the change using the identity $\mathrm{Var}(\log V) - \mathrm{Var}(\log A) = \mathrm{Var}(\log P) + 2\,\mathrm{Cov}(\log P, \log A)$; within each group, bars report the average contribution of $\mathrm{Var}(\log P)$ and $2\,\mathrm{Cov}(\log P, \log A)$.

inequality when moving from area to value is therefore driven by two components: dispersion in land prices and the covariance between land size and land prices.

Figure 3.2, Panel B shows that this covariance is negative in all countries and becomes more negative as land-area inequality rises. The negative covariance reflects the inverse value–size relationship: large landholdings tend to command lower prices per hectare. Crucially, the

covariance can be written as $\text{Cov}(\log P, \log A) = \beta \text{Var}(\log A)$ where β denotes the elasticity of land prices with respect to land size. The equalizing effect of the inverse value–size relationship therefore depends both on the steepness of the price gradient (the elasticity) and on the dispersion of landholdings over which this gradient operates.

While the development literature has long emphasized an inverse relationship between farm size and productivity (e.g., Vollrath, 2007), we document that a similar inverse relationship holds for land prices. Our contribution is to show that this price gradient is spread across countries and that it plays a central role in shaping the mapping from land distribution to land wealth inequality. Controlling for observable measures of land productivity reduces the magnitude of the elasticity only modestly (Appendix Table C.8), indicating that the inverse value–size relationship is not solely driven by productivity differences.

Panel C decomposes this relationship by grouping countries according to whether measured inequality rises or falls when moving from area to value. We define the "descending group" as countries where $\text{Var}(\log V) - \text{Var}(\log A) < 0$ (inequality falls) and the "ascending group" as countries where $\text{Var}(\log V) - \text{Var}(\log A) > 0$ (inequality rises). In the descending group, the negative covariance term is large enough to more than offset price dispersion, leading to a decline in inequality when moving from area to value. In the ascending group, the covariance term is smaller in absolute value and price dispersion dominates.

To quantify the relative roles of the elasticity and initial land-area inequality in generating these differences, Appendix C.3.1 implements a Shapley decomposition of the covariance term. The results show that the primary driver of the more negative covariance in the descending group is their substantially higher dispersion of landholdings, with a secondary contribution from slightly more negative elasticities. In other words, the inverse value–size relationship is widespread, but its inequality-compressing effect is magnified in countries with pronounced upper tails in landholdings.

To shed light on the cross-country differences in land-area dispersion, we draw on two complementary sources of evidence. At the household level, Appendix C.3.3 shows that crop-intensive households have significantly less dispersed farm sizes than animal-intensive households, and that the dispersion effect operates specifically through the upper tail of the distribution, consistent with pastoral and ranching operations inflating $\text{Var}(\log A)$ in countries with more extensive land use. At the country level, Appendix C.3.2 exploits FAO data on agricultural land composition. Countries with a lower share of arable land in total agricultural land exhibit significantly higher land-area inequality and a more negative covariance between land size and prices, while arable share is not systematically related to the price elasticity. Although these correlations are descriptive, they align closely with our decomposition results: cross-country differences in the covariance between land size and prices appear to be associated more strongly with differences in land-area dispersion than with systematic variation in the price elasticity. This suggests that land-use structure shapes the dispersion of landholdings,

particularly the prevalence of large, extensive estates, while the inverse value–size gradient itself remains a broadly shared feature across countries.

Taken together, our findings present land inequality as a wealth inequality phenomenon. Standard area-based measures overstate inequality in countries characterized by large, extensive holdings and understate it in countries with smaller but higher-value plots. Once land is measured by value and landless households are incorporated, agricultural land emerges as an asset class, at least as concentrated as other forms of wealth, with less cross-country heterogeneity than area-based measures would suggest. The inverse value–size relationship provides a unifying economic mechanism linking the distribution of land area, land prices, and the resulting structure of land wealth inequality.

3.5 Land Value, Asset Poverty, and Minimum Endowments

While land value inequality is relatively more similar across countries than land area inequality, this does not imply that land-based poverty is similar. Inequality measures summarize relative dispersion, but they do not reveal how many households lack a minimally sufficient land endowment. In this section, we move from relative inequality to land-based asset poverty, using economically meaningful thresholds expressed in land value rather than land area.

A large literature in development economics emphasizes poverty traps, whereby households remain poor because they lack the minimum asset endowment required to access higher-return activities (Banerjee and Newman 1993; Galor and Zeira 1993). Consistent with these theories, randomized evaluations of "big-push" programs that combine one-time productive asset transfers with complementary support document sizable gains in income, earnings, and wealth in the short to medium run, with long-run persistence in some settings and attenuation in others (Barker et al. 2024). Such programs have been widely adopted, reaching more than 20 million households across 75 countries by 2020 (Andrews et al. 2021).

In rural economies, land is the dominant form of household wealth and the primary productive asset for the poor. Evidence from countries with detailed wealth data shows that land accounts for the largest share of rural household assets (e.g., Bharti 2018), and is tightly linked to overall household wealth. This motivates a focus on land value as a natural metric of asset poverty among agricultural households. Land value captures a fundamental stock of productive wealth for agricultural households and shapes occupational choice, access to credit, and risk smoothing. Figure C.12 in the Appendix decomposes household wealth and shows that plot value accounts for a substantial share of total assets, and an even larger share of productive assets excluding the dwelling, across all four wealth surveys. This confirms that land value serves as a meaningful proxy for productive wealth among agricultural households and motivates our focus on land-based asset poverty thresholds.

Motivated by these findings, we study the distribution of land wealth relative to a minimum

productive-asset threshold. Unlike commonly used indicators of land fragmentation, such as farm-size cutoffs at two hectares, which the literature considers as rather arbitrary (Lowder et al. 2016; Lowder et al. 2025), our thresholds are not based on land area per se, but are economically motivated and grounded in evidence from poverty-trap models and asset-transfer experiments. Conceptually, land-based asset poverty differs from consumption poverty: whereas consumption captures realized living standards, asset poverty proxies the minimum productive endowment required to access higher-return activities and escape persistent low-income states.

We use a benchmark of US$1,000 (PPP) per household, which aligns with the scale of one-off asset transfers studied in "big-push" programs. Across settings, these programs typically provide households with productive asset bundles valued between roughly US$500 and US$1,400 in PPP terms (e.g., Balboni et al. 2022; Banerjee et al. 2021; Blattman et al. 2020; Barker et al. 2024).[9] Our baseline definition of households below the threshold includes all agricultural households, both landowners and the landless, with landless households assigned a land value of zero. In the Appendix, we show results for more restrictive definitions of landlessness and alternative thresholds, including US$500 and US$1,500 (PPP).

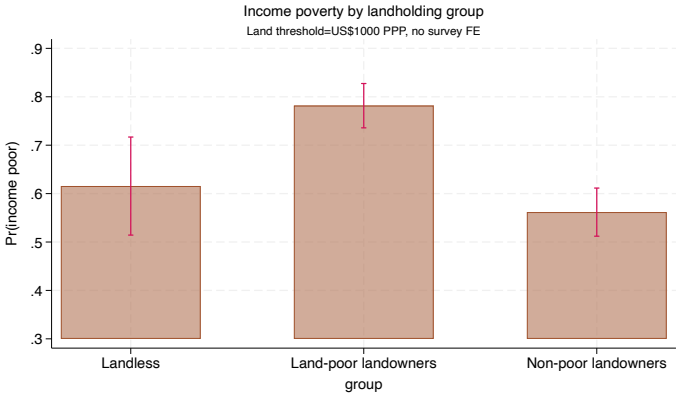
A natural validation of our land-based asset-poverty measure is whether it is systematically related to household income poverty. If land value captures a minimum productive endowment, then households below this threshold should face substantially higher income poverty. Panel (a) of Figure 3.3 examines this relationship within agricultural households by comparing predicted income-poverty rates across three mutually exclusive landholding groups: landless households, land-poor landowners (with land value below US$1,000), and non-poor landowners (above the threshold). Income poverty is measured using the World Bank's US$2.15 (PPP) per-day line.

The results reveal a clear non-monotonic pattern. Land-poor landowners are the most economically vulnerable group, with roughly 72 percent living below the income-poverty line, compared with about 51 percent of landless households and 48 percent of non-poor landowners. Thus, owning a small amount of low-value land is associated with higher income poverty than landlessness itself. This pattern is robust to the inclusion of survey fixed effects. These findings complement the household-level evidence in Section 3.3, which shows that landowners are overrepresented at the bottom and top of the income distribution, compared to landless households. Together, the results underscore that land-based deprivation in rural economies is more closely related to the prevalence of landholdings that fall below a minimum productive scale than to landlessness.

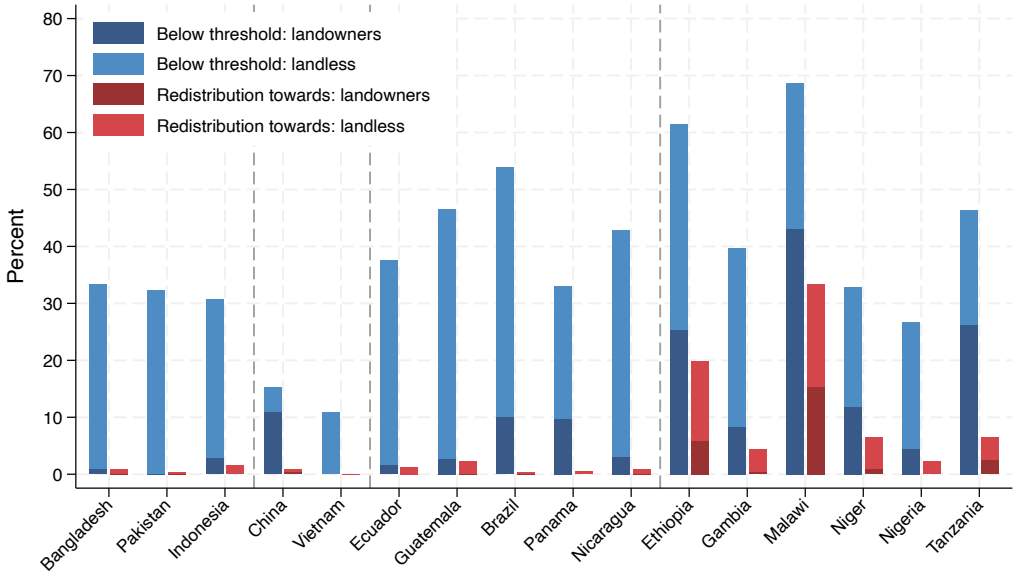
Using this benchmark, we quantify (i) the share of agricultural households whose land

[9]Direct asset transfers in these studies are of roughly comparable magnitude: about US$1,200 (PPP 2018) per household in India (Banerjee et al., 2021), US$1,371 (PPP 2021) per household in Ethiopia (Barker et al., 2024), roughly US$490 (PPP 2007) per household in Bangladesh (Balboni et al., 2022), and US$400 per treated individual in Uganda (Blattman et al., 2020), implying household-level capital transfers that can reach similar magnitudes when multiple household members are treated.

Figure 3.3. Asset Poverty and Potential Redistribution



(a) Income Poverty by Landholding Group



(b) Asset Poverty and Potential Redistribution, by Country

Notes: Panel (a) reports predicted income-poverty rates (US$2.15 PPP per day) by landholding group among agricultural households, based on an OLS regression of the poverty indicator on mutually exclusive indicators for land-poor landowners (below US$1,000 PPP) and non-poor landowners (at or above the threshold), with landless households as the omitted category. Estimates use survey weights and standard errors clustered at the survey level; bars display point estimates and 95 percent confidence intervals. Results are very similar when survey fixed effects are included (Appendix Figure C.13). Panel (b) reports, for the same threshold, (i) the share of agricultural households below the land-value benchmark and (ii) the total land value required to raise all households to the threshold as a share of aggregate land value ("Redistribution"). Bars are stacked by household type; landless households are assigned zero land value.

wealth falls below the threshold, and (ii) the total value of land that would be required to close these gaps, expressed as a fraction of aggregate land value – which we call the redistribution share. It is important to emphasize that this redistribution exercise is not a policy proposal. Redistributing land is complex and raises issues of feasibility, political economy, and efficiency.

Our objective is instead descriptive: to provide orders of magnitude for the land value shortfalls faced by households at the bottom of the distribution, and to benchmark these gaps relative to aggregate land wealth.

Panel (a) of Figure 3.3 summarizes these results at the US$1,000 threshold. It reports country-level estimates of (i) the share of agricultural households below the threshold and (ii) the corresponding land-value shortfall required to raise all households to the benchmark, expressed as a fraction of aggregate land value.

Two broad patterns stand out. First, the incidence of land-based asset poverty varies enormously across countries. At the US$1,000 threshold, the share of agricultural households below the benchmark ranges from below 15 percent in China and Vietnam to above 60 percent in Malawi and Ethiopia. These differences are large even relative to the cross-country variation in land value inequality documented in the previous section, underscoring that inequality measures alone provide an incomplete picture of land-based deprivation. Countries with similar levels of inequality can differ sharply in the mass of households clustered near the bottom of the land-value distribution. Second, both the incidence and the aggregate depth of asset poverty are largely influenced by landless households. Across countries and regions, landlessness accounts for the majority of households below the threshold as well as for a large share of the total land-value shortfall relative to the benchmark. When looking only at landowners, the largest share of land-poor people is concentrated in Sub-Saharan Africa, consistent with this region also accounting for two-thirds of the global income poor (World Bank 2024).

At the same time, and perhaps more strikingly, the aggregate land value required to close these gaps is often modest relative to total land wealth. In many countries, raising all agricultural households to the US$1,000 threshold would require less than 5 percent of aggregate land value, and in several cases well under 1 percent. Even in low-income agrarian economies such as Ethiopia and Malawi, where asset poverty is widespread, the implied redistribution shares remain well below half of aggregate land value. This reflects the highly skewed distribution of land value documented earlier: large mass deprivation at the bottom can coexist with limited aggregate resource requirements.

Taken together, these results highlight a sharp distinction between the prevalence of asset poverty and its aggregate cost. In some countries, a large fraction of households fall below economically meaningful land thresholds, yet the total land value required to close these gaps is small relative to existing land wealth. In others, asset poverty is both widespread and more resource-intensive to address, reflecting deeper structural constraints rather than extreme concentration alone. This distinction is obscured by inequality statistics, but emerges clearly once land holdings are evaluated relative to minimum productive endowments.

More broadly, our approach offers a complementary lens on rural poverty that parallels the logic of consumption-based poverty measurement. Just as consumption poverty rates

summarize how many individuals fall below a subsistence benchmark, land-based asset poverty rates quantify how many households lack the minimum productive wealth needed to access higher-return activities. By anchoring thresholds in evidence from asset-transfer programs, our approach speaks directly to debates about the scale and feasibility of asset-based interventions in rural economies.

These magnitudes also provide perspective on the long-standing debate over land concentration and agricultural productivity. A large literature emphasizes that misallocation and fragmentation can depress agricultural productivity and that consolidation may raise output (e.g., Restuccia et al. 2008; Adamopoulos and Restuccia 2014). Our accounting exercise does not evaluate the efficiency effects of redistribution. However, the fact that lifting all agricultural households above a minimum land-asset threshold often requires only a small fraction of aggregate land wealth suggests that land-based deprivation can coexist with highly concentrated wealth at the top. In value terms, the aggregate resource gap facing the rural poor is frequently modest relative to existing land wealth. This reframes the equity–efficiency tradeoff in quantitative terms: improving the asset position of the poorest households need not mechanically entail large aggregate reallocations of land wealth.

3.6 Conclusion

This paper constructs a harmonized household-level database covering seventeen developing countries to study the distribution of agricultural land as a form of wealth. Three main findings emerge. First, land value, not land area, is the economically relevant dimension of land endowments: it consistently predicts household income, wealth, and access to credit, while land area adds little explanatory power once value is accounted for. Second, when land inequality is measured by value and landless households are incorporated, it reaches levels comparable to overall wealth inequality, with a cross-country Gini of 0.77. Cross-country differences in land inequality shrink substantially under value-based measurement, driven by a robust inverse relationship between landholding size and price per hectare that compresses land wealth inequality in countries with highly unequal land-area distributions. Third, land-based asset poverty is widespread across developing countries, yet the aggregate land value required to bring all agricultural households above a minimum productive threshold is often modest relative to existing land wealth, suggesting that the equity costs of land deprivation need not entail large aggregate reallocations.

These findings have implications for both measurement and policy. For measurement, our results show that area-based indicators of land inequality, while useful, are insufficient for capturing the economic dimension of land concentration. Value-based measures, inclusive of landless households, provide a more complete picture of land as a component of wealth inequality. For policy, the evidence underscores that landless agricultural households are a highly heterogeneous group that warrants particular attention: landlessness is not synonymous

with poverty, but land-poor landowners, those holding land below a minimum productive scale, face the highest rates of income poverty. Future research could extend this framework to study the dynamics of land value inequality over time, exploit within-country variation to identify causal channels linking land endowments to household welfare, and investigate the role of land markets and tenure reform in shaping the distribution of land wealth.

References

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly Journal of Economics 117*(4), 1231–1294.

Adamopoulos, T. and D. Restuccia (2014). The size distribution of farms and international productivity differences. *American Economic Review 104*(6), 1667–1697.

Alvaredo, F., A. B. Atkinson, L. Bauluz, T. Blanchet, L. Chancel, M. Fisher-Post, I. Flores, B. Garbinti, J. Goupille-Lebret, C. Martínez-Toledano, M. Morgan, T. Neef, T. Piketty, A.-S. Robilliard, E. Saez, L. Yang, and G. Zucman (2021). Distributional national accounts (dina) guidelines: Methods and concepts used in the world inequality database. *WID Working Paper 2016/2*.

Andrews, C., A. de Montesquiou, I. Arevalo Sánchez, P. V. Dutta, B. V. Paul, S. Samaranayake, J. Heisey, T. Clay, and S. Chaudhary (2021). *The State of Economic Inclusion Report 2021: The Potential to Scale*. Washington, DC: World Bank.

Balboni, C., O. Bandiera, R. Burgess, M. Ghatak, and A. Heil (2022). Why do people stay poor? *The Quarterly Journal of Economics 137*(2), 785–844.

Banerjee, A., E. Duflo, and G. Sharma (2021). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights 3*(4), 471–486.

Banerjee, A. V. and E. Duflo (2007). The economic lives of the poor. *Journal of economic perspectives 21*(1), 141–167.

Banerjee, A. V. and A. F. Newman (1993). Occupational choice and the process of development. *Journal of Political Economy 101*(2), 274–298.

Barker, N., D. Karlan, C. Udry, and K. Wright (2024). The fading treatment effects of a multifaceted asset-transfer program in ethiopia. *American Economic Review: Insights 6*(2), 277–294.

Berry, A. and W. Cline (1979). *Agrarian Structure and Productivity in Developing Countriess*. John Hopkins University Press.

Besley, T. (1995). Property rights and investment incentives: Theory and evidence from ghana. *Journal of Political Economy 103*(5), 903–937.

Bharti, N. K. (2018). Wealth inequality, class and caste in india, 1951-2012. *PSE Work Pap 2018*, 1–71.

Binswanger, H. P. and K. Deininger (1997). Explaining agricultural and agrarian policies in developing countries. *Journal of economic literature 35*(4), 1958–2005.

Binswanger, H. P., K. Deininger, and G. Feder (1995). Power, distortions, revolt and reform in agricultural land relations. In J. Behrman and T. Srinivasan (Eds.), *Handbook of Development Economics*, Volume 3, pp. 2659–2772. Elsevier.

Blattman, C., N. Fiala, and S. Martinez (2020). The long-term impacts of grants on poverty: Nine-year evidence from uganda's youth opportunities program. *American Economic Review: Insights 2*(3), 287–304.

Carter, M. R. and P. Olinto (2003). Getting institutions "right" for whom? credit constraints and the impact of property rights on the quantity and composition of investment. *American Journal of Agricultural Economics 85*(1), 173–186.

Chancel, L., T. Piketty, E. Saez, and G. Zucman (Eds.) (2021). *World Inequality Report 2022*. Paris, France. Report on global income and wealth inequality based on the World Inequality Database (WID.world).

Davies, J. B., R. Lluberas, and A. F. Shorrocks (2017). Estimating the level and distribution of global wealth, 2000–2014. *Review of Income and Wealth 63*(4), 731–759.

Davies, J. B., S. Sandström, A. Shorrocks, and E. N. Wolff (2011). The level and distribution of global household wealth. *The Economic Journal 121*(551), 223–254.

Deininger, K. (2003). *Land policies for growth and poverty reduction*. World Bank Publications.

Deininger, K. and G. Feder (2001). Land institutions and land markets. In B. L. Gardner and G. C. Rausser (Eds.), *Handbook of Agricultural Economics*, Volume 1, pp. 288–331. Elsevier.

Deininger, K. and G. Feder (2009). Land registration, governance, and development: Evidence and implications for policy. *The World Bank Research Observer 24*(2), 233–266.

Deininger, K. and L. Squire (1998). New ways of looking at old issues: inequality and growth. *Journal of development economics 57*(2), 259–287.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of development economics 84*(2), 755–776.

Engerman, S. and K. Sokoloff (1997). Factor endowments, institutions and differential paths of growth among the new world economies," in stephen haber, ed., how latin america fell behind, stanford: Stanford university press.

Engerman, S. L. and K. L. Sokoloff (2002). Factor endowments, inequality, and paths of development among new world economies. *Economía 3*(1), 41–109.

FAO (2016). *The State of Food and Agriculture: Climate Change, Agriculture and Food Security*. Rome, Italy: Food and Agriculture Organization of the United Nations.

FAO (2023). *World Food and Agriculture – Statistical Yearbook 2023*. Rome: FAO.

Frankema, E. (2010). The colonial roots of land inequality: geography, factor endowments, or institutions? *The Economic History Review 63*(2), 418–451.

Galor, O., O. Moav, and D. Vollrath (2009). Inequality in landownership, the emergence of human-capital promoting institutions, and the great divergence. *The Review of economic studies 76*(1), 143–179.

Galor, O. and J. Zeira (1993). Income distribution and macroeconomics. *The review of economic studies 60*(1), 35–52.

Lowder, S., J. Skoet, and T. Raney (2016). The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Development 87*, 16–29.

Lowder, S. K., G. Bhalla, and B. Davis (2025). Decreasing farm sizes and the viability of smallholder farmers: Implications for resilient and inclusive rural transformation. *Global Food Security 45*, 100854.

Lowder, S. K., M. V. Sánchez, and R. Bertini (2021). Which farms feed the world and has farmland become more concentrated? *World Development 142*, 105455.

Rapsomanikis, G. (2015). *The Economic Lives of Smallholder Farmers: An Analysis Based on Household Data from Nine Countries*. Rome: Food and Agriculture Organization of the United Nations.

Restuccia, D., D. Yang, and X. Zhu (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics 55*(2), 234–250.

Schlager, E. and E. Ostrom (1992). Property-rights regimes and natural resources: A conceptual analysis. *Land Economics 68*(3), 249–262.

Sen, A. K. (1962). An aspect of indian agriculture.

UBS (2024). Global wealth report 2024. Report, UBS Group AG. Accessed 2026-02-18.

Vollrath, D. (2007). Land distribution and international agricultural productivity. *American Journal of Agricultural Economics 89*(1), 202–216.

World Bank (2008). *World Development Report 2008: Agriculture for Development*. Washington, DC: World Bank.

World Bank (2024). *Poverty, Prosperity, and Planet Report 2024: Pathways Out of the Polycrisis*. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO.

Appendix A

Appendix to Chapter 1

A.1 Supplemental Tables & Figures

Table A.1. Coffee Production by Department, Selected Years (sacks of 62.5 kg)

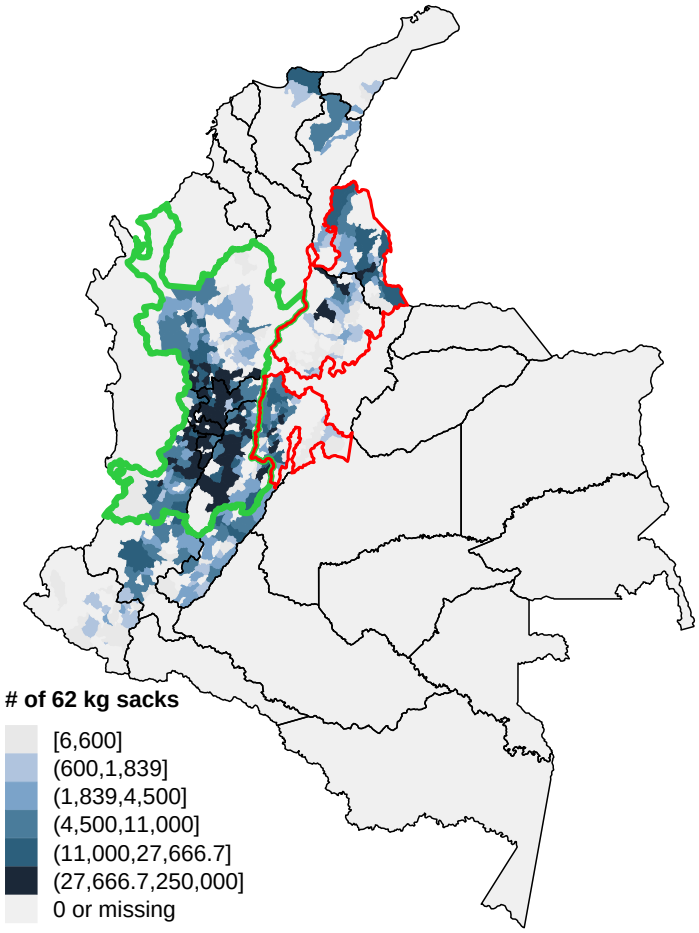
	1874	1888	1913	1924[a]	1932	1944	1955[b]
Old Frontier							
Cundinamarca	8,000	38,000	200,000	331,876	405,500	528,530	414,274
Norte de Santander	90,000	120,000	223,000	179,630	270,000	332,800	152,144
Santander	10,000	60,000	105,000	92,773	150,400	126,350	124,542
New Frontier							
Antioquia	1,200	5,200	185,000	493,672	617,500	833,022	864,915
Caldas	1,300	800	199,000	703,805	1,003,700	2,099,705	1,700,835
Tolima	1,000	2,000	60,000	214,215	448,400	901,970	813,760
Valle del Cauca	600	4,000	50,000	190,272	354,400	1,036,547	726,210
Others							
Rest	2,100	—	63,000	89,453	171,800	307,570[c]	511,998[c]
Total	114,200	230,000	1,085,000	2,295,696	3,421,700	6,166,494	5,308,678

Notes: This table reports raw coffee production values underlying the production shares plotted in Figure 1.2a. Production is measured in sacks of 62.5 kg, the standard Colombian unit. The shares in the main text figure are calculated as each department's (or frontier group's) production divided by the national total in each year.

[a]Production figures for 1924 are not directly available. I estimate them by applying the 1932 productivity rate (production per tree) to the 1924 tree count reported by (Monsalve, 1927): $\text{prod}_{d,1924} = \text{trees}_{d,1924} \times (\text{prod}_{d,1932}/\text{trees}_{d,1932})$. [b]The 1955 data from (United Nations. Economic Commission for Latin America, 1958) report production in metric tons; I convert to sacks using the factor $\times 907/62.5$. [c]The "Rest" aggregate excludes departments with missing data for that year.

Author's calculations based on (Arango, 1977; Beyer, 1947; Ocampo, 2011; Palacios, 2002). 1874 and 1913 are calculated from (Beyer, 1947) (p.388) and (Arango, 1977) (Cuadro no. II.8). 1888 comes from (Palacios, 2002) (p.21) and (Ocampo, 2011) (p.279–280 and p.369). For 1888, (Ocampo, 2011) (p.369) notes 40k for Cundinamarca and Tolima, while (Palacios, 2002) notes 38k for Cundinamarca, thus I arrive at 2k for Tolima. The same logic applies for Antioquia and Caldas as the source for the Caldas 800 value for 1888. (Palacios, 2002) has 181k for both Norte de Santander and Santander, which tracks with (Ocampo, 2011) values for both distinct departments which I take. Data for 1924 onward are based on national coffee censuses: (Monsalve, 1927) for 1924, (Federación Nacional de Cafeteros de Colombia, 1933) for 1932, (Federación Nacional de Cafeteros de Colombia, 1945) for 1944, and (United Nations. Economic Commission for Latin America, 1958) for 1955.

Figure A.1. Coffee Production Intensity by Municipality, 1944



Note: Shading indicates municipal coffee production in 1944, measured in sacks of 62.5 kg. Green borders delineate the new frontier (Antioquia, Caldas, Tolima, Valle del Cauca); red borders delineate the old frontier (Cundinamarca, Santander, Norte de Santander). Municipalities with zero production or missing data are shown in light grey.
Source: Author's calculations based on (Federación Nacional de Cafeteros de Colombia, 1945).

Table A.2. Bandit Groups by Department, 1962

	Bandit Groups	Share (%)
Old Frontier	**28**	**16.4**
Cundinamarca	16	9.4
Santander	12	7.0
New Frontier	**124**	**72.5**
Antioquia	34	19.9
Caldas	33	19.3
Tolima	27	15.8
Valle del Cauca	30	17.5
Others	**19**	**11.1**

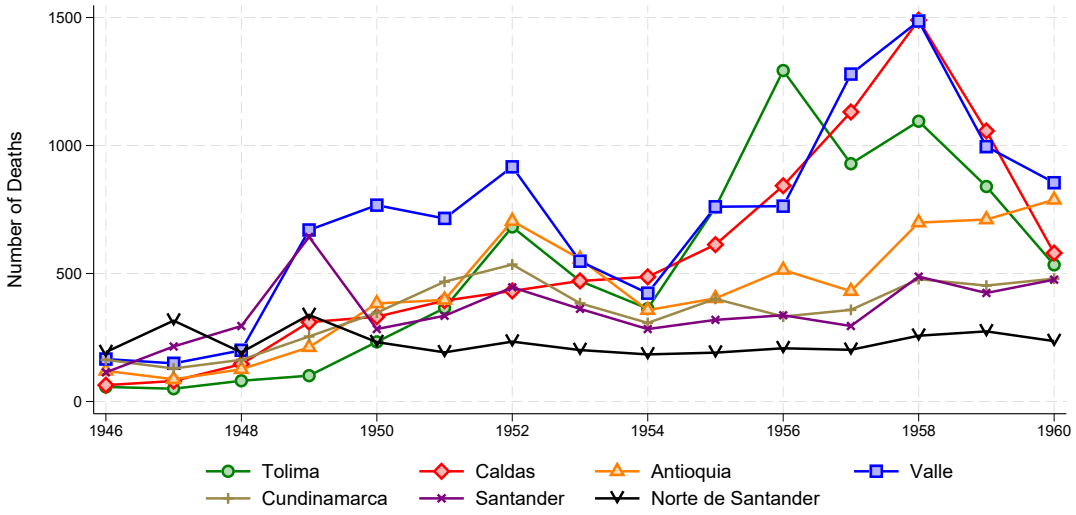
Note: Total bandit groups = 171. All identified groups are concentrated in the coffee-producing departments. "Others" comprises Cauca and Huila, the next-largest coffee producers after the old and new frontier departments.
Source: Author's calculations based on (Policía Nacional de Colombia, 1965).

Table A.3. International Intentional Death Rates per 100,000 (Top 5 Countries)

Rank	Country	Deaths per 100,000	Year
1	Colombia	34.0	1960
2	Mexico	31.1	1958
3	Nicaragua	22.8	1959
4	South Africa	21.2	1959
5	Burma	10.8	1959

Note: Colombia's rate of 34.0 per 100,000 in 1960 was the highest in the world, driven by the late-period violence of *La Violencia*.
Source: (United Nations, 1961), item BE 50, pp. 398–471.

Figure A.2. Intentional Deaths by Department, 1946–1960



Note: This figure disaggregates the frontier-level trends in Figure 1.4 by individual department. New frontier departments: Tolima, Caldas, Antioquia, Valle del Cauca. Old frontier departments: Cundinamarca, Santander, Norte de Santander.
Source: Author's calculations based on (Gutiérrez Tovar, 1961).

Table A.4. Coffee × Distance to Early Violence Hub: OLS and PPML

	(1) OLS	(2) PPML
Any coffee	2.1970***	0.4126
	(0.5416)	(0.2898)
Distance to hub (km)	-0.0141**	-0.0639**
	(0.0065)	(0.0259)
Coffee × Distance	-0.0220***	0.0374
	(0.0049)	(0.0264)
Dept. FE	Y	Y
Competing theories	Y	Y
State controls	Y	Y
Geo. controls	Y	Y
Pop. controls	Y	Y
N	741	590

Notes: DV: violent deaths 1962. Col. (1): OLS. Col. (2): PPML. Distance = km to nearest early violence hub (1949–53). HC1 robust SEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5. Role of Bandits: Coffee × Armed Group Presence

	(1) OLS	(2) PPML
Coffee (10k sacks)	0.004	0.027
	(0.206)	(0.082)
Neighbor has leader	1.438**	1.416***
	(0.636)	(0.485)
Has leader	3.398***	2.118***
	(1.297)	(0.457)
Coffee × Neighbor	0.268	-0.082
	(0.236)	(0.096)
Coffee × Has leader	2.416***	-0.004
	(0.784)	(0.082)
Dept. FE	Y	Y
Competing theories	Y	Y
State controls	Y	Y
Geo. controls	Y	Y
Pop. controls	Y	Y
N	741	590

Notes: DV: violent deaths 1962. Coffee in 10,000 sacks (mean among producers ≈ 13,000). Leader3: 0 = no leader and no neighbor with leader, 1 = neighbor has leader but municipality does not, 2 = municipality has leader. Col. (1): OLS. Col. (2): PPML (ppmlhdfe), 155 separated/singleton obs dropped. HC1 robust SEs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.1.1 Coffee Revenue per Farm: Construction and Sources

This section documents the construction of the coffee revenue per farm estimates reported in the main text (Figure 1.8) and in Table A.6 below.

Objective. The goal is to measure how much nominal revenue a single coffee farm generated relative to the prevailing agricultural wage, for each frontier group and for four benchmark years: 1924, 1932, 1944, and 1955. The formula is:

$$\text{Revenue per farm}_t = \frac{\text{Total production (sacks)}_t \times \text{Price per sack}_t}{\text{Number of coffee farms}_t}$$

and the ratio plotted in Figure 1.8 is this quantity divided by the standard annual agricultural wage.

Production data. Department-level coffee production is drawn from successive national coffee censuses. The 1932 census reports production directly in sacks of 62.5 kg (Federación Nacional de Cafeteros de Colombia, 1933). The 1944 figure comes from (Federación Nacional de Cafeteros de Colombia, 1945). The 1955 census (Departamento Administrativo Nacional de Estadística and Charry Lara, 1962) reports production in metric tons; I convert to 62.5 kg sacks using sacks = tons $\times$ (907.185/62.5).

No direct production data exists for 1924. I estimate 1924 production by multiplying the 1924 tree count reported in (Monsalve, 1927) by the department-specific 1932 trees-to-production ratio:

$$\text{Production}_{d,1924} = \text{Trees}_{d,1924} \times \frac{\text{Production}_{d,1932}}{\text{Trees}_{d,1932}}$$

This assumes that per-tree yields remained approximately stable within each department over the eight years between 1924 and 1932. The assumption is reasonable given that coffee is a perennial crop with relatively stable yields once trees reach maturity (typically 4–5 years after planting).

Farm counts. The number of coffee farms at the department level is available for 1924 (originally digitized from (Monsalve, 1927)), 1932 (Federación Nacional de Cafeteros de Colombia, 1933), and 1955 (Departamento Administrativo Nacional de Estadística and Charry Lara, 1962). No farm census was conducted in 1944.

Price data. Coffee prices are expressed in current Colombian pesos per 62.5 kg sack. The price is a five-year moving average derived from (Urrutia-Montoya et al., 2000) and converted to the units reported in the coffee censuses (62.5 kg sacks). For 1905–1925, prices are taken

from p. 31, Table 5, column 3 of the same work (external coffee price in pesos multiplied by 60%, the ratio of price kept by Colombian producers). For 1926 onward, prices follow Table 3.3, column 1. The benchmark prices used are:

	1924	1932	1944	1955
Price per sack (pesos)	15.35	12.91	25.39	188.05

Wage benchmark. The denominator of the ratio is the standard annual agricultural wage, computed as the daily agricultural wage times 305 workdays per year, following the convention in (Palacios, 2002, p. 116). For 1924 and 1932, daily wages are based on figures from (Palacios, 2002, p. 260) for a typical hacienda worker in Antioquia. Values for 1944 and 1955 are taken from (Berry, 1972) for nationally representative agricultural wage averages. The benchmark wages are:

	1924	1932	1944	1955
Annual agricultural wage (pesos)	329	247	286	1,055

The 1944 bounds. Because no farm census exists for 1944, revenue per farm in that year cannot be computed directly. Instead, I compute an upper and a lower bound:

- *Upper bound*: divides 1944 revenue by the 1932 farm count (fewer farms ⇒ higher revenue per farm).
- *Lower bound*: divides 1944 revenue by the 1955 farm count (more farms ⇒ lower revenue per farm).

The value plotted in Figure 1.8 and reported in Table A.6 is the midpoint of these two bounds. The full range is shown in the "1944 Range" column of the table.

Aggregation. For the frontier-level figure, department-level production, farms, and revenue are summed within each frontier group (New Frontier: Antioquia, Caldas, Tolima, Valle; Old Frontier: Cundinamarca, Santander, Norte de Santander; Others: all remaining departments). The ratio is then computed at the frontier level.

Note: All values in current pesos. "1944 Range" reflects the observed range between two data points (1944–32 and 1944–55). # of Farms refers to the number of farms recorded in coffee censuses.

Source: Author's calculations based on archival coffee census production and farm data and secondary sources for prices and standard agricultural wages. See text for explanation.

Table A.6. Coffee Revenue per Farm and Number of Farms by Department, Selected Years

	Coffee Revenue per Farm				**# of Coffee Farms**		
	1924	1932	1944 Range	1955	1924	1932	1955
Old Frontier							
Cundinamarca	1,808	379	[451, 972]	2,619	2,817	13,812	29,749
Norte de Santander	820	437	[1,361, 1,060]	4,609	3,361	7,972	6,207
Santander	1,085	638	[280, 1,054]	2,045	1,313	3,045	11,454
New Frontier							
Antioquia	849	279	[637, 740]	4,899	8,923	28,589	33,203
Caldas	953	323	[1,161, 1,327]	6,964	11,336	40,174	45,930
Tolima	876	453	[842, 1,793]	5,627	3,755	12,771	27,197
Valle del Cauca	978	228	[1,098, 1,311]	5,696	2,987	20,069	23,975
Agricultural Wage	329	247	286	1,055			

A.1.2 Coffee's Share of Agricultural Value

Table A.7 presents the full value of Colombia's agricultural wealth by department in 1955, using estimates from (United Nations. Economic Commission for Latin America, 1958). This source provides detailed estimates of the "Gross Farm Value of Coffee and of the Principal Commodities"—denominated in U.S. dollars—for each department, broken down by crop type. The "Other products" column encompasses all major export crops (sugarcane, tobacco, cocoa, agave), all staple foods (maize, rice, beans, cassava, plantains, vegetables), and all livestock and associated products—in other words, the entirety of Colombia's non-coffee agricultural economy.

Table A.7. Coffee vs. Other Agricultural Products by Colombian Department (1955)

Department	Coffee	Other products	Total	Coffee share
Old Frontier				
Cundinamarca	24,699	17,160	41,859	0.59
Norte de Santander	9,060	1,774	10,834	0.84
Santander	7,416	6,075	13,491	0.55
New Frontier				
Antioquia	51,505	16,694	68,199	0.76
Caldas	101,282	13,432	114,714	0.88
Tolima	48,458	7,167	55,625	0.87
Valle del Cauca	43,245	11,806	55,051	0.79
Others				
Huila	13,393	3,963	17,356	0.77
Cauca	8,179	5,849	14,028	0.58
Magdalena	4,289	790	5,079	0.84
Boyacá	2,383	10,183	12,566	0.19
Nariño	2,246	1,857	4,103	0.55

Notes: All values in thousands of U.S. dollars. "Coffee share" = Coffee/(Coffee + Other products). "Other products" includes all major export crops (sugarcane, tobacco, cocoa, agave), all staple foods (maize, rice, beans, cassava, plantains, vegetables), and all livestock and associated products.
Source: Author's calculations based on the appendix of (United Nations. Economic Commission for Latin America, 1958).

A.1.3 Historical Evidence: Bandit Groups and the Coffee Economy

The following passages from the Colombian historiography document how bandit groups operated in coffee-producing regions during *La Violencia*. I organize them by theme to illustrate the different channels through which armed groups interacted with the coffee economy.

Theft of harvests and the seasonal dynamics of violence. The most direct channel was the theft of coffee itself, timed to coincide with harvest periods.

> "The appearance of these gangs coincides with the coffee harvests. Coffee pickers and dryers are murdered, and the harvest is stolen.
>
> Many landowners were forced to enlist the "services" of these bandits to protect their estates; others did so to settle personal scores or to acquire property at a fraction of its value.
>
> The stewards or estate managers maintain a climate of unrest and encourage violence to keep the owners away, as their presence would deprive the stewards of their substantial profits. The rivalry and ambition among the stewards to replace those in more favorable positions have led to hundreds of deaths." (Policía Nacional, 1959) (p.29)

> "The coffee crop had to be commercialized in an armed fashion in violent areas or it might be stolen several times on the way to market. This violence did not represent national economic losses, but it did imply changes in the appropriation of the economic benefits of the crop. The coffee always got to market." (Oquist, 1980) (p.229)

The following passage describes the concrete methods through which coffee was collected by armed groups:

> "The coffee taken to the processing centers had been collected in various ways: sometimes it was the quota that the guerillas demanded [. . .] from the immediate farm managers (small landowners or tenant farmers); in certain villages in Caicedonia and Génova, for example, there were even designated spots where the bundle of coffee intended for the guerrillas had to be left periodically; sometimes they were sacks of dried coffee from the warehouse or coffee that was still damp, piled up in the stack, which was attacked and stolen; in the mountains, gangs carrying sacks [. . .] or ghost jeeps and masked men in the areas around Armenia, Montenegro, and Quimbaya, continued their nightly raids. Sometimes, in short,

> the coffee for sale had been picked straight from the tree, particularly on land that had been abandoned in a hurry; pickers would enter the coffee plantations at night for this purpose, often the local residents themselves; this was a very common practice during the early years in areas where liberals were being persecuted; and in the final years this was still the case along the riverbanks, when it was especially the conservatives who had left; the liberal guerrillas would spread the word: go harvest the coffee from such-and-such an abandoned farm; out of every ten arrobas, three or four had to be handed over to the gang; the remaining six or seven arrobas belonged to the pickers; both groups would sell it to the merchant at half price." (Ortiz Sarmiento, 1995) (pp.296–297)

(Roldán, 2002) documents how these operations followed the seasonal rhythm of the coffee economy in southwestern Antioquia:

> "[The guerrillas] continued to force Conservative peasants off lands [. . .] and penetrated and took control of important coffee plantations [. . .] Attacks occurred most frequently in July and August before the onset of the main coffee harvest, a period in which large numbers of men were without jobs or partially unemployed and beginning to be desperate for food and money. Guerrilla activity was next most intense in February, March, and April, just before and at the onset of the *travesia*, during the coffee crop's second flush [. . .] Guerrilla activities were preceded by the *boleteo* (written messages sent to threaten potential victims and exact a forced contribution); the penetration of isolated or unpatrolled villages and settlements; the confiscation of land, arms, and food; widespread cattle-rustling; and selective assassination and arson [. . .]
>
> Complaints regarding the presence of the guerrillas focused primarily on their extensive cattle-rustling and coffee theft rather than on the assassination or displacement of peasants [. . .] The complaint's principal concern was once more the economic effect of violence: 'several incidents of coffee looting had already occurred, as had orders to hand over significant sums of money.' Affluent hacienda owners from Urrao, for instance, had begun to relocate to Betulia to seek protection [. . .] similar incidents of economically motivated out-migration [were reported]." (Roldán, 2002) (pp.248–252)

Economic organization: sharecroppers, merchants, and intermediaries. The appropriation of coffee rents operated through an economic organization that linked armed groups to the broader rural economy through sharecroppers, farm managers, and village merchants.

> "The owner of a large coffee estate in La Camara complained that guerrillas had taken over private properties in the hamlet and appropriated parts of the

> coffee harvest since 1950 and that they were planning to do the same in 1953. In contrast to what was happening in Urrao, the estates in La Camara were not being destroyed or taken over directly by the guerrillas. The majority of La Camara's inhabitants were Liberal, as were the majority of *mayordomos*, but the estate owners were sometimes Conservative. To impede outright land confiscation, *mayordomos* and guerrillas from Urrao had worked out a deal: in return for not destroying haciendas, *mayordomos* received half the coffee production while the owner's share was given to the guerrillas. Witnesses also began to report that guerrillas were paid a wage and that they alternated between making war and sowing fields on empty public lands and usurped properties." (Roldán, 2002) (p.266)

The chain of intermediaries that connected stolen coffee to the export market is described in detail by (Ortiz Sarmiento, 1995):

> "This situation can be explained in large part by the chain of intermediaries that had always characterized the flow of coffee between the direct producers and the exporters. The latter [. . .] neither gained nor lost anything from La Violencia, nor were they interested in scrutinizing the ethical circumstances surrounding the origin of the beans they received.
>
> What set exporters apart, just as much as banks, was that the social influence they wielded depended very little on local vicissitudes, not even during such a turbulent period; as we have noted elsewhere, they had no need to involve themselves in the quarrels of local politics; they seemed to float safely above all the storms, and yet a large portion, if not the largest, of regional wealth remained in their hands [. . .]
>
> At the other end of the chain, very close to the smallholder, was undoubtedly the most crucial link: the village merchant. His new role in facilitating the laundering of looted coffee was the irrefutable expression of his very essence: bringing two distinct worlds into contact, village production and global trade; in his hands, the plunder of the villages became exportable merchandise. The middleman, of course, reaped a handsome profit. After all, coffee sourced through theft generally cost half the going price. Since the guerillas, or the mere opportunistic looters (who were the most numerous), had not been required to work the same hours or meet the same demands as under normal conditions, and since they also viewed the legalization that would allow them to purchase other goods as a favor, they never haggled over the price." (Ortiz Sarmiento, 1995) (pp.296–297)

> "Hence, after 1954–1955, following the outbreak of bandolerismo, two practices emerged that would characterize the relationship between coffee and La Violencia:

> (1) the eviction of peasants from their land or the purchase of it at derisory prices just as a promising harvest appeared at hand; and (2) the theft or "confiscation" of coffee once peasants had harvested and processed it.
>
> The first practice was carried out by non-bandoleros—hacendados, other peasants, but especially merchants—who took advantage of the prevailing apprehensions to frighten producers; the second became a typical means of support for cuadrillas. Sharecroppers would play a key role in both cases, since they were the only actors who continued to have an ongoing physical presence on fincas whose owners had been killed or forced to flee. For absentee landowners, the only guarantee of retaining a minimal portion of the harvest, or at least of forestalling the immediate loss of their property, was to contract a sharecropper who was accepted by 'those who controlled the area.' Sharecroppers had all the means of production at their disposal and would determine the fate of the harvests, though naturally under the pressure and manipulation both of merchants, with whom they were customarily in debt, and of bandoleros, who threatened them or simply convinced them of the advantages of complicity.
>
> The sharecropping system, then, allowed cuadrillas to appropriate part of the harvests without affecting the interests of peasant cultivators. This eventually led sharecroppers to understand that cooperating with bandoleros meant less interference from landowners in their daily work and a greater share of the earnings from the harvests, or even the possibility, as many of them foresaw, of appropriating the fincas. The vereda bosses, organically linked to bandas in Quindío, provide clear evidence of this phenomenon. Since booty in the form of coffee is easy to expropriate and market, bandoleros could not only survive but even become profitable." (Sánchez et al., 2001) (p.30)

The same source describes how extortion and intermediaries operated within these local economies:

> "Bandoleros on both sides were equally inclined to resort to procedures that would become classic, such as the *boleteo* (extortion through the delivery of a letter) to pressure the owner of a finca to sell or abandon it [. . .] After the bandoleros would come the *usufructuarios* (literally, "beneficiaries"): local merchants, neighboring landowners who wanted to extend their boundaries, bosses who wanted to get rid of a sharecropper or avoid paying wages. The most conflictive relationship linked to La Violencia appears to have arisen between landowners and sharecroppers [. . .] The owner of a farm in Génova [. . .] complained that in the preceding year he had gone through four sharecroppers: the first had been killed, the second had worked 'about fifteen days picking coffee . . . [then] he picked up the coffee and

ran off and didn't give me any accounting, all he gave me was one arroba [about 25 pounds] of coffee. And the next one did give me a little coffee and told me he wasn't staying any longer, and gave me the *boleta*.' The landowner had not been able to appear on his finca for a year (Sumario 9, 109). Taking advantage of this climate of terror, many administrators came to own not only the harvests but also, although temporarily, the land they administered or adjoining parcels." (Sánchez et al., 2001) (pp.39–41)

Territorial division and financing of armed groups. As the conflict evolved, rival armed groups divided coffee-producing territory among themselves and developed systematic financing structures.

"When Efraín González left Quindío, fellow Conservatives Melco and "Polancho" replaced him. They and their most famous Liberal adversary, Chispas, reached a tacit understanding to divide the area that runs along the Central Cordillera, including part of Quindío and two large coffee-growing centers of northern Valle del Cauca: Sevilla and Caicedonia. Melco and Polancho set up their base of operations in corregimientos (an administrative designation for villages lying outside a cabecera, or municipal seat) in Pijao and in Caicedonia. This area formed a politically homogeneous Conservative zone, referred to sardonically as the "Independent Republic" or "Sovereign State" of Aures. Chispas established Quebradanegra, at that time within the jurisdiction of the wealthy coffee-growing municipality of Calarcá and today part of the municipality of Córdoba, as his center of operations.

These rival bandas generally used the same technique to defend their areas of control or to penetrate those of their adversaries: physical elimination, usually of innocent peasants, and with a deliberately competitive brutality so as to sow terror. Within their respective dominions, the social bases of their militants and the support networks were not very different either. One printed comment regarding Chispas's cuadrilla could be considered valid for all cuadrillas, at least in their initial stage: '[I]t exercised an illegitimate power discussed neither by the landowners nor by the illiterate laborers' (La Nueva Prensa 96, May 3, 1963, 25). And the finca administrators, who appear to have been the persons most able to obtain an economic advantage from the anxiety that had taken hold in the countryside, discussed this 'illegitimate power' even less." (Sánchez et al., 2001) (pp.39–41)

"To maintain such an organization obviously required large sums of money. Where did they get this money? We can distinguish three analytically distinct

sources from different times in the cuadrillas' existence: Initially, funding was not a serious problem because, as mentioned, Chispas had been recruited and bankrolled by Liberal landowners who had, more or less formally, agreed to pay a periodic "protection quota." Later, funding became more complicated, since the landowners began to renege on their original commitments or fulfill them only grudgingly, leading the cuadrillas to demand obligatory contributions based on the size of the landholding and to resort to the boleteo, that is, to issue death threats for those who failed to make the required contribution by the peremptorily established deadline. Finally, when the landholders attempted to transfer the burden of the payments to sharecroppers, and thus keep part of the remuneration that contractually belonged to them, the affected parties openly sided with the cuadrillas. In this alliance, they inverted the traditional relationship of domination, placing landholders in a position of subordination." (Sánchez et al., 2001) (p.64)

A.2 Robustness

A.2.1 Persistence and Anti-Persistence

Conflict Incidence (Table 1.1)

The composite early violence indicator used in Table 1.1 combines two independent sources: (Policía Nacional, 1959) (police reports) and (Guzmán et al., 1962) (Guzman reports). A concern with the composite measure is that one of its two components might drive the entire persistence result. Table A.8 reports the results. The decomposition confirms that the persistence result is not an artifact of a single source.

Table A.8. Robustness: Early Violence Components (LPM)

	Criminality (Dummy)				**Guzman Rep. (Dummy)**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Civil wars battles (dummy)	-0.011				0.028			
	(0.028)				(0.034)			
1000 Days battles (dummy)		0.074**				0.041		
		(0.036)				(0.038)		
Land conflicts 19th c. (dummy)			0.189**				0.160**	
			(0.076)				(0.079)	
Any 19th c. conflict				0.059**				0.089***
				(0.029)				(0.033)
Dept. FE	Y	Y	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y	Y	Y
HGU level	1892	1892	1892	1892	1892	1892	1892	1892
N	651	651	651	651	651	651	651	651

Notes: Same specification as Table 1, Panel A. All independent variables enter as dummies. Each column is a separate regression. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Conflict Intensity (Table 1.2)

Alternative Outcome Measure. The main text established that nineteenth-century conflict does not predict the intensity of late-period violence, measured through violent deaths in 1962. A natural concern is that this null might be specific to that particular outcome measure. The police death statistics cover a single year and capture only one dimension of violence, lethal outcomes reported to law enforcement. If the null reflected a limitation of this source rather than a genuine absence of persistence, replacing the outcome with an independently compiled measure should produce different results.

I address this concern by replacing violent deaths with clashes from Guzman Reports of the Guzman reports (1958–1963). This alternative measure differs from the baseline in three respects: it covers a broader time window, it captures armed confrontations rather than deaths, and it was compiled from a different archival source. Table A.9 reports the results. None of the three nineteenth-century conflict measures (civil war battles, Thousand Days battles, or land conflicts) predicts the intensity of clashes. All coefficients are small and statistically insignificant.

Table A.9. Robustness: Alternative Intensity Measure

	Clashes 1958–63 (PPML)		
	(1)	(2)	(3)
Civil wars battles (count)	0.044		
	(0.057)		
1000 Days battles (count)		-0.035	
		(0.060)	
Land conflicts 19th c. (dummy)			0.170
			(0.500)
Dept. FE	Y	Y	Y
State controls	Y	Y	Y
Geographic controls	Y	Y	Y
Pop. controls	Y	Y	Y
HGU level	1892	1892	1892
N	651	651	651

Notes: Pseudo-Poisson Maximum Likelihood (PPML). Dependent variable: total clashes 1958–63. Robust standard errors (HC1) in parentheses. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. Land conflicts 19th c. always enters as a dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Extensive vs. Intensive Margin. The PPML specification in Table 1.2 pools two distinct questions: whether a municipality experiences any late-period violence at all, and how many deaths occur conditional on experiencing some. A concern is that persistence might operate on one of these margins but not the other, and the pooled specification might obscure this by averaging across both. To examine this possibility, I decompose the count model into its extensive component (the probability of any late-period deaths) and its intensive component (the number of deaths conditional on positive values).

Table A.10 reports both margins for violent deaths. On the extensive margin (left panel), Thousand Days battles do predict the probability of any late-period deaths. On the intensive margin (right panel, $N = 220$), however, past conflict has no predictive power for the number of deaths conditional on experiencing any violence. The coefficient is small and statistically insignificant for all three nineteenth-century conflict measures. Table A.11 replicates this decomposition for clashes, with the same pattern with land conflicts now entering significantly, but conditional on experiencing any clashes, past conflict does not predict how many occur.

The extensive-margin persistence is consistent with the incidence results from Section 1.4: municipalities with a history of conflict are more likely to see any late-period violence. But this extensive-margin effect does not extend to meaningful variation in intensity. Municipalities that experienced nineteenth-century conflict may see a few additional deaths, but past conflict does not explain why some municipalities suffered death rates comparable to the world's most violent countries while others experienced minimal violence. This is confirmed in the next robustness section on threshold-varying persistence.

Table A.10. Robustness: Extensive vs. Intensive Margin of Late Violence

	Extensive			**Intensive (DV > 0)**		
	Any deaths (dummy) (LPM)			Deaths count (PPML)		
	(1)	(2)	(3)	(4)	(5)	(6)
Civil wars battles (dummy)	-0.027					
	(0.034)					
1000 Days battles (dummy)		0.132***				
		(0.036)				
Land conflicts 19th c. (dummy)			0.042			0.231
			(0.068)			(0.367)
Civil wars battles (count)				0.032		
				(0.039)		
1000 Days battles (count)					-0.042	
					(0.064)	
Dept. FE	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y
HGU level	1892	1892	1892	1892	1892	1892
N	651	651	651	220	220	220

Notes: Left panel: OLS (LPM) on full sample. Right panel: Pseudo-Poisson Maximum Likelihood (PPML) on violent deaths > 0 subsample. Robust standard errors (HC1) in parentheses. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. Land conflicts 19th c. always enters as a dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11. Robustness: Extensive vs. Intensive Margin — Alternative Measure

	Extensive			**Intensive (DV > 0)**		
	Any clashes (dummy) (LPM)			Clashes count (PPML)		
	(1)	(2)	(3)	(4)	(5)	(6)
Civil wars battles (dummy)	-0.039					
	(0.031)					
1000 Days battles (dummy)		0.066**				
		(0.034)				
Land conflicts 19th c. (dummy)			0.215***			-0.207
			(0.065)			(0.478)
Civil wars battles (count)				0.079		
				(0.053)		
1000 Days battles (count)					-0.038	
					(0.064)	
Dept. FE	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y
HGU level	1892	1892	1892	1892	1892	1892
N	651	651	651	150	150	150

Notes: Left panel: OLS (LPM) on full sample. Right panel: Pseudo-Poisson Maximum Likelihood (PPML) on clashes > 0 subsample. Robust standard errors (HC1) in parentheses. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. Land conflicts 19th c. always enters as a dummy. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Threshold-Varying Persistence. The extensive-intensive margin analysis showed that Thousand Days battles predict whether any late-period deaths occur but not how many across both conflict measure. A remaining question is whether this extensive-margin persistence extends to low but nontrivial levels of violence. If municipalities with nineteenth-century conflict experience modestly higher death counts, not just the jump from zero to one, then persistence might be more substantively meaningful than the extensive-margin might suggest.

I test this by estimating the same extensive margin regression from the previous robustness section, but changing the threshold of the dummy which is the outcome. This amounts to a series of regressions in which the dependent variable is a binary indicator equal to one if violence exceeds a threshold k, varying k from low values (where the question is close to the extensive margin) to high values (where only the most violent municipalities are coded as one). If persistence extends beyond the basic extensive margin, the coefficient on Thousand Days battles should remain positive and significant as k increases. If, instead, the extensive-margin result reflects only a few additional deaths in otherwise low-violence municipalities, the coefficient should shrink toward zero as the threshold rises.

Figures A.3 and A.4 show the results for violent deaths and clashes, respectively. In both cases, the coefficient is positive and significant at the lowest thresholds, consistent with the extensive-margin persistence documented above, but shrinks monotonically as the threshold increases. For deaths, the coefficient becomes statistically insignificant by the threshold of 3 deaths, well below the average of 8 deaths in municipalities with any deaths. The pattern for clashes is even clearer. Once the threshold is changed from 0 to 1 clash, the coefficient is no longer siginificant and turns negative quickly. This confirms that the extensive-margin persistence is transmitting specific information: past conflict predicts a few additional deaths in otherwise low-violence municipalities, but it has no predictive power for the intense violence that characterized the worst-affected areas during *La Violencia*.

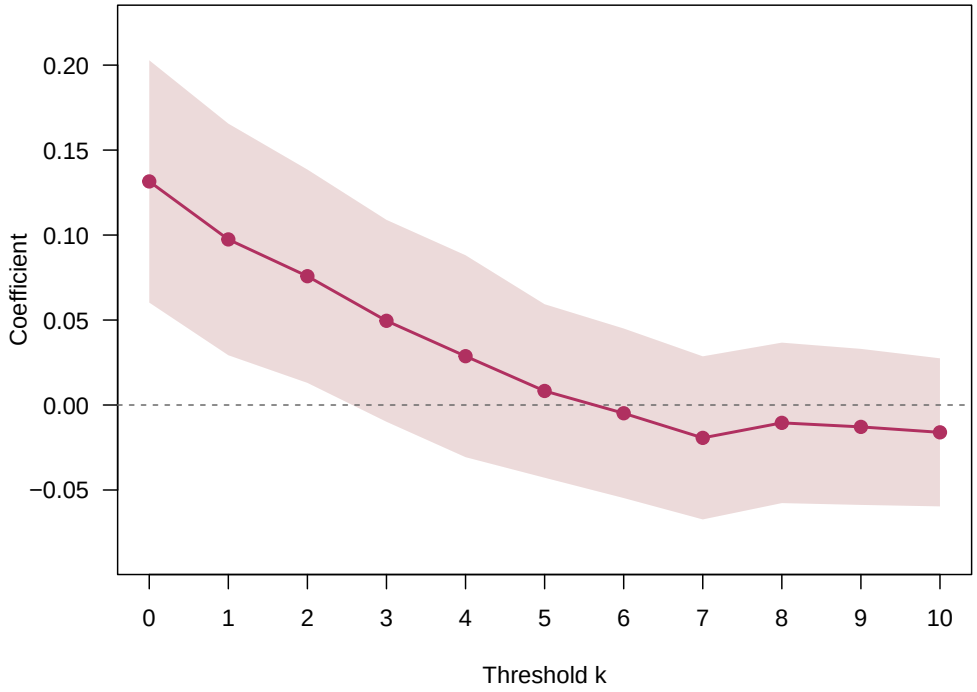


Figure A.3. Threshold-Varying Persistence: Violent Deaths

Notes: Each point plots the OLS (LPM) coefficient from regressing a binary indicator equal to one if violent deaths in 1962 exceed threshold k on a 1000 Days War battles dummy. The shaded band shows the 90% confidence interval (HC1 robust standard errors). Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to department capital, log population 1905. HGU level: 1892 ($N = 651$).

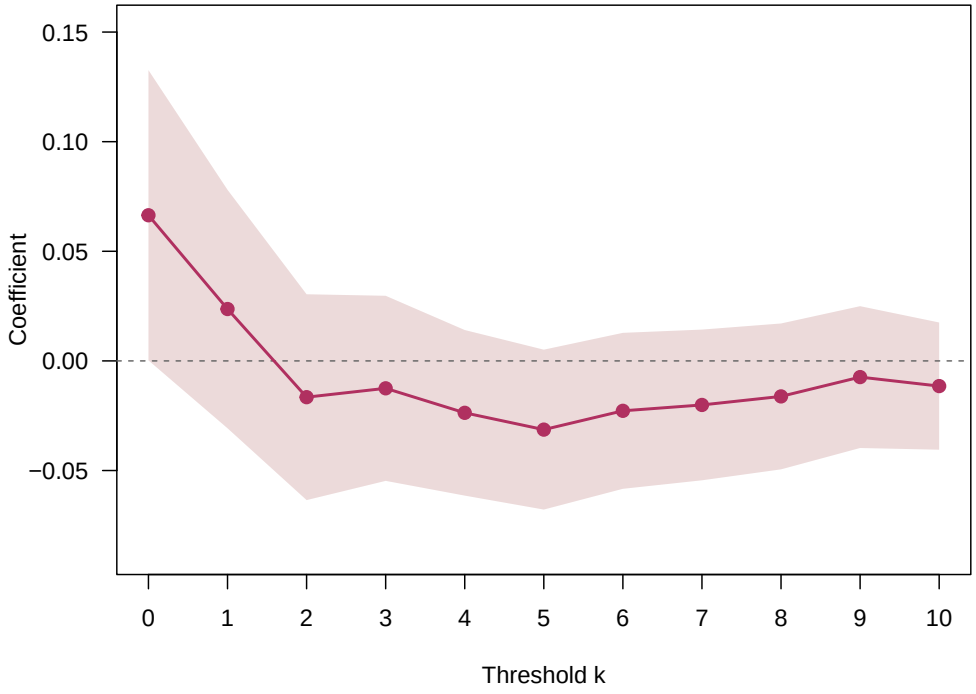


Figure A.4. Threshold-Varying Persistence: Clashes

Notes: Each point plots the OLS (LPM) coefficient from regressing a binary indicator equal to one if total clashes (1958–63) exceed threshold k on a 1000 Days War battles dummy. The shaded band shows the 90% confidence interval (HC1 robust standard errors). Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to department capital, log population 1905. HGU level: 1892 (N = 651).

Quantile Regression. The threshold-varying exercise examined whether persistence appears at different levels of the unconditional violence distribution. A related but distinct concern is that persistence might be concentrated at specific quantiles of the *conditional* distribution. For instance, past conflict might matter for the most violent municipalities conditional on observables, even if it does not matter on average. Quantile regressions address this concern by estimating the effect of Thousand Days battles on violent deaths at each quantile of the conditional distribution, from the 10th to the 90th percentile.

I estimate quantile regressions at five quantiles ($\tau = 0.10, 0.25, 0.50, 0.75, 0.90$) with the same controls as the baseline specification. Figure A.5 restricts the sample to municipalities with at least one death ($N = 220$), focusing on the intensive margin among municipalities that experienced violence. Figure A.6 broadens the sample to municipalities that either experienced deaths or had nineteenth-century battles ($N = 317$), which captures municipalities that were "at risk" of persistence. In both samples, the coefficient on Thousand Days battles is near zero and statistically insignificant across the entire conditional distribution. There is no quantile at which past conflict predicts violence intensity: not at the median, not at the upper tail, and not at the lower tail. The null on intensity persistence holds throughout the conditional distribution. This rules out the possibility that the average null result masks meaningful heterogeneity across quantiles.

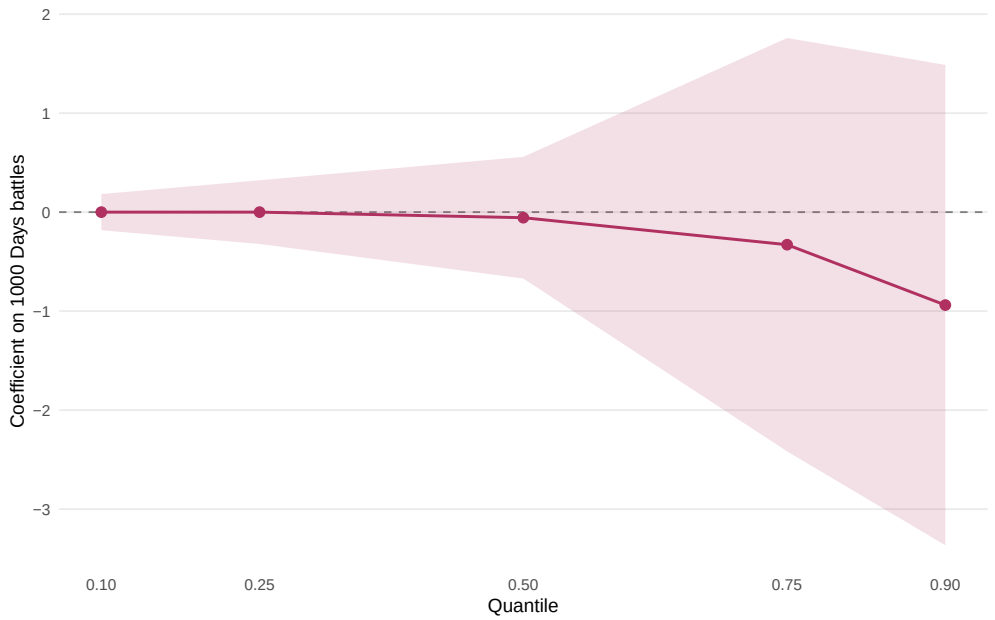


Figure A.5. Quantile Regression: Deaths > 0 Sample

Notes: Each point plots the quantile regression coefficient of violent deaths (1962) on 1000 Days War battles (count) at quantile τ. Sample restricted to municipalities with at least one violent death in 1962. The shaded band shows the 95% confidence interval based on xy-pair bootstrap standard errors ($R = 200$). Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to department capital, log population 1905. HGU level: 1892.

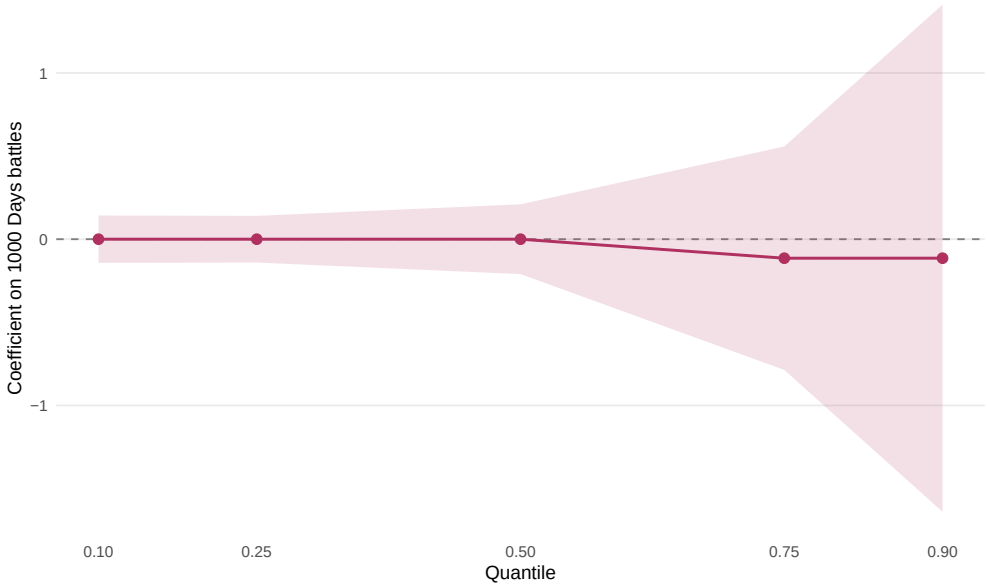


Figure A.6. Quantile Regression: Deaths > 0 or Battles > 0 Sample

Notes: Each point plots the quantile regression coefficient of violent deaths (1962) on 1000 Days War battles (count) at quantile τ. Sample restricted to municipalities with at least one violent death in 1962 or at least one 1000 Days War battle. The shaded band shows the 95% confidence interval based on xy-pair bootstrap standard errors ($R = 200$). Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to department capital, log population 1905. HGU level: 1892.

Frontier Subsamples. The null on intensity persistence pools the entire country, but the new frontier (Antioquia, Caldas, Tolima, Valle) and the rest of Colombia had very different conflict experiences. A key argument in the paper is that the new frontier was the epicenter of late-period *La Violencia*, concentrating the majority of violent deaths despite containing a minority of municipalities. A concern is that the pooled null might mask regional heterogeneity: perhaps persistence holds in one region but not the other, and averaging across them produces a misleading null.

I address this by splitting the sample into the new frontier ($N = 157$) and the rest of Colombia ($N = 494$) and re-estimating the intensity specification in each subsample. Table A.12 reports the results for the new frontier. All three nineteenth-century conflict measures (civil war battles, Thousand Days battles, and land conflicts) yield coefficients that are small and statistically insignificant. Table A.13 reports the results for the rest of Colombia. Here, land conflicts show a marginally significant positive coefficient of 0.719, significant at the 10% level, while battle counts remain insignificant.

The subsample analysis reinforces the main finding. The coffee frontier, where late-period violence was most intense and where the variation that Section 1.5 seeks to explain is concentrated, is entirely disconnected from nineteenth-century conflict patterns in terms of intensity. The weak positive association for land conflicts in the rest of Colombia does not alter this conclusion, as it is marginal in significance and absent in the region that matters most for the paper's argument.

Table A.12. Frontier Subsample: New Frontier (Antioquia, Caldas, Tolima, Valle)

	(1)	(2)	(3)
Civil wars battles (count)	-0.002		
	(0.068)		
1000 Days battles (count)		-0.020	
		(0.062)	
Land conflicts 19th c. (dummy)			-0.174
			(0.580)
Dept. FE	Y	Y	Y
State controls	Y	Y	Y
Geographic controls	Y	Y	Y
Pop. controls	Y	Y	Y
HGU level	1892	1892	1892
N	157	157	157

Notes: Pseudo-Poisson Maximum Likelihood (PPML). Dependent variable: violent deaths 1962. New Frontier. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13. Frontier Subsample: Rest of Colombia

	(1)	(2)	(3)
Civil wars battles (count)	0.068		
	(0.054)		
1000 Days battles (count)		0.119	
		(0.087)	
Land conflicts 19th c. (dummy)			0.719*
			(0.402)
Dept. FE	Y	Y	Y
State controls	Y	Y	Y
Geographic controls	Y	Y	Y
Pop. controls	Y	Y	Y
HGU level	1892	1892	1892
N	494	494	494

Notes: Pseudo-Poisson Maximum Likelihood (PPML). Dependent variable: violent deaths 1962. Rest of Colombia. Controls: department FE, colonial state presence dummy (1794), encomienda dummy, altitude, ruggedness, distance to dept. capital, log population 1905. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Liberal Voting (Table 1.3)

The main text established that Liberal vote share in 1946 is a strong predictor of early *La Violencia*, even after controlling for past conflict measures and competing theories. A concern with this result is reverse causality: if municipalities already experiencing partisan tensions in the late 1940s voted more heavily Liberal as a consequence, then the direction of causality would run from violence to Liberal voting rather than the other way around. Section 1.4 addressed this concern by replacing the 1946 measure with Liberal vote share in 1856, recorded a full century before *La Violencia*, and showing that this deeper measure also predicts early violence. This appendix table provides the full set of results underlying that argument.

Table A.14 reports the direct effect of Liberal vote share in 1856 on early violence outcomes. The first step is to confirm that partisan identity itself persists across the century: column (1) shows that Liberal vote share in 1856 predicts Liberal vote share in 1946 by 22.5 percentage points, statistically significant at the 1% level. Municipalities that were Liberal before the civil wars of the late nineteenth century remained Liberal on the eve of *La Violencia*. The second step asks whether this deep partisan alignment predicts violence directly. Under baseline controls (columns 2–4), Liberal vote share in 1856 predicts any early violence by 14.0 percentage points (significant at the 1% level), Guzman Reports violence by 9.4 percentage points (significant at the 5% level), and Criminality Reports by 17.5 percentage points (significant at the 1% level). Adding past conflict measures and the full set of competing theories as controls (columns 5–7) attenuates the coefficients modestly (to 10.9, 5.2, and 14.3 percentage points respectively), but all remain statistically significant.

These results assuage the reverse causality concern. A municipality's partisan alignment in the 1850s cannot be caused by conditions that arose in the 1940s, because those conditions did not yet exist when the 1856 vote was cast. The fact that this deep historical measure predicts

violence across all three outcomes, even after controlling for observable confounders, confirms that partisan identity is a genuine channel of conflict persistence rather than an artifact of contemporaneous confounding.

Table A.14. Liberal 1856 and Early Violence: Reduced Form

		Baseline			**Full Controls**		
	Liberal 1946 (1)	Any Early V. (2)	Guzman Rep. (3)	Criminality (4)	Any Early V. (5)	Guzman Rep. (6)	Criminality (7)
Liberal 1856	0.225*** (0.035)	0.140*** (0.047)	0.094** (0.046)	0.175*** (0.039)	0.109*** (0.040)	0.079** (0.039)	0.147*** (0.032)
Dept. FE	Y	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y	Y
Geographic controls	N	Y	Y	Y	Y	Y	Y
Past conflict	N	N	N	N	Y	Y	Y
Competing theories	N	N	N	N	Y	Y	Y
HGU level	1892	1892	1892	1892	1892	1892	1892
N	638	651	651	651	638	638	638

Notes: OLS. All columns use liberal vote share 1856 as IV. Col (1): persistence of liberal voting. Cols (2)–(4): baseline controls (pop 1851, state, geography). Cols (5)–(7): full controls (pop 1938, state, geography, past conflict, competing theories). HC1 robust SEs in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Antioquia Event Study (Figure 1.5)

The event study in the main text exploits the timing of Laureano Gómez coming to power in 1950 to show that liberal municipalities experienced a sharp increase in violence relative to conservative municipalities after this date. The identifying assumption is that liberal and conservative municipalities would have followed parallel violence trajectories absent the political shock. Two concerns arise: first, that pre-existing differences between the two groups, rather than partisan persecution, might explain the results; and second, that the event study design, which uses annual violence counts, might not capture the cumulative effect of partisan identity on total deaths.

I address the first concern with a balance test. Figure A.7 compares liberal-majority and conservative-majority municipalities in Antioquia on pre-treatment observables: population, altitude, distance to Medellín, colonial state presence, and pre-1948 economic indicators. The distributions are comparable across all covariates except for altitude and land conflicts in the 20th century, supporting the assumption that liberal and conservative municipalities were very similar on observable dimensions before the onset of violence. The difference in altitude seems unlikely to drive such large violence, absent any historiographical argument explaining it. The difference in land conflicts on the other hand could be actively contributing to the effect seen on the event study. However, as is shown in the following cross-sectional robustness, land conflict is not significant predictor of total deaths in Antioquia during this period. Hence, while some of the liberal effect in the event study might be driven by an overlap of land conflict locations, it seems unlikely that land conflicts dominates the liberal channel, even more so considering the overwhelming historiographical evidence pointing in the direction of the liberal channel.

I address the second concern by estimating a PPML cross-sectional regression of total deaths during the early period (1949–1953) on Liberal vote share in 1946 for the 99 Antioquia municipalities. Table A.15 reports the result: the coefficient is 3.655, statistically significant at the 5% level. This confirms the event study finding in a count-data framework that accounts for the non-negative, skewed distribution of death counts. The cross-sectional and panel approaches tell the same story: within Antioquia, liberal municipalities experienced significantly more violence during the partisan phase of *La Violencia*.

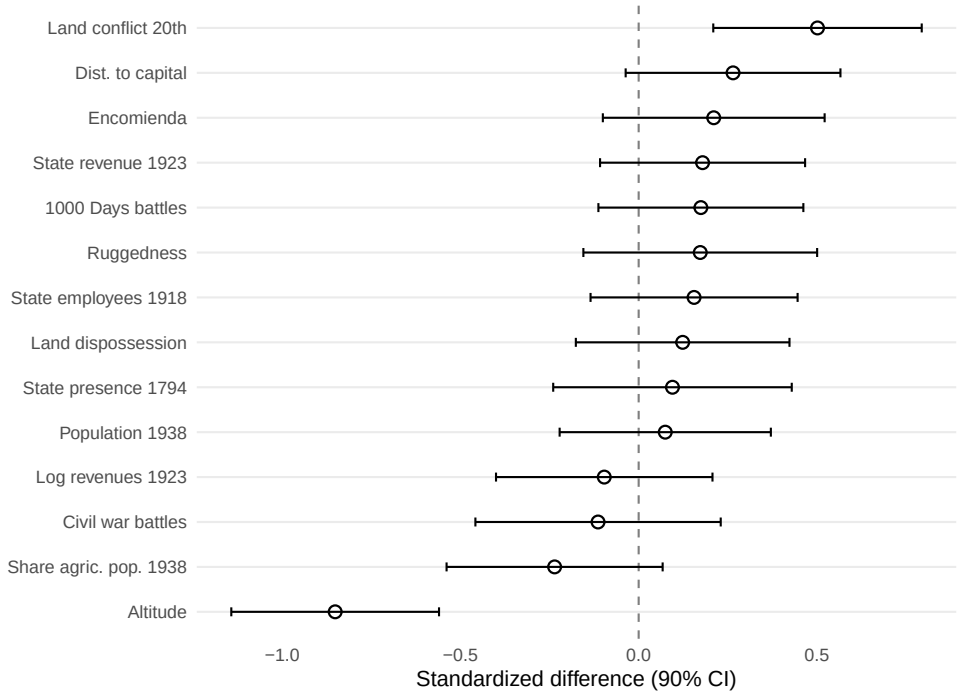


Figure A.7. Balance Test: Covariates by Liberal Treatment (Antioquia)

Notes: Each panel shows the distribution of a pre-treatment covariate for liberal-majority and conservative-majority municipalities in Antioquia. Treatment is defined as having a liberal vote share above the median in the 1946 election.

Table A.15. Liberal Voting and Violence: PPML (Antioquia)

	(1) Total Deaths (Roldan)
Liberal 1946	3.655** (1.436)
Land conflict 20th	0.371 (0.592)
Dept. FE	N
State controls	Y
Geographic controls	Y
Pop. controls	Y
Past conflict	Y
Competing theories	Y
HGU level	1946
N	99

Notes: Pseudo-Poisson Maximum Likelihood (PPML). Dependent variable: total violent deaths 1949–1953 (Roldan). Antioquia municipalities only. HC1 robust SEs in parentheses. No department FE (single department). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2.2 Coffee and Violence

Alternative Measures

The main result in Table 1.4 uses violent deaths in 1962 as the dependent variable and coffee production in sacks (1944) as the key independent variable. Two concerns arise with these specific measurement choices. First, the result might be specific to one particular measure of violence: if the relationship holds for police-reported deaths but not for independently compiled conflict records, it could reflect a reporting artifact rather than a genuine association. Second, the result might be specific to the 1944 coffee production measure: if the relationship depends on the timing or unit of the coffee variable, it could reflect a confound particular to that year rather than the structural geography of coffee cultivation. I address each concern in turn.

Alternative Conflict Measures. I replace the dependent variable in the baseline specification with four alternatives that differ in source, time coverage, and measurement unit. Table A.16 reports the results. Columns (1)–(2) use clashes from Guzman Reports (1958–1963), an independently compiled source covering a broader time window and measuring armed confrontations rather than deaths. The standardized coffee coefficient is 0.766 in the full sample and 0.685 among coffee-producing municipalities, both statistically significant at the 1% and 5% levels respectively. If the baseline result reflected a peculiarity of the police death statistics, switching to an entirely different source and conflict dimension should weaken the association. It does not.

Column (5) uses deaths per 100,000 population (using the 1951 census as baseline) to account for the possibility that larger municipalities mechanically experience more deaths. The coefficient is 0.310, statistically significant at the 5% level, confirming that the relationship is not driven by population size. Column (6) uses a binary indicator for any deaths, estimated via a linear probability model with a coffee dummy rather than the continuous standardized measure: coffee-producing municipalities are 13.0 percentage points more likely to experience any violence, significant at the 1% level. The coffee-violence relationship holds across independently compiled sources, different time windows, per capita adjustments, and both the extensive and intensive margins. The finding is not an artifact of the specific outcome measure.

Table A.16. Robustness: Alternate Measures of Conflict

	OLS (Std)		**PPML (Counts)**†		**OLS (Std)**	**LPM**
	(1)	(2)	(3)	(4)	(5)	(6)
DV:	Full Clashes	Coffee > 0 Clashes	Full Clashes	Coffee > 0 Clashes	Full Deaths/100k	Full Any Deaths
Coffee 1944	0.766***	0.685**	0.430***	0.491***	0.310**	–
	(0.260)	(0.275)	(0.123)	(0.123)	(0.136)	
Coffee dummy	–	–	–	–	–	0.130***
						(0.047)
Dept. FE	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y
Competing theories	Y	Y	Y	Y	Y	Y
N	741	382	704	352	724	741
R^2 / Pseudo R^2	0.427	0.429	0.741	0.700	0.243	0.502
Within R^2 / Pseudo R^2	0.169	0.176	0.483	0.502	0.067	0.123
HGU level	1918	1918	1918	1918	1918	1918

Notes: Dependent variable varies by column. Cols. (1)–(2): OLS on standardized clashes (1958–63); standardized coefficients as in Table 1. Cols. (3)–(4): PPML on raw clash counts. † PPML coffee coefficient $\times 10^{-5}$. Col. (5): OLS on standardized deaths per 100,000 (1951 pop), no population control. Col. (6): LPM on $\mathbf{1}$(violent deaths > 0) with binary coffee indicator. HC1 robust SEs. Competing theories: liberal 1946, state revenue, state employees, pol. comp., land dispos., land conflict. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Alternative Coffee Measures. The baseline specification uses coffee production in sacks from the 1944 census. A concern is that this particular measure, recorded during the height of political polarization in the 1940s, might capture something other than the structural geography of coffee cultivation. If, for example, municipalities with more violence had their coffee production recorded differently, or if production in 1944 reflected short-run political conditions rather than long-run coffee cultivation patterns, the association could be spurious.

I address this concern by replacing the 1944 sacks measure with two alternatives that differ in both timing and unit. Table A.17 reports the results. Columns (1)–(4) use coffee area in hectares from the 1960 agricultural census. Although this measure might be endogenous to conflict given it is a measure taken during *La Violencia*, it provides a useful benchmark to verify that coffee locations during the conflict hadn't fundamentally been relocated. The OLS coefficient is 0.936, statistically significant at the 1% level and larger than the baseline estimate, indicating that the relationship is if anything stronger. Columns (5)–(8) use the number of coffee trees from the 1932 census, an even earlier historical measure of coffee cultivations that predates both *La Violencia* and the political polarization of the 1940s by more than a decade. The OLS coefficient is 0.485 in the full sample, statistically significant at the 5% level.

The fact that both alternatives (measured in different units and from different censuses) yield significant coefficients confirms that the coffee-violence relationship reflects the structural geography of coffee cultivation. The association is not an artifact of the specific timing of the 1944 measure, nor of the particular unit in which coffee is recorded.

Table A.17. Robustness: Alternate Coffee Measures

	Coffee Area (1960)				**Coffee Trees (1932)**			
	OLS (Std)		PPML		OLS (Std)		PPML	
	(1) Full	(2) Cof>0	(3) Full	(4) Cof>0	(5) Full	(6) Cof>0	(7) Full	(8) Cof>0
Coffee	0.936***	0.880***	0.439*	0.488**	0.485**	0.448**	0.164***	0.185***
	(0.256)	(0.263)	(0.224)	(0.224)	(0.217)	(0.226)	(0.042)	(0.046)
Dept. FE	Y	Y	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y	Y	Y
Competing theories	Y	Y	Y	Y	Y	Y	Y	Y
N	741	552	590	486	741	420	590	387
R^2 / Pseudo R^2	0.431	0.416	0.713	0.696	0.349	0.339	0.666	0.651
Within R^2 / Pseudo R^2	0.210	0.202	0.477	0.472	0.096	0.100	0.391	0.404
HGU level	1918	1918	1918	1918	1918	1918	1918	1918

Notes: DV: Violent deaths 1962. Cols. (1)–(2), (5)–(6): OLS; standardized coefficients as in Table 1. Cols. (3)–(4), (7)–(8): PPML on raw counts. † PPML coffee area coefficient $\times 10^{-4}$; coffee trees $\times 10^{-6}$. HC1 robust SEs. Cols. (1)–(4) control for log pop. 1938; cols. (5)–(8) use log pop. 1918 (pre-treatment for 1932 coffee trees). Competing theories: liberal 1946, state revenue, state employees, pol. comp., land dispos., land conflict. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Geographic Concentration

Both coffee production and violence are geographically concentrated in the new frontier departments (Antioquia, Caldas, Tolima, Valle). This geographic overlap raises the concern that the coffee-violence relationship might be driven by a single department with unusually high values of both variables, or that the result might be confined to the coffee heartland and not generalize to the rest of the country. I address these concerns with two analyses: a leave-one-department-out exercise and a frontier subsample split.

Leave-One-Department-Out. If the coffee-violence relationship is driven by one particular department (for example, if Tolima alone, with its extreme violence and large coffee sector, accounts for the entire association), then dropping that department should substantially weaken the coefficient. I test this by sequentially dropping each department that contains at least one coffee-producing municipality and re-estimating the baseline specification (column 1 of Table 1.4).

Figure A.8 plots the coffee coefficient from each leave-one-out regression. The dotted maroon line marks the full-sample estimate of 0.783. The coefficient remains positive and statistically significant regardless of which department is excluded. The estimates are tightly clustered around the baseline, with no single department producing a notable drop. This confirms that the coffee-violence relationship is not an artifact of any one department's extreme values; it reflects a pattern that holds across the full geography of coffee cultivation in Colombia.

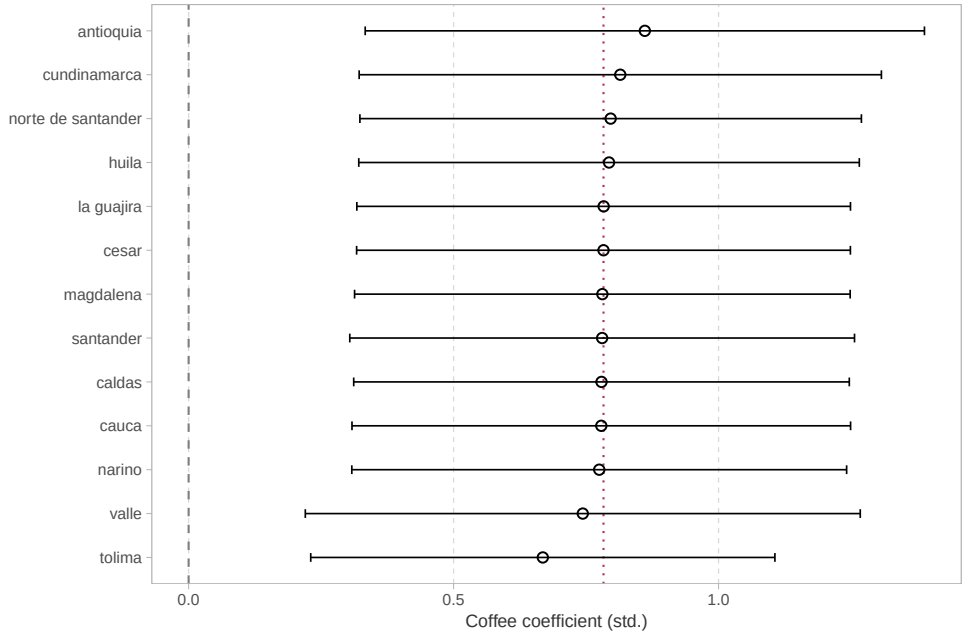


Figure A.8. Leave-One-Department-Out: Coffee Coefficient on Violent Deaths

Notes: Each point shows the standardized OLS coffee coefficient when one department is dropped from the sample. The dotted maroon line marks the baseline estimate (0.783) from column (1) of Table 1.4. Horizontal bars show 95% confidence intervals (HC1 robust standard errors). HGU level: 1918.

Frontier Subsamples. The leave-one-department-out analysis shows that no single department drives the result, but a related concern is whether the relationship is confined to the new frontier as a whole. The new frontier comprises only 26.2% of municipalities but concentrates 73.6% of violent deaths, 76.6% of clashes, and 75.9% of total coffee production. If the coffee-violence association existed only within these four departments, it might reflect something specific to the frontier's settlement history, institutional environment, or conflict dynamics rather than a general relationship between coffee rents and violence.

I address this by re-estimating the specification separately for the new frontier and all other departments. Table A.18 reports the results. Coffee is statistically significant at the 1% level in both subsamples: the standardized coefficient is 0.740 in the new frontier and 0.325 in the rest of Colombia. The effect is larger in the coffee heartland, which is consistent with the rents mechanism: where coffee production is more concentrated, the surplus available for armed group extraction is larger, and the association with violence is correspondingly stronger. But the relationship extends well beyond the frontier. Even in departments where coffee is less dominant and violence was less intense, municipalities with more coffee production experienced more violent deaths. The coffee-violence association is a national pattern, not a frontier-specific one.

Table A.18. Robustness: Frontier Subsamples

	New Frontier		**All Others**	
	(1) Deaths	(2) Clashes	(3) Deaths	(4) Clashes
Coffee 1944	0.740***	0.676**	0.325***	0.211**
	(0.262)	(0.281)	(0.114)	(0.083)
Dept. FE	Y	Y	Y	Y
State controls	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y
Competing theories	Y	Y	Y	Y
N	194	194	547	547
R^2	0.368	0.428	0.197	0.192
Within R^2	0.231	0.260	0.102	0.103
HGU level	1918	1918	1918	1918

Notes: OLS on z(violent deaths 1962). Standardized coefficients as in Table 1. HC1 robust SEs. New Frontier: Tolima, Caldas, Antioquia, Valle. All Others: all remaining departments. Competing theories: liberal 1946, state revenue, state employees, pol. comp., land dispos., land conflict. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Spatial Correlation

Cross-sectional regressions with geographic data are susceptible to spatial dependence. Recent critiques (Conley and Kelly, 2025) of empirical historical studies emphasize the potentially large confounding role of spatial autocorrelation, particularly in regressions such as main coffee specification, where coffee cultivation (1944) is measured well before the outcome (violent deaths in 1962). For example, coffee and violence might exhibit spatial trends: if both variables increase smoothly from east to west, a regression of one on the other could find a significant association even if no causal or structural relationship exists. I address these two concerns with Conley standard errors and a spatial unit root test.

Conley Standard Errors. Conley standard errors allow for arbitrary spatial correlation among residuals within a specified distance bandwidth, providing inference that is robust to the clustering of unobserved shocks across nearby municipalities. The key question is whether the significance of the coffee coefficient survives when standard errors account for this spatial dependence. I re-estimate the baseline specification using Conley standard errors at distance bandwidths ranging from 50 to 500 km, progressively allowing for correlation among municipalities that are farther apart.

Table A.19 reports the results. At all bandwidths, the coffee coefficient of 0.783 remains statistically significant at the 1% level. The largest spatial standard error is 0.218 at the 50 km bandwidth, which is in fact slightly smaller than the baseline HC1 robust standard error of 0.238. As the bandwidth increases, allowing for spatial correlation among increasingly distant municipalities, the standard errors remain stable or decline. This pattern indicates that spatial clustering of unobserved shocks does not inflate the baseline inference. If anything, accounting for spatial dependence slightly tightens the estimates.

Table A.19. Spatial Robustness of Coffee Effect (Standard Errors)

Bandwidth cutoff	Std. error on coffee (std.)
50 km	(0.2177)***
100 km	(0.1092)***
150 km	(0.1681)***
200 km	(0.0212)***
250 km	(0.0864)***
300 km	(0.0393)***
350 km	(0.0346)***
400 km	(0.0479)***
450 km	(0.0581)***
500 km	(0.0520)***

Notes: Table reports Conley spatial robust standard errors for the coffee coefficient (std. = 0.783) at varying distance cutoffs. Specification: column (1) of Table 1. Baseline HC1 SE = 0.2377. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Spatial Unit Root Tests. Conley standard errors address the concern that nearby municipalities share unobserved shocks, but they do not rule out a more fundamental problem: the possibility that the regression itself is spurious. In time-series econometrics, regressing one trending variable on another can produce significant coefficients even when the two series are unrelated. The spatial analogue arises when both the dependent and independent variables exhibit spatial trends: if coffee production increases smoothly from one region to another and violence does the same, a cross-sectional regression might find a significant association that reflects nothing more than shared spatial structure.

I test for this possibility by applying the Local Spatial Unit Root Test to the OLS residuals from the benchmark specification. The logic is as follows: if the regression captures a genuine relationship between coffee and violence, the residuals should be spatially stationary and should not exhibit the smooth spatial trends that would indicate a spurious regression. If, on the other hand, the regression is spurious, the residuals will inherit the spatial non-stationarity of the underlying variables. I use the recent diagnostic toolkit from (Becker et al., 2025). Table A.20 reports the results. The I(0) stationarity test statistic is 1.335 with a p-value of 0.873, indicating that the residuals are spatially stationary; the null of stationarity cannot be rejected. The I(1) unit root test statistic is 15.418 with a p-value of 0.001, rejecting the null of a spatial unit root. Together, these tests confirm that the coffee-violence relationship is not a spurious spatial regression. The association between coffee and violence reflects a genuine cross-sectional relationship, not an artifact of shared spatial trends.

Table A.20. Spatial Unit Root Tests on Benchmark Residuals

	Test Statistic	P-value
I(0) Stationarity Test (LFST)	1.335	0.873
I(1) Unit Root Test (LFUR)	15.418	0.001
HGU level	1918	

Notes: Results from the Local Spatial Unit Root Test (Becker et al., 2025). OLS residuals from the fully saturated benchmark regression (column 1 of Table 1.4) are tested for spatial stationarity. High p-value for the I(0) test indicates residuals are stationary; the low p-value for the I(1) test rejects a spatial unit root.

Shape of the Coffee Effect

The baseline specification imposes a linear relationship between standardized coffee production and violence. A concern is that the large standardized coefficient might be driven by a few extreme coffee-producing municipalities with unusually high death counts. If the relationship between coffee and violence is concentrated entirely in the right tail of the coffee distribution (for instance, if only the top 5% of coffee municipalities experience elevated violence), then the linear specification would overstate the breadth of the association. Alternatively, the relationship might be genuinely increasing across the full coffee distribution, which would strengthen the rents interpretation. Two analyses examine the functional form of the relationship.

Marginal Effects and Binscatter. I first characterize the shape of the coffee-violence relationship semi- and non-parametrically. Figure A.9 presents two complementary views. Panel A replaces the continuous coffee measure with quartile dummies among coffee-producing municipalities, using non-producers as the reference group, and estimates the PPML average marginal effect for each quartile. This approach allows the relationship to take any shape: the marginal effect of moving from the first to the second quartile need not equal the effect of moving from the third to the fourth. The gradient is increasing: each successive quartile of coffee production is associated with more additional deaths. This first view suggests that increases in coffee production monotonically increase death counts.

Panel B presents a non-parametric binscatter (Cattaneo et al., 2024) of violent deaths on log coffee, conditioning on all controls and department fixed effects. I also plot the histogram of coffee production of each bin behind the point estimates. This visualization shows the relationship after partialing out the influence of all covariates, allowing the reader to see the association between coffee and violence directly. Deaths roughly double from the lowest bins to the penultimate coffee bins. The top bin (which accounts for roughly half of all coffee production) concentrates the highest death counts more than doubling from the penultimate bin. Rents appear to bite hardest at the very top, consistent with the largest coffee-producing municipalities generating the most extractable surplus for armed groups, but the association extends well below the top of the distribution.

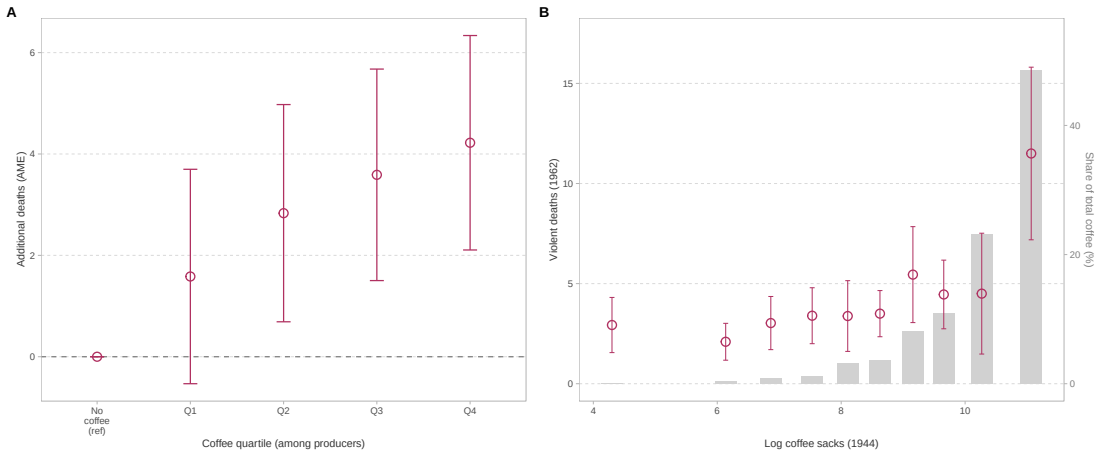


Figure A.9. Shape of the Coffee-Violence Relationship

Notes: Panel A: Average marginal effects from PPML with coffee quartile dummies (no coffee = reference). 90% confidence intervals shown. Controls: competing theories, geographic controls, population controls, and department FE. Panel B: Binscatter (Cattaneo et al., 2024) of violent deaths on log coffee, conditioning on all controls and department FE. Secondary axis shows the share of total coffee production in each bin.

Winsorization. The binscatter and quartile analysis show that the relationship is increasing across the coffee distribution, but the largest marginal effects are concentrated at the top. A remaining concern is that a handful of extreme observations in the upper tail of the *violence* distribution, where municipalities with exceptionally high death counts might disproportionately influence the linear coefficient. If the coefficient completely collapses when these extreme municipalities are removed, the result would be driven by outliers rather than by a broad-based association.

I address this by progressively winsorizing the outcome at the 99th, 95th, 90th, and 80th percentiles, effectively capping and then removing municipalities in the upper tail of the violence distribution. Table A.21 reports the results. For deaths, the coffee coefficient declines to 0.593 at the 99th percentile to 0.407 at the 80th percentile. For clashes, the decline is to 0.587 and to 0.327. At every winsorization threshold, the coefficient remains statistically significant: at the 1% level for deaths across all cutoffs, and at the 5% level for clashes at the most aggressive cutoff.

The pattern is informative. The top of the violence distribution does contribute to the magnitude of the coefficient (as the binscatter confirmed, the most violent municipalities are also the most coffee-intensive), but the association between coffee and violence survives the removal of up to 20% of the most extreme municipalities. The relationship is broad-based: coffee matters not only for the handful of municipalities that experienced catastrophic violence, but across the full range of the violence distribution.

Table A.21. Robustness: Winsorization of Outliers

	Deaths				**Clashes**			
Drop:	(1) >p99	(2) >p95	(3) >p90	(4) >p80	(5) >p99	(6) >p95	(7) >p90	(8) >p80
Coffee 1944	0.593***	0.528***	0.493***	0.407***	0.587***	0.360**	0.468***	0.327**
	(0.184)	(0.126)	(0.136)	(0.151)	(0.203)	(0.148)	(0.169)	(0.154)
Dept. FE	Y	Y	Y	Y	Y	Y	Y	Y
State controls	Y	Y	Y	Y	Y	Y	Y	Y
Geographic controls	Y	Y	Y	Y	Y	Y	Y	Y
Pop. controls	Y	Y	Y	Y	Y	Y	Y	Y
Competing theories	Y	Y	Y	Y	Y	Y	Y	Y
N	738	728	716	693	739	732	724	709
R^2	0.389	0.405	0.405	0.404	0.430	0.377	0.404	0.375
Within R^2	0.138	0.142	0.139	0.131	0.128	0.096	0.109	0.070
HGU level	1918	1918	1918	1918	1918	1918	1918	1918

Notes: OLS; standardized coefficients as in Table 1. Percentiles computed among observations with DV > 0 only; observations above the cutoff are dropped (zeros retained). Deaths ($N_{>0}$=255): p99=50, p95=29, p90=17, p80=10. Clashes ($N_{>0}$=171): p99=27, p95=18, p90=12, p80=8. Re-standardized within each subsample. HC1 robust SEs. Competing theories: liberal 1946, state revenue, state employees, pol. comp., land dispos., land conflict. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.2.3 Mechanisms

PPML Distance Margins

The main text presented OLS predicted margins showing that the gap in violent deaths between coffee and non-coffee municipalities is wide near early violence hubs and closes gradually with distance (Figure 1.6). A concern is that this distance-decay pattern might be an artifact of the linear OLS specification. Violent deaths are a count variable with many zeros and a long right tail, and a linear model might produce misleading predicted values, particularly at the extremes of the distance distribution, where predictions could turn negative or exhibit leverage from a few observations.

I address this by replicating the margins analysis using the PPML estimator, which models the conditional mean as an exponential function and naturally accommodates the count nature of the data. If the distance-decay pattern is robust to estimator choice, it should appear in the PPML margins as well. Figure A.10 plots the PPML predicted deaths by coffee status across the 0–100 km range. The pattern is qualitatively similar to the OLS results: coffee-producing municipalities are predicted to have more deaths at short distances, and the gap between coffee and non-coffee municipalities narrows with distance from early violence hubs. The nonlinear PPML specification does not find a significant difference of violence at the hub (even though the p-value is very close at 12.7%) and generates wider confidence intervals at larger distances, a consequence of the exponential functional form amplifying uncertainty in the tails. Despite this wider uncertainty, the qualitative pattern (rents matter most where fighting is feasible) holds across both linear and count-data estimators, confirming that the distance-decay result is not an artifact of the OLS specification.

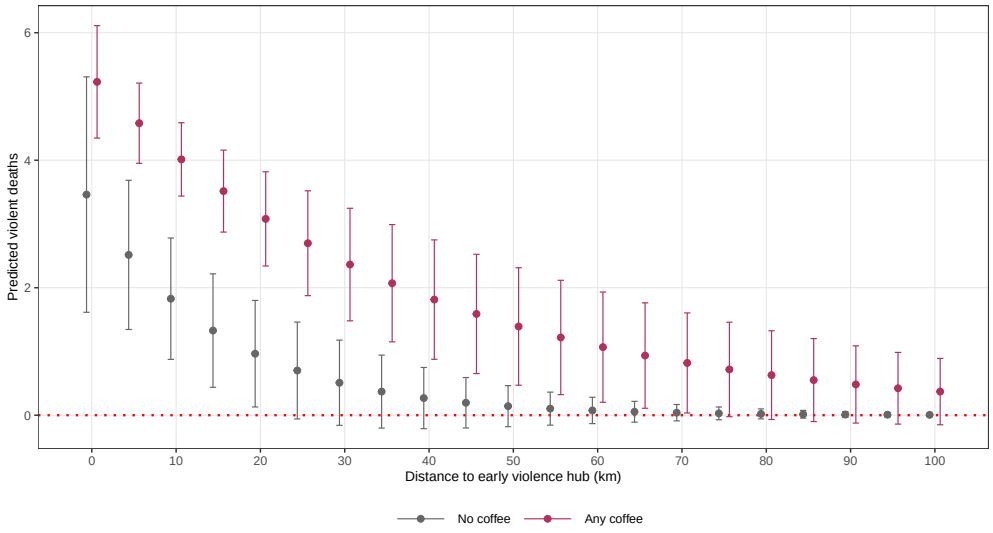


Figure A.10. PPML Predicted Violent Deaths by Coffee Status and Distance to Hub

Notes: Predicted violent deaths (1962) from PPML estimation. Margins evaluated at 0–100 km in 5 km steps. Whiskers show 95% confidence intervals. Controls: competing theories, state controls, geographic controls, population controls, and department FE.

A.3 Data & Measurement

A.3.1 Data Sources & Variables

Table A.22. Data Sources and Variables

Variable	Description	Coverage	Source
Conflict Data			
Number of battles during the Thousands Days War	Author provides a list of municipal locations of every battle occuring during the Thousands Days War at the month-year level for the whole time period of the war. The winner of each battle (Conservative, Liberal or unknown) is also given.	1899–1902	Own digitization from (Jaramillo Castillo, 1991) (pp. 378–408)
Number of battles	(Grueso, 1950) recorded the number of municipalities in each municipality-year from independence through to the end of the nineteenth century. These include national-level partisan civil wars as well as other partisan conflicts. (Fergusson et al., 2013) code each municipality-year mentinoned in the original work as 1 if any battle occurred that year and 0 otherwise. This data was shared by Leopoldo Fergusson and Juan Vargas.	1820–1885	Originally from (Grueso, 1950), coded by (Fergusson et al., 2013) and (Fergusson and Vargas, 2023)
Number of land dispute incidents	Disputes between squatters and landowners lodged in formal legal system. This variable is presented at the municipality-year level in the appendix of (LeGrand, 1986) originally found in National Archives of Colombia. Afterwards, I collapse the data into total amount of disputes per municipality for two periods, 1827-1900 and 1900-1936.	1827–1936	Own digitization from (LeGrand, 1986) (pp. 185–204)

Variable	Description	Coverage	Source
Dummy for *La Violencia* presence in Early Period	The first “Guzman Report" provides lists of every region of the country which experienced Early Period violence. I take the municipalities reported in this report and code each municipality as 1 if it is mentioned and 0 otherwise.	1948–1953 (1958 for Tolima)	Own digitization from (Guzmán et al., 1962) (pp. 117-140)
Count of Conflict Events (Clashes) during the Late Period	The second “Guzman Report" provides a list of month-municipality conflict events for the Late Period. I collapse the number of conflict events for each municipality for this period.	1958–1963	Own digitization from (Guzmán et al., 1964) (Section 3, Chapter 3, Appendix, Table III)
Dummy for *La Violencia* presence in Early Period	National Police statistics provide lists for presence of early-period violence in the 1959 volume. Municipalities are coded as 1 if mentioned and 0 otherwise.	1948–1953	Own digitization from (Policía Nacional, 1959) (pp. 23–26)
Count of violent deaths due to *La Violencia*	National Police statistics provide a comprehensive count of violent deaths caused by bandit or any organized armed groups for the entire Colombian territory in the 1962 volume. Some locations’ counts are hard to discern in printed version available in archives, thus for these locations, I complement the information from 1962 with a re-printing of the data in 1965 volume (pp.24-25). I take the count reported for each municipality.	1962	Own digitization from (Policía Nacional, 1962) (pp. 120–121)

Variable	Description	Coverage	Source
Number of bandit leaders	National Police statistics provide a comprehensive count of each bandit leader for the entire Colombian territory at the municipal level. I collapse the number of bandit leaders at the municipal level.	1962	Own digitization from (Policía Nacional de Colombia, 1965) (pp. 34–41)
Coffee Data			
Production of 62.5kg bags of coffee	National coffee censuses provide periodic municipal-level production estimates covering the entire Colombian territory. If a year does not provide direct production estimates, the figure is calculated indirectly by multiplying the number of coffee trees in a municipality by the departmental productivity per tree. The 1944 census year serves as the baseline measure.	1923, 1932, 1944, 1960	Own digitization from (Sociedad de Productores de Café and Sociedad de Agricultores de Colombia, 1923), (Monsalve, 1927), (Federación Nacional de Cafeteros de Colombia, 1933), (Federación Nacional de Cafeteros de Colombia, 1945), and (Departamento Administrativo Nacional de Estadística and Charry Lara, 1962)

Variable	Description	Coverage	Source
Number of coffee trees & coffee area	National coffee censuses provide municipal-level counts of coffee trees and area under cultivation. Used as alternative coffee measures in robustness checks.	1932, 1960	Own digitization from (Federación Nacional de Cafeteros de Colombia, 1933) and (Departamento Administrativo Nacional de Estadística and Charry Lara, 1962)
Competing Theories			
Municipal revenues	Municipal-level government budget data from official statistical reports (*Anuario Estadístico*). Proxy for pre-conflict state capacity.	1923	Shared by Diana Ricciulli-Marín and Mateo Uribe-Castro
State employees	Number of state employees at the municipal level, from the 1918 population census. Proxy for pre-conflict bureaucratic capacity.	1918	Own digitization from (Dirección Nacional de Estadística., 1923)
Political competition & Liberal vote share	From the 1946 presidential election: (i) Liberal vote share as the combined share of Turbay and Gaitán, and (ii) political competition as 1 − |Conservative share − Liberal share|. Liberal vote share in 1856 is also used as a deeper historical measure of partisan identity.	1856, 1946	Own digitization from official election results

Variable	Description	Coverage	Source
Land disputes (20th century)	Municipal-level counts of formal land disputes between squatters and landowners, covering 1900–1931.	1900–1931	Own digitization from (LeGrand, 1986) (pp. 185–204)
Land dispossession	Measure of municipal-level land dispossession, registered by local authorities as formal complaints from the General National Archives.	1914–1946	Public repository of (Lopez-Uribe and Sanchez Torres, 2024)
Historic, Demographic, and Geographic Controls			
Encomienda assignment	Dummy for whether a municipality had a colonial *encomienda*, a forced-labor institution assigning indigenous tributary populations to Spanish settlers. Digitized from a 1560 colonial survey (*visita de la tierra*) by Tomás López.	1560	(Faguet et al., 2024)
Colonial state presence	Dummy for the presence of Crown employees and fiscal institutions (state monopolies, consumption taxes, postal service) at the municipal level. Originally compiled by (Durán y Díaz, 2012), digitized by (García Jimeno, 2005).	1794	(García Jimeno, 2005)
Municipal population counts	Municipal-level population counts for the universe of Colombian municipalities. Raw, unadjusted population figures reported in each census year.	1851, 1905, 1938	Own digitization from national population censuses
Agricultural employment share	Share of municipal population employed in agriculture, computed from census occupational data.	1938	Own digitization from (Dirección General de los Censos., 1942)

Variable	Description	Coverage	Source
Geographic characteristics	Altitude, distance to departmental capital, and latitude/longitude. Derived from the CEDE municipal dataset.	2023	CEDE (Universidad de los Andes)
Terrain ruggedness index	Municipal-level measure of terrain ruggedness.	2012	(Nunn and Puga, 2012)

A.3.2 Measurement Error

Given that conflict statistics are inherently a hard thing to measure, measurement error in the intensity of violence might be a large concern. In order to gage the extent to which the data I use in the intensity of violence sections might be afflicted by this, I try to gage how much different sources align as to the trends and locations of the highest relative violence. First, I do this at the higher level of aggregation of departments, and then at the municipal level.

Department-Level Trends. First, we have already seen in Figure 1.4, that estimates at the national level (from demographic study (Romero-Prieto and Meisel-Roca, 2019)) track trends at the department and new frontier levels (from archival source (Gutiérrez Tovar, 1961)) extremely closely. However, there might be a large disconnect between these sources and the municipal level sources used as outcomes in the empirical section.

In order to test this, I use the panel aspect of the 1958-1963 conflict event dataset from the Guzman Reports (Guzmán et al., 1964), which permits me to decompose the amount of events per year. This allows me to see if the trends exhibited in the (Gutiérrez Tovar, 1961) are also present in the municipal level data. I aggregate this municipal data to the frontier level and plot both sources in Figure A.11 which plots deaths (solid line, left axis) against clash events (dashed line, right axis). Both series rise and fall in tandem, indicating that the broad geographic pattern of violence is consistent across sources. Figure A.12 disaggregates this comparison across seven major departments (Tolima, Caldas, Valle, Antioquia, Cundinamarca, Santander, and Norte de Santander) and shows that the co-movement holds within each department rather than being an artifact of aggregation. The only department with very different trends is Norte de Santander, the least violent department in the sample.

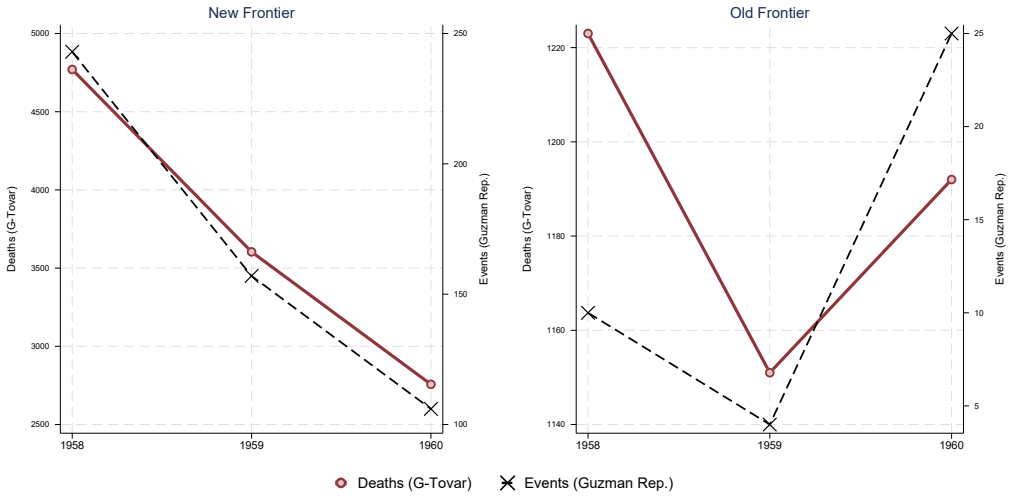


Figure A.11. Violence Trends by Frontier Type: Deaths vs. Clashes

Notes: Solid line = intentional deaths from (Gutiérrez Tovar, 1961) (left axis). Dashed line = clash events from Guzman Reports (right axis). Left panel: new frontier departments (Antioquia, Caldas, Tolima, Valle). Right panel: old frontier departments (Cundinamarca, Santander, Norte de Santander).

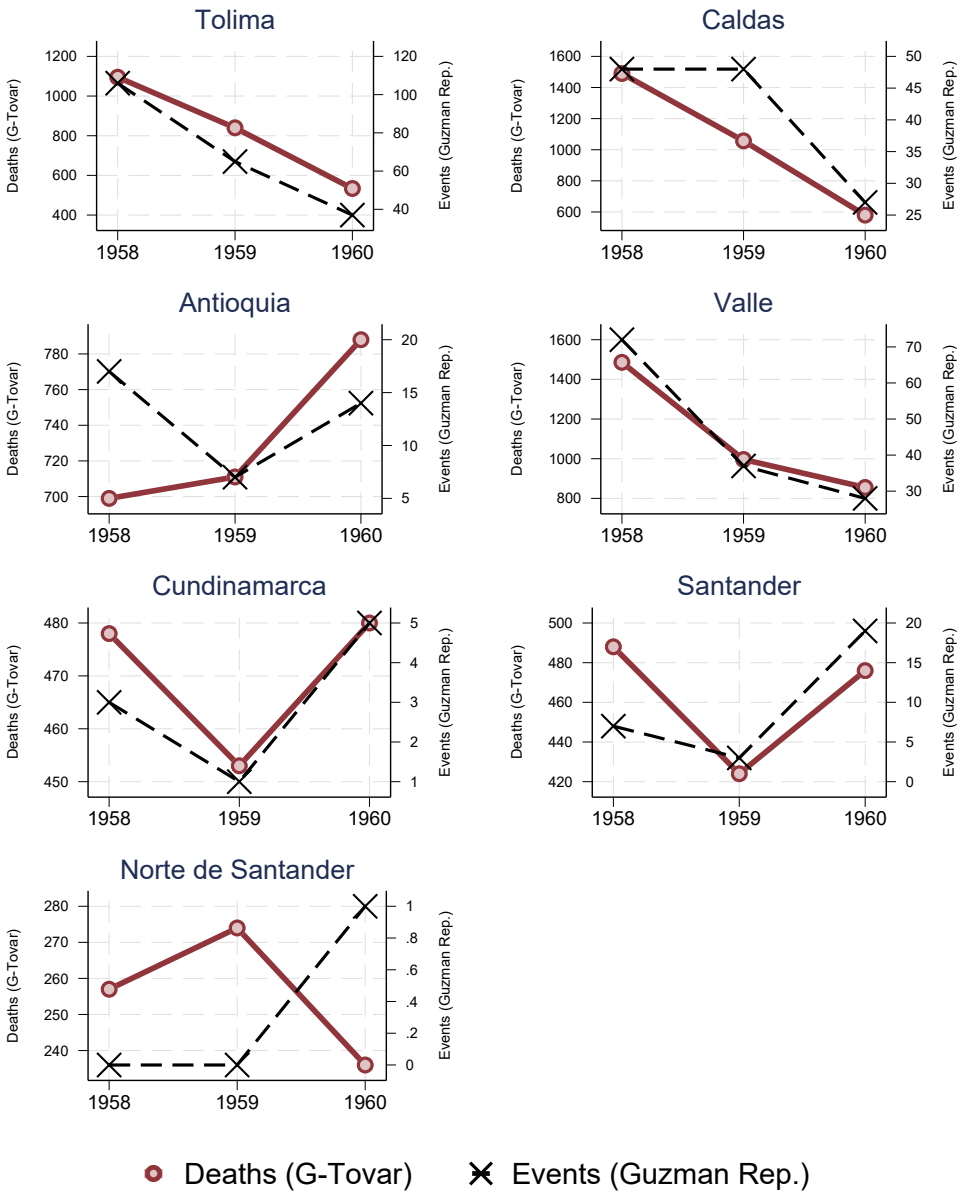


Figure A.12. Violence Trends by Department: Deaths vs. Clashes

Notes: Each panel shows one department. Solid line = intentional deaths from (Gutiérrez Tovar, 1961) (left axis). Dashed line = clash events from Guzman Reports (Guzmán et al., 1964) (right axis).

Municipality-Level Concordance. While the department level comparison assuages concerns measurement error concerns at regional level, the unit of analysis in the empirical analysis is the municipal level. Hence, if within departments, municipalities exhibit a large reshuffling of relative intensity, it is possible the empirical results could be spurious. In order to examine this possibility, I focus on the two main sources of mid-twentieth century violence intensity (police reports from 1962 and the Guzman Reports clash events). These two data measure different dimensions of conflict over slightly different periods. If they capture the same underlying phenomenon, municipalities ranked as highly violent by one source should also rank highly by the other. This subsection validates that the two sources agree on which locations experienced the most violence.

At the municipal level, the Spearman rank correlation between the two sources is $\rho = 0.639$ ($p < 0.001$), indicating that municipalities are ranked similarly by both measures despite differences in timing and data collection. Table A.23 reports overlap statistics for the tails of the distribution: among the top decile of death-ranked municipalities, 62.3% also fall in the top decile of clash-ranked municipalities; the reverse overlap is 64.7%. Restricting to municipalities with positive values, the top-decile overlap rises to 59.1% (deaths to clashes) and 86.7% (clashes to deaths). Figure A.13 plots the two measures against each other at the municipal level, confirming the positive relationship. Given that these are two independently compiled, historical sources the fact that both match so well assuages concerns of severe measurement error as to the relative positioning of municipalities with respect to violence intensity.

Table A.23. Source Concordance: Police Reports (1962) vs. Guzman reports Clashes (1958–63)

Statistic	Value
Rank correlation	
Spearman ρ	0.639***
Top decile overlap — all municipalities (N = 651)	
Deaths top 10% also in clashes top 10%	33 of 53 (62.3%)
Clashes top 10% also in deaths top 10%	33 of 51 (64.7%)
Top decile overlap — positive values only	
Deaths top 10% also in clashes top 10%	13 of 22 (59.1%)
Clashes top 10% also in deaths top 10%	13 of 15 (86.7%)
N (deaths > 0)	220
N (clashes > 0)	150

Notes: Deaths = violent deaths in 1962 from police criminalidad reports. Clashes = violent events 1958–63 from Guzman reports (Fals Borda, Umaña Luna et al.). Spearman rank correlation on non-missing pairs (N = 651). All-municipalities panel: top decile defined as strictly above the 90th percentile of each source. Positive-only panel: deciles computed on municipalities with > 0 values for each source separately. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.



Figure A.13. Source Concordance: Deaths (1962) vs. Clashes (1958–63)

Notes: Each point is a municipality. Horizontal axis: violent deaths in 1962 (Criminality Reports). Vertical axis: total clash events 1958–63 (Guzman Reports). OLS fitted line shown. Sample restricted to municipalities with at least one positive value in either source.

A.3.3 Historical Geographic Units

Colombian municipalities have undergone frequent reorganization since the late nineteenth century, primarily through the subdivision of existing entities into smaller ones. This poses a fundamental challenge for longitudinal analysis: the geographic unit recorded in one census may not correspond to the unit recorded in the next. Three types of changes can occur between any two periods:

1. **Name changes**: A municipality changes its name, so records from different periods cannot be linked by name alone.
2. **Splits**: A municipality X at time t is divided into X and Y at time $t+1$. Comparing X across periods is invalid because X at $t+1$ covers a smaller area than X at t.
3. **Mergers**: Two municipalities are consolidated into one, creating the reverse problem.

To address this, I construct a *crosswalk*, a mapping that links each municipality in one period to its counterpart(s) in the adjacent period, explicitly recording every name change, split, and merger. Following the approach described by (Eckert et al., 2020) for (Horan and Hargis, 1995), the crosswalk aggregates recently partitioned municipalities back into larger *Historical Geographic Units* (HGUs) that can be compared across time. When a municipality recorded in a later census is identified as having split from a parent municipality, it is merged back into the parent for earlier periods. The process is iterated backwards: HGUs expand in size the further back in time the analysis goes, but spatial consistency is preserved.

For T data periods, $T-1$ crosswalks must be created, each linking adjacent periods. Once the full chain is established, any two periods can be compared on consistent geographic units.

Why HGUs matter: the problem of splits. The most consequential boundary change is the split. Figure A.14 illustrates the problem and its solution using a real example from this paper's data. In 1950, the municipality of El Águila (Valle del Cauca) split off from Ansermanuevo. Before the split, in 1944, Ansermanuevo produced 150,000 sacks of coffee, one of the highest figures in the department. After the split, in 1962, the police recorded 22 violent deaths in Ansermanuevo and 14 in El Águila. The 1960 agricultural census reveals that El Águila had 5,476 hectares of coffee land compared to Ansermanuevo's 6,375, meaning on may safely assume El Águila accounted for nearly half of the original municipality's coffee area 15 years before.

The diagram highlights two approaches. The naive approach matches municipalities by name across periods. This links Ansermanuevo in 1944 to Ansermanuevo in 1962, but El Águila (which did not exist in 1944) receives zero coffee production. A regression of violence on coffee would find no relationship for El Águila, even though it inherited nearly half of the original coffee area. The HGU approach instead aggregates El Águila and Ansermanuevo back

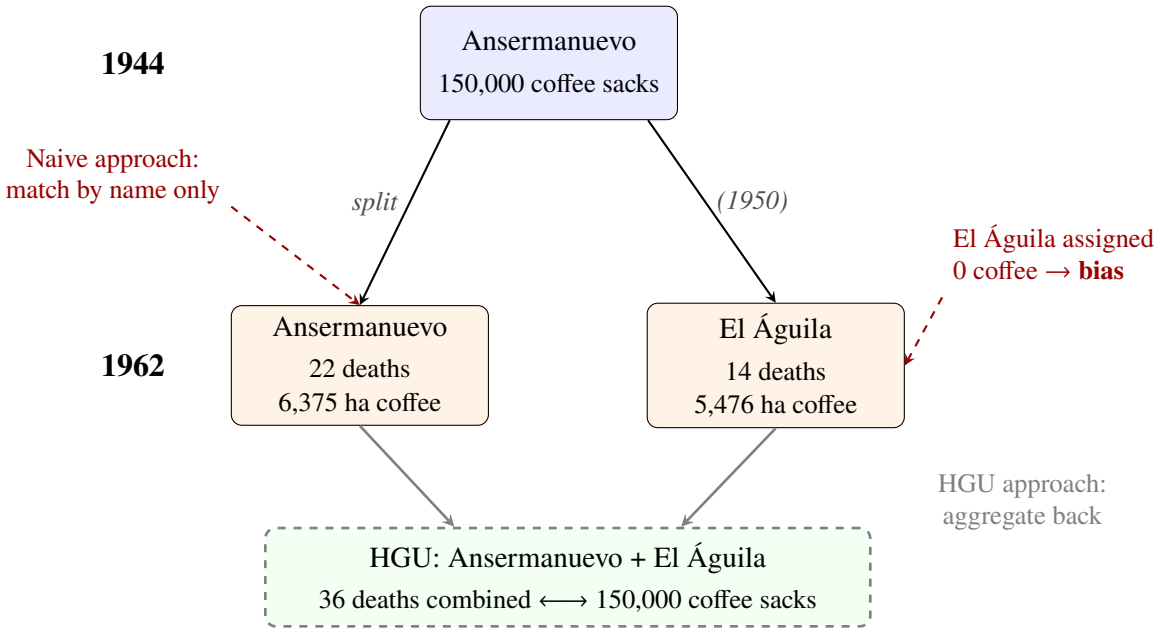


Figure A.14. HGU aggregation: the case of El Águila splitting from Ansermanuevo

into a single unit for all periods after the split. The combined unit records 36 deaths in 1962 and 150,000 sacks in 1944, preserving the correct geographic comparison without requiring any assumption about how coffee production was divided between the two municipalities after separation.

Sources. The construction of HGUs relies on triangulating information across multiple historical sources. Each entry in the lists below is tagged with one or more of the following source codes:

[P] 1918 Population Census: provides a universe of all municipalities and their *corregimientos*, smaller administrative entities that are not independent municipalities. When a corregimiento shares its name with a future municipality, this indicates a likely split from the parent.

[C] 1924 Coffee Census (Monsalve, 1927): contains similar municipality and sub-entity lists, allowing cross-validation of splits.

[W] Wikipedia: modern municipality pages were systematically scraped for mentions of administrative splits, parent municipalities, and founding dates.

[I] IGAC *Diccionario Geográfico*: the Instituto Geográfico Agustín Codazzi provides local histories for many municipalities. This source is particularly useful for geolocating battle sites from the Thousand Days War, which historical sources record by geographical location rather than municipality.

[D] DANE: from 1962 onward, the national statistical office provides an official universe of municipal splits and name changes.

[O] Other: triangulation from online search, comparison of modern and historical municipal maps, and geographic analysis of locations where new municipalities arose.

The sections that follow present (i) the ordered list of all documented name changes, (ii) the list of splits and new municipal creations, and (iii) the list of mergers.

List of name changes.

I separate name changes into two groups. The pre-1900 entries correspond primarily to the Thousand Days War. Many of the battle locations listed in (Jaramillo Castillo, 1991) are recorded as geographical place names (rivers, haciendas, mountain passes) rather than municipalities. These entries therefore represent geolocalization (mapping a historical place name to the municipality that contains it), while others are actual name changes. The post-1900 entries are all literal changes in official municipal names.

List of name changes: Pre-Thousand Days War

1. **Lumbi → Mariquita.** [O]
2. **Camala (Tolima) → Flandes.** [W]
3. **Jegua → municipality in San Benito Abad.** [O]
4. **Río Recio → Ambalema.** [W]
5. **Pela Huevos (Tolima) → Rovira.** [O]
6. **Macueque → Sácama.** [O]
7. **Potosí (Boyacá) → Gameza.**
8. **San Jorge (Tolima) → Ibagué.** [O]
9. **Calibío (batalla de) → Popayán.** [W,P]
10. **Papayal (batalla de) → Palmira (Valle).** [W]
11. **La Dorada (Antioquia) → Caldas.**
12. **Aguas Vivas (Bolívar) → Montería (Córdoba).** [W]
13. **Colorados (Bolívar) → San Juan Nepomuceno.** [W]
14. **Coloso → Ricaurte.** [W]
15. **El Desconsolado (Bolívar) → San Pedro Consolado (San Juan Nepomuceno).** [W]
16. **Santa Cruz (Bolívar) → Lorica (Córdoba).**
17. **Carmen (Boyacá) → Carmen de Carupa.**
18. **El Puerto (Boyacá) → Puerto Boyacá.**
19. **Guadualón → Yacopí (Cundinamarca).** [O,I]
20. **Lagunaseca (Boyacá) → Laguna de Marcura (Paipa).** [O]
21. **Medina (Boyacá).**
22. **Anaime/Amaime (Cauca) → Palmira or Cerrito (assign Palmira).** [P]
23. **Rioblanco (Cauca) → Rioblanco (Tolima).**
24. **Cumaca (Cundinamarca) → Tibacuy.**
25. **El Hato (Cundinamarca) → Carmen de Carupa.** [O]
26. **La Virgen (Cundinamarca) → Quipile.** [O]
27. **Las Abras de Une → Une.**
28. **Socotá → Boyacá (from Cundinamarca).**

29. **San Miguel de Sema → San Miguel (from Cundinamarca).**
30. **Jegua → corregimiento of San Benito Abad.**
31. **Momil / Montería (Magdalena) → Córdoba / Bolívar.**
32. **Cerro de Girón → Girón.**
33. **La Palma (Santander) → Hacarí.**
34. **Anaime (Tolima) → Cajamarca.**
35. **Miraflores (Tolima) → Rovira.**
36. **Guaduas (Tolima) → Cundinamarca.**
37. **San Miguel (Tolima) → Cajamarca.**
38. **Guayabal → Guayabal de Síquima (Cundinamarca).**
39. **Chita (Cundinamarca) → Boyacá.**
40. **Vianicito → Vianí.**
41. **Macueque → Sácama (Boyacá).** [I]
42. **Mesa Grande (Boyacá) → Guayabetal (Cundinamarca).** [I,O]
43. **Quipe (Boyacá) → Caldas (Boyacá).** [I]
44. **Las Huertas (Bolívar) → corregimiento of Sincelejo.** [P]
45. **El Piñal (Bolívar) → Espinal (Cartagena).** [P]
46. **Playones (Magdalena) → Sitionuevo.** [I]
47. **Flor de Monte (Bolívar) → Ovejas.** [P]
48. **Juan Gordo (Bolívar) → Since (Bolívar).** [P]
49. **La Roya (Bolívar) → La Raya (Majagual).** [P]
50. **Riofrío (Bolívar) → Ciénaga (Magdalena).** [P]
51. **Hatonuevo → Barrancas (Magdalena).** [P]
52. **Santana (Cauca) → Miranda or Cartago (choose Miranda).**
53. **El Trapiche (Cauca) → San Miguel (La Vega).** [O]
54. **Las Delicias (Cauca) → El Rosario (Nariño).** [P]
55. **Los Chancos (Cauca) → Andalucía (Valle).** [P]
56. **Pitayo (Cauca) → Silvia (Cauca).** [P]
57. **Bazán (Cauca) → Buenaventura (Valle).** [I]
58. **Bocas de Guengue → Puerto Tejada.** [I]
59. **Bocas de la Paila → La Paila (Zarzal, Valle).** [O]
60. **El Morro → Playa del Morro (Tumaco, Nariño).** [O]
61. **La Monigua = Jardín Botánico Manigua → Buenaventura.** [O]
62. **Tres Tulpas → Córdoba (Nariño).** [O]
63. **Simancas (Cauca) → Guachucal (Nariño).** [I]
64. **Bosavita (Cundinamarca) → Villapinzón.** [O]
65. **Cambao (Cundinamarca) → San Juan de Río Seco.**
66. **Chorrillo (Ambalema, Tolima).** [P]
67. **La Florida (Cundinamarca) → Anolaima.** [P]
68. **La Virginia (Cundinamarca) → Tocaima.** [P]
69. **Alto de la Cruz → Arbelaez.** [O]

70. **Alto de Minipi → La Peña / Nimaima / La Palma.** [O]
71. **Alto del Trigo → Guaduas.** [O]
72. **Alto las Tibayes → Guaduas.** [O]
73. **Cerro del Muerto → Melgar (Tolima).** [I]
74. **Amoladero (Cundinamarca) → Guatavita.** [I]
75. **Boquerón → Tibacuy.** [O]
76. **Guavio → Gachetá.** [I,O]
77. **Guiña → Macheta.** [I,W]
78. **Hacienda Liberia → Viotá.** [O]
79. **Izacar → La Palma.** [W]
80. **La Cabuya → Cáqueza.** [W]
81. **La Palestina → Pulí.** [W]
82. **Agualarga → Albán.** [W]
83. **Montefrío → Chaguaní.** [W]
84. **Novillero → Fusagasugá.** [I,O]
85. **Portachuelo → Zipaquirá.** [I,O]
86. **La Aguadita → Fusagasugá.** [I,O]
87. **San Nicolás → San Juan de Río Seco.** [I,O]
88. **Sibaté → Soacha.** [W]
89. **Cerro Blanco → Ciénaga (Magdalena) → Zona Bananera.** [I]
90. **Chivolo → Tenerife.** [O]
91. **Los Obispos (Magdalena) → Morales (Bolívar) / Olaya Herrera.** [I]
92. **Santa Catalina → Bolívar (change department).** [O]
93. **Chicamocha → Aratoca.**
94. **Alto La Paja → Rionegro (Santander).** [O]
95. **Vetas (Santander) → California.** [P]
96. **Girón → Lebrija.** [O]
97. **Helechales → Floridablanca.** [I,O]
98. **Palenquillo (Santander) → Gamarra (Magdalena).** [I,O]
99. **Palonegro → Lebrija.** [W]
100. **Carazúa → Riohacha.** [W]
101. **Peralonso → Cúcuta.** [W]
102. **Mamarruca → Suaita.** [W]
103. **San Jorge → Rionegro.** [O]
104. **Santurbán → Salazar de las Palmas.** [O]
105. **Four Santander towns (Palau) → Rionegro + Girón.**
106. **Amoya/Amaya → Chaparral.** [O,I]
107. **Ata → Río Ata (Ataco/Planadas).**
108. **Bermejo → Dolores.** [I]
109. **Chicoral → Espinal.** [W]
110. **Corozal (Tolima) → Palermo (Huila).** [O]

111. **Doima → Piedras (Tolima).** [P]
112. **Garrapata → Mariquita.** [O]
113. **Tetuán → San Antonio.** [W]
114. **Santa Lucía → Ambalema.** [O]
115. **Rinconsanto → Guamo.** [O,I]
116. **Río Páez → Tesalia.** [O]
117. **Río Guarino → Mariquita.** [O]
118. **Pescaderías → Coello.** [O]
119. **Patio Bonito + La Florida → Fresno.** [O]
120. **La Vieja → Villavieja (Huila).**
121. **La Sierra → Lérida.** [O]
122. **Hotel Mi Casa → Gualanday (Coello).** [O]
123. **Matamundo → Neiva.** [O]
124. **Alto del Bledo → Lérida.** [W]
125. **Ambato → Espinal.** [O]
126. **La Morada → Alpujarra.** [O]
127. **El Real → Natagaima.** [O]
128. **Camala → Flandes.** [W]

List of name changes: Post-Thousand Days War

1. **Albán (Cundinamarca) → Chimbe.** [I]
2. **Anaime → San Miguel de Perdomo → Cajamarca / Quezada / Herrera.** [I]
3. **Aspasica → La Playa.** [O,W]
4. **Baudo (Chocó) → Pizarro.**
5. **Belén de Umbría (Caldas) → Mocatán → Belén de Umbría.**
6. **Berruecos (Nariño) → Arboleda.** [W]
7. **Briceño (Tolima) → Anzoátegui.**
8. **Caldas (Tolima) → Alvarado.**
9. **Carlosama (Nariño) → Cuaspud.** [W]
10. **Carnicerías → Tesalia.**
11. **Cincelada (Santander) → Coromoro.** [W]
12. **Cité → Barbosa, Santander.**
13. **Concepción (Huila, 1905/1912) → Acevedo.**
14. **Concordia (Norte de Santander) → Ragonvalia.** [O]
15. **Córdoba (1912, Norte de Santander) → Durania.**
16. **Dolores (Cauca) → Rosas.**
17. **El Retiro → Teruel.**
18. **El Rosal → San Sebastián.** [P]
19. **Espejuelo → Miranda.**
20. **Espíritu Santo → Codazzi → Robles → La Paz.** [O,W]
21. **Guagua (Huila) → Palermo.**

22. **Guagua → Palermo.**
23. **Guayabal (Tolima) → San Lorenzo → Armero.** [W]
24. **Hato (Huila) → Tarqui.** [O,I]
25. **Hato (Huila) → Toa → Tarqui.** [O]
26. **Huasano (Valle) → Riofrío.** [I,C]
27. **Iscuandé → Santa Bárbara de Iscuandé.**
28. **La Capilla de Cocuy / La Capilla (Boyacá) → San Mateo.**
29. **La Cruz (Norte de Santander) → Ábrego.** [W]
30. **La Jagua → corregimiento of Garzón.**
31. **La Palma (Norte de Santander) → Hacarí.**
32. **La Paz (Boyacá) → Paz del Río.**
33. **La Santísima Trinidad/La Trinidad (Boyacá) → Almeida.**
34. **La Unión (Valle del Cauca) → Lemos → La Unión.** [W]
35. **Manare → Hato Corozal.** [O]
36. **Margento (Antioquia) → Caucasia.** [W,I]
37. **María → Villamaría.**
38. **Marroquín → El Morro → Yopal.** [O]
39. **Miraflores (Tolima) → Rovira.** [W]
40. **Morales (Bolívar, 1912) → Bodega Central → Olaya Herrera → Morales.** [W]
41. **Moreno (Casanare) → Paz de Ariporo.** [I]
42. **Moreno (Casanare) → Paz de Ariporo.** [I]
43. **Nazareth → Quinchía.** [P]
44. **Olaya Herrera → Morales.**
45. **Ospina (Cundinamarca) → Venecia.**
46. **Pavarandocito → Mutata.**
47. **Pavas (Valle) → La Cumbre.** [I,W]
48. **Pinillos (Bolívar, 1912) → Palomino → Pinillos → Palomino → Pinillos (multiple changes).**
49. **Pueblo Viejo (Boyacá) → Aquitania.**
50. **Purísima (Bolívar) → Momil → Purísima / Momil split.** [C,W,P]
51. **Rafael Reyes (Cundinamarca) → Apulo.**
52. **Restrepo (Valle) → Conto → Restrepo.**
53. **Ricaurte (Bolívar) → Colosó.**
54. **Riosucio (Antioquia, 1905) → Murindó.**
55. **Robles (Magdalena) → La Paz.** [I]
56. **Salado → Dagua.** [W,O]
57. **Salazar (Santander) → Salazar de Las Palmas.**
58. **San Agustín (Caldas) → Samaná.**
59. **San Eduardo (Boyacá) = Berbeo.** [P]
60. **San Francisco (Caldas) → Chinchiná.** [W]
61. **San Joaquín (Caldas) → Risaralda.** [I]
62. **San José (Nariño) → Roberto Payán.**

63. **San Mateo (Huila) → Rivera.**
64. **San Miguel (Cauca, 1918) → López de Micay.**
65. **San Miguel de Perdomo → Cajamarca.**
66. **San Pedro (Santander) → Villacaro.** [I]
67. **San Rafael (Boyacá) → Rondón.**
68. **San Rafael → Rondón.**
69. **San Vicente (Valle) → Andalucía.** [I,W]
70. **San Zenón → Buenavista → San Zenón → Buenavista segregates.** [O]
71. **Santa Elena (appears 1892/1905) → deleted.**
72. **Santa Librada (Huila) → Suaza.** [W,I]
73. **Santa Rosa (Cundinamarca/Tolima, 1905) → Suárez.**
74. **Santana → Espejuelo.**
75. **Santana (Tolima / Santa Ana) → Falan.**
76. **Santuario (Caldas) → Tatamá → Santuario.** [I]
77. **Segovia (Caldas) → Marsella.**
78. **Soledad (Tolima) → Herveo.** [O,W]
79. **Sucre (Antioquia) → Olaya.**
80. **Sucre (Putumayo) → Colón.**
81. **Tequendama (Cundinamarca) → San Antonio.**
82. **Tequía (Santander) → San José de Miranda.** [W]
83. **Tetón → Córdoba (Bolívar, 1912).**
84. **Tunia (Cauca) → Piendamó.**
85. **Unión (Huila) → Tello.** [W]
86. **Ventaquemada → Padua → Ventaquemada.** [W,P]
87. **Villa Caro → Bucarasica → Villa Caro.**
88. **Zapatosa (Boyacá) → Aguazul.** [W]

List of splits

Determining exact split years is difficult, as splits are identified when merging adjacent data periods. The groupings below reflect the decade in which each split was identified with high confidence.

Pre-1900

1. **Armenia Mantequilla (1894) → split from Heliconia.**
2. **Gómez Plata (1896) → was a corregimiento of Carolina.**
3. **Salgar (1894) → split from Concordia.**
4. **Albán (Cundinamarca) → split from Guayabal.** [I]
5. **Peñón → under jurisdiction of Pacho since 1856, separated in 1903.**

1900s–1910s

1. **La Unión → split from La Ceja.**

2. **Pueblorrico → split from Jericó.**
3. **Piojó → split from Santo Tomás.** [I]
4. **Repelón → split from Sabanalarga.**
5. **San Carlos (Bolívar) → split from Ciénaga de Oro.** [W]
6. **San Marcos (Bolívar) → split from Caimito.** [W]
7. **Belén de Umbría (Caldas) → split from Anserma.** [W]
8. **Circasia (1906) → split from Filandia.**
9. **Montenegro → split from Circasia (but originally from Filandia).**
10. **San Agustín (will later become Samaná) → split from Pensilvania.** [W]
11. **Santuario (Caldas) → split from Apía.** [W]
12. **Aldana (Nariño, 1911) → split from Cuaspud/Carlosama.**
13. **Córdoba (Nariño, 1911) → split from Puerres.**
14. **La Palma (Norte de Santander, 1908) → split from Aspasica.**
15. **Santa Isabel (1912) → split from Venadillo.**
16. **Alejandría (1907) → split from Guatapé.** [W]
17. **Caicedo (1908) → split from Urrao.** [O]
18. **Venecia (1909) → split from Fredonia.** [O]
19. **La Trinidad (1907) → split from Somondoco (this becomes Almeida).** [O]
20. **Herran (1911) → split from Concordia (Norte de Santander).** [O]

1910s–1920s

1. **Abriaquí → split from Frontino.** [O]
2. **Chigorodó → split from Turbo.**
3. **Montebello → split from La Ceja.**
4. **Nariño (Antioquia) → split from Sonsón.** [O]
5. **Peque → split from Ituango.** [O]
6. **Toledo → split from San Andrés.** [W]
7. **Valdivia → split from Yarumal.** [O]
8. **San Andrés y Providencia → split from Bolívar (and into San Andrés + Providencia).** [W]
9. **Malambo → split from Soledad (Atlántico).** [O]
10. **Polonuevo → split from Baranoa (Atlántico).** [W]
11. **Berbeo → split from Miraflores (Boyacá).** [W]
12. **Palestina → split from Chinchiná (then called San Francisco).** [O]
13. **Pueblo Rico (Caldas) → split from Pueblorrico (Chocó).** [O]
14. **San Joaquín (future Risaralda/Caldas) → split from Anserma.** [O]
15. **La Sierra → split from Rosas (called Dolores at this time).** [O]
16. **Puerto Tejada → split from Caloto (Cauca).** [O]
17. **San Miguel (Cauca) → split from Timbiquí.** [O]
18. **Aracataca → split from Pueblo Viejo.** [W]
19. **Santa Cruz → split from Samaniego (Nariño).** [I]
20. **Pamplonita → split from Pamplona.** [I]

21. **Briceño (later Anzoátegui) → corregimiento of Alvarado (then named Caldas).** [O]
22. **Icononzo → split from Melgar.** [O]
23. **San Antonio (Tolima) → corregimiento of Chaparral.** [W]
24. **Sevilla (Valle) → corregimiento of Bugalagrande.** [W]
25. **Betania (Antioquia, 1922) → split from Bolívar (Antioquia) and Andes.** [W]
26. **Cumaca (Cundinamarca) → corregimiento split from Tibacuy.** [C]
27. **Topaipí → corregimiento split from El Peñón.** [C]

1920s–1930s

1. **Cisneros (Antioquia) → split from Santo Domingo.** [W]
2. **Balboa (Caldas, 1923) → split from Santuario.** [I]
3. **Quimbaya (1923) → split from Filandia.** [I]
4. **Colón (Nariño, 1921) → split from San Pablo.** [O]
5. **Barrancabermeja (Santander, 1922) → split from San Vicente.** [I]
6. **Alcalá (1919) → corregimiento split from Cartago.** [O]
7. **Guática → split from Anserma.** [O]
8. **Ansermanuevo (Valle) → split from Cartago.** [I,P]
9. **Caicedonia (Valle) → split from Sevilla.** [W,P]
10. **Restrepo (Valle, 1924) → split from Vijes.** [W,P]
11. **San Bernardo → split from Pandi.** [P]
12. **Obando (Caldas, 1928) → split from La Victoria.** [I,W,C]
13. **Trujillo (Valle) → split from Riofrío (via Huasano).** [C]
14. **Ulloa (also called Villa Sucre) → split from Alcalá.** [I,C]
15. **Mistrató (Caldas) → split from Belén de Umbría.** [I,W,C]
16. **Pijao (1932 census) → split from Calarcá.** [I,P]
17. **Marquetalia (Caldas) → split from Manzanares.** [P,I,W]
18. **San Agustín (Huila, 1930) → split from Pitalito.** [W,I,C]
19. **San Juanito (Huila, 1930) → split from Campoalegre.** [I,C]
20. **Lourdes (Norte de Santander) → split from Gramalote.** [I,C]
21. **Charta (Santander, 1930) → split from Matanza.** [C]

1930s–1940s

1. **Tarso (Antioquia) → split from Jericó.** [W,C]
2. **Achí (Bolívar, 1934) → split from Majagual.** [W,C]
3. **Génova (Caldas, 1937) → split from Pijao.** [O]
4. **La Dorada (Caldas) → split from Victoria (in 1923).** [W,P]
5. **Puerto Salgar (called Liévano in 1938) → split from Caparrapí.** [I]
6. **Gamarra (Magdalena) → split from Aguachica.** [P]
7. **María La Baja (Bolívar) → split from San Juan Nepomuceno.** [C,P]
8. **Rivera (Huila, 1943) → split from Neiva.** [I,W]

9. **Maceo (Antioquia, 1943) → split from Puerto Berrío.** [I,W]

1940s–1950s

1. **San Bernardo (Bolívar, 1945) → split from Lorica.** [I,W]
2. **San Pedro (Bolívar / Sucre, 1939) → split from Sincelejo.** [C]
3. **Tierralta (Bolívar, 1948) → split from Montería.** [I,W]
4. **Támara (Casanare) → split from Pore.** [I]
5. **San José del Palmar (Chocó, 1948) → split from Nóvita.** [I]
6. **Rafael Reyes (Cundinamarca) → split from Anapoima.** [W]
7. **Silvania (Cundinamarca, 1940s) → split from Fusagasugá.** [I]
8. **Saladoblanco (Huila) → split from Elías.** [I]
9. **Fundación (Magdalena) → split from Aracataca.** [I]
10. **Acacías (Meta) → split from San Martín.**
11. **Roncesvalles (Tolima) → split from San Antonio.**
12. **Chaparral (Tolima) → large population drop 1938→1951, likely due to Río Blanco split.** [O]
13. **Villarica (Tolima, 1949) → split from Cunday.** [O]
14. **El Águila (Valle, 1950) → split from Ansermanuevo.** [I]
15. **Calima (Valle, 1949) → split from Yotoco.** [I]
16. **Ginebra (Valle) → split from Guacarí.** [O]
17. **El Cairo (Valle) → split from Versalles.** [O]
18. **Ospina (Cundinamarca) → split from Pandi.** [W]
19. **Montelíbano (1954) → split from Ayapel (Bolívar).** [D,I]
20. **San Luis de Gaceno (Boyacá, 1954) → split from Campohermoso.** [D]
21. **Tauramena → split from Aguazul (formerly Zapatosa).** [W]
22. **Córdoba (Quindío) → split from Calarcá.** [D,I]
23. **La Celia (Risaralda) → split from Santuario.** [D,I]
24. **La Tebaida (Caldas/Quindío) → split from Armenia.** [D,W]
25. **Viterbo (Caldas) → split from Apía.** [I]
26. **La Virginia (Risaralda) → split from Belalcázar.** [D,W]
27. **Suárez (Cauca) → split from Buenos Aires.** [D,I]
28. **Bojayá (Chocó) → split from Quibdó.** [D]
29. **Cabrera (Cundinamarca, 1963) → split from Pandi.** [D,I]
30. **La Argentina (Huila) → split from La Plata.** [I,D]
31. **Cubarral (Meta, 1960) → split from San Martín.** [I]
32. **El Zulia (Norte de Santander) → split from Cúcuta.** [D]
33. **El Guacamayo (Santander) → split from Aguada.** [D,I]
34. **Rioblanco (Tolima) → split from Chaparral.** [W]
35. **Argelia (Valle) → split from Toro.** [D]
36. **El Dovio (Valle) → split from Roldanillo.** [D,I]

1950s–1960s

1. **Arboletes (Antioquia, 1959) → split from Turbo.** [D,I]
2. **Yalí (Antioquia, 1959) → split from Yolombó.** [D,I]
3. **Luruaco (Atlántico, 1959) → split from Luruaco.** [D,I]
4. **Planeta Rica (Córdoba) → split from Montería.**
5. **Pueblonuevo (Córdoba) → split from Sahagún.**
6. **Flandes (Tolima) → split from Espinal.** [I,D]
7. **Valencia (Córdoba) → split from Tierralta.** [I,D]
8. **Castilla La Nueva (Meta, 1961) → split from Guamal.** [I,D]
9. **Fuente de Oro (Meta) → split from San Martín.** [I]
10. **Cumaral (Meta) → split from Restrepo.** [D]
11. **San Sebastián de Buenavista (Magdalena) → split from San Zenón.** [D]
12. **San Miguel de Sema (Boyacá) → split from Chiquinquirá.** [I,D]
13. **Argelia (Antioquia) → split from Sonsón.** [I,D]
14. **Santa María (Boyacá) → split from Macanal.** [D,I]
15. **Alto Baudó (Chocó) → split from Baudó (Bajo Baudó).** [I]
16. **Puerto Escondido (Córdoba) → split from Cereté.** [D]

List of Merges

Municipal merges represent the opposite problem of splits: municipalities that existed independently for many years suddenly disappear from the administrative structure. In a merge, historical municipalities lose administrative autonomy and become integrated into another municipality, meaning we must assign their population, coffee production, or other characteristics to the absorbing municipality. In order to simplify the process, all values are merged from the earliest year possible (i.e. when both municipalities first exist), instead of the "merging year".

Merges — individual reasoning

1. **Naranjal → merged into Timaná (Huila).** [O,P]
2. **Zea → merged into Anorí in 1905.**
3. **San José → merged into Cúcuta between 1905–1912.** Based on populations: San José (15k) + Cúcuta (2k) → Cúcuta (20k) in 1912.
4. **La Paz → merged into Guaduas.** [O,W]
5. **El Yucal → merged into Calamar (Bolívar).** [W]
6. **Sibundoy + San Francisco → merged into Mocoa (1918).**
7. **Cumaca → merged into Tibacuy (1922).** [C]
8. **San Bernardo → merged into Pandi pre-1924, then separated again (1932). This is because these two municipalities start off un-merged, then merge, and un-merge again, so I assume they are merged until final split in 1932**
9. **Albán (Valle) → merged into Versalles (1924).** [P]
10. **Naya → merged into Buenaventura.**
11. **La Jagua → merged into Garzón.**

12. **Tame (San Lope) → merged into Arauca.** [W]
13. **Puerto Boyacá + Otanche + Borbur → merged into Territorio Vásquez.**
14. **Bogotá metropolitan merge (historical): Bosa + Engativá + Fontibón + Suba + Usaquén + Usme → Bogotá.**
15. **Umpalá → merged into Piedecuesta (Santander).** [W]

References

Arango, M. (1977). *Café e Industria, 1850-1930*. C. Valencia Editores.

Becker, S. O., P. D. Boll, and H.-J. Voth (2025). Spatial Unit Roots in Regressions: A Practitioner's Guide and a Stata Package.

Berry, R. A. (1972). Some determinants of changing income distribution in colombia: 1930-1970. Center Discussion Paper 137, Yale University, Economic Growth Center, New Haven, CT.

Beyer, R. C. (1947). *The Colombian Coffee Industry: Origins and Major Trends, 1740-1940*. Ph. D. thesis, University of Minnesota.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2024). On Binscatter. *American Economic Review 114*(5), 1488–1514.

Conley, T. G. and M. Kelly (2025). The standard errors of persistence. *Journal of International Economics 153*, 104027.

Departamento Administrativo Nacional de Estadística and A. Charry Lara (1962). *Directorio Nacional de Explotaciones Agropecuarias (Censo Agropecuario) 1960: Resumen Nacional*. Bogotá: DANE.

Dirección General de los Censos. (1940–1942). Censo general de población, 5 de julio de 1938, ordenado por la ley 67 de 1917 / Dirección General de los Censos. In *Censo general de población, 5 de julio de 1938, ordenado por la ley 67 de 1917 /*. Bogotá: Imprenta Nacional, / Colombia. Dirección General de los Censos.

Dirección Nacional de Estadística. (1923). Censo de población de la República de Colombia, levantado el 14 de octubre de 1918 ; con especificación de los varones mayores y menores de 21 años / Dirección Nacional de Estadística. In *Censo de población de la República de Colombia, levantado el 14 de octubre de 1918 ; con especificación de los varones mayores y menores de 21 años* (2a. ed revisada y corregida. ed.)., Misceláneas-Colecciones ; no. 610/3. Bogotá: Imp. Nacional,.

Durán y Díaz, J. ([1794] 2012). *Estado General de todo el Virreynato de Santafé de Bogotá en el presente año de 1794*. Bogotá: Banco de la República.

Eckert, F., A. Gvirtz, J. Liang, and M. Peters (2020). A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790.

Faguet, J.-P., C. Matajira, and F. Sánchez (2024). *Encomienda*, the colonial state, and long-run development in Colombia. *The Economic Journal*.

Federación Nacional de Cafeteros de Colombia (1933). *Boletín de Estadística de La Federación Nacional de Cafeteros, Número 5*, Volume Vol. I. Federación Nacional de Cafeteros de Colombia.

Federación Nacional de Cafeteros de Colombia (1945). *Boletín de Estadística de La Federación Nacional de Cafeteros, Número 26*, Volume Vol. VI. Federación Nacional de Cafeteros de Colombia.

Fergusson, L. and J. F. Vargas (2023). Colombia: Democratic but Violent? In F. Valencia Caicedo (Ed.), *Roots of Underdevelopment: A New Economic and Political History of Latin America and the Caribbean*, pp. 285–316. Cham: Springer International Publishing.

Fergusson, L., J. F. Vargas, D. Salazar, and G. Vesga (2013). Don't make war, make elections franchise extension and violence in XIXth-Century Colombia.

García Jimeno, C. (2005). Colonial institutions and long-run economic performance in Colombia: Is there evidence of persistence? *Documentos CEDE No. 59 Octubre de 2005*.

Grueso, E. R. (1950). *Geografía guerrera de Colombia*. Impr. Bolivariana.

Gutiérrez Tovar, G. (1961). *Cinco Años de Criminalidad Aparente (1955-1959): Contribución al Estudio de La Criminalidad En Colombia*. Bogotá: Ministerio de Justicia - República de Colombia.

Guzmán, G., O. Fals Borda, and E. Umaña Luna (1962). *La Violencia En Colombia: Estudio de Un Proceso Social. Tomo I.* Number 10 in Colección Actores de La Violencia En Colombia. Bogotá, Colombia: Ediciones Tercer Mundo.

Guzmán, G., O. Fals Borda, and E. Umaña Luna (1964). *La Violencia En Colombia: Estudio de Un Proceso Social. Tomo II* (Segunda edición ed.). Number 11 in Colección Actores de La Violencia En Colombia. Bogotá, Colombia: Ediciones Tercer Mundo.

Horan, P. M. and P. G. Hargis (1995). County longitudinal template, 1840-1990.

Jaramillo Castillo, C. E. (1991). *Los guerrilleros del novecientos*. CEREC.

LeGrand, C. (1986). *Frontier Expansion and Peasant Protest in Colombia, 1850-1936* (1st ed ed.). University of New Mexico Press.

Lopez-Uribe, M. D. P. and F. Sanchez Torres (2024). Ideology and Rifles: The Agrarian Origins of Civil Conflict in Colombia. *World Development 173*, 106387.

Monsalve, D. (1927). *Colombia Cafetera*. Barcelona: Artes Gráficas.

Nunn, N. and D. Puga (2012). Ruggedness: The Blessing of Bad Geography in Africa. *The Review of Economics and Statistics 94*(1), 20–36.

Ocampo, J. A. (2011). *Colombia y la economía mundial 1830-1910*. Bogotá: Ediciones Uniandes.

Oquist, P. H. (1980). *Violence, Conflict, and Politics in Colombia*. New York : Academic Press.

Ortiz Sarmiento, C. M. (1995). *Estado y Subversión En Colombia: La Violencia En El Quindío Años 60*, Volume 4 of *Historia Contemporánea*. Cerec.

Palacios, M. (2002). *Coffee in Colombia, 1850-1970: An Economic, Social and Political History*. Cambridge University Press.

Policía Nacional (1959). *Criminalidad Colombiana 1959*. Number No. 5. Bogotá: Editorial El Voto Nacional.

Policía Nacional (1962). *Criminalidad Colombiana 1962*. Number No. 5. Bogotá: Editorial El Voto Nacional.

Policía Nacional de Colombia (1965). *Estadística de Criminalidad En 1965*. Number No. 8. Bogotá: Editorial El Voto Nacional.

Roldán, M. (2002). *Blood and Fire: La Violencia in Antioquia, Colombia, 1946-1953*. Duke University Press.

Romero-Prieto, J. E. and A. Meisel-Roca (2019). Análisis demográfico de la Violencia en Colombia. Technical report, Banco de la República, Bogotá, Colombia.

Sánchez, G., D. Meertens, and A. Hynds (2001). *Bandits, Peasants, and Politics: The Case of ""La Violencia"" in Colombia*. Austin: University of Texas Press.

Sociedad de Productores de Café and Sociedad de Agricultores de Colombia (1923). *Revista Nacional de Agricultura. Órgano de La Sociedad de Productores de Café*, Volume Vol. 230. Bogotá: La Sociedad.

United Nations (1961). *Demographic Yearbook*, Volume 8. United Nations.

United Nations. Economic Commission for Latin America (1958). *Coffee in Latin America: Colombia and El Salvador*, Volume 1 of *Coffee in Latin America: Productivity Problems and Future Prospects*. United Nations.

Urrutia-Montoya, M. et al. (2000). Comercio exterior y actividad económica de Colombia en el siglo XX: Exportaciones totales y tradicionales. Borradores de Economía 163, Banco de la República.

Appendix B

Appendix to Chapter 2

B.1 Description of the data

For each country we use data from the most recent agricultural census for which data are available: Ecuador (2000), Chile and Mexico (2007), El Salvador, Paraguay (2008), Uruguay (2011), Perú (2012), Bolivia (2013), Colombia (2014). The Peruvian and Uruguayan censuses do not collect information on agricultural production and Chile only collects output data for cereals, leguminous vegetables, root crops, and industrial crops. See Table B.1 for a summary.

Sample

Farms are the unit of observation on the nine agricultural censuses. It is defined as a plot or group of plots, not necessarily contiguous, under the same decision-making unit (firm or household). Our analysis concentrates on private farms. Hence, we eliminate from the sample: (i) indigenous reservations and other type of land access arrangements for indigenous communities (Colombia); (ii) collective territories of afro-descendants or ancestral territories of *raizales* (Colombia); (iii) National Parks (Chile and Colombia); (iv) land with environmental restrictions for agricultural exploitation (Colombia); (v) agricultural communities (Bolivia and Chile); (vi) historical communities (Chile); and native communities (Peru).

All censuses, with the exception of Colombia and Ecuador, cover rural and urban areas, albeit without identifying where the plot is located. In addition, the Paraguayan and Uruguayan censuses exclude farms below 0.1 and 1 hectare, respectively. Our main results include farms in the whole range of areas for which data is available and eliminate non-agricultural farms and those without land, which are usually fish farms, poultry producers or greenhouses in urban areas. Lastly, in Perú we eliminate from the data *El Callao*, a district with only urban population.

Labor Input Variables

To measure labor inputs, we would ideally have the aggregate number of hours dedicated to agricultural activities on the farm, yet time use data is not collected on agricultural censuses. Hence, we calculate the yearly number of workers: total, family and hired. For estimating the number of workers, we define family workers as those household members or persons living in the farm and working permanently or temporarily on the farm.

In Bolivia, the census does not discern between permanent or temporary workers, but

makes the distinction between paid and unpaid labor. We assume that all unpaid workers are family workers. If the farm does not register workers in the employment module, the number of workers is assigned according to the household members that report participating in farm activities. For Mexico, once the number of workers from the employment module is calculated, the number of workers is imputed to the farms that do not register this, as follows:

- If the producer reports that she alone is in charge of the farm activities, one worker is assigned to her farm.
- If the producer reports that her family is involved in the farm activities, the number of workers is assigned according to the average number of family workers per state, according to the 2010 Population Census.
- If the producer reports that she is part of a group or cooperative, the number of workers is assigned according to the total number of members of the group. Note that this provides an upper bound for the total number of workers of farms that belong to cooperatives or groups of producers.

In Colombia, the census includes the number of daily temporary workers (*jornaleros*) employed during the previous month. We transform this variable to an equivalent of permanent workers by assuming that a permanent worker works 6 days a week for 4 weeks a month. This overestimates the number of agricultural workers by 7% when compared to the total number of agricultural workers estimated from the GEIH (3.6 million vs 3.4 million of workers, respectively).

Finally, the Uruguayan census asks for the number of *zafrales* hired in the census year, according to census guidelines, a temporary worker is equivalent to 200 *zafrales*.

Output per worker and land yields

We use the information on agricultural output to calculate land yields and output per worker. The percentage of farm producing more than one crop is 77% in Bolivia, 40% in Chile, 81% in Colombia, 85% in Ecuador, 60% in El Salvador, 52% in Mexico, 94% in Paraguay, and 86% in Peru. Hence, for aggregating production at the farm level, we use revenue instead of the quantity produced. We use FAO producer prices to construct measures of farm revenue. We transform output variables to USD constant prices using 2011 as base year.

In Paraguay, the production of some crops is measured in boxes, bushells, etc. We first calculate the average weight of those units weight and then convert it in ton. In El Salvador, the census registers input information at the producer level and not at the farm level. For producers that have more than one farm (0.82%), we allocate workers across farm using the percentage of land that each farm represents.

Equations 2.1 and 2.2 in the text present our definitions of land yields and output per worker. These measures use total farm revenue, the number of hectares cultivated for the crops with price information, and the number of hired and family workers allocated to crops with price information, which we calculate as the total number of workers multiplied by the percentage

of land cultivated on crops with price information.

The crops without price information represent a small percentage of total production. For Bolivia, 74.65% of crops do not have price information, which amounts to 2.96% of total production. For Chile, these numbers are 45.45% and 4.12%, respectively. For Colombia, 88% and 8.55%. For Ecuador, 73.38% and 12.13%. In El Salvador 69.01% and 1.29%. In Mexico, 39.07% and 20.4%. In Paraguay, 82.73% and 6.71%. In Peru, 57.07% and 4.59%.

B.2 Derivations of the theoretical framework

There are N farmers with heterogeneous productivity z_i and a total land endowment L. Farmers use land ℓ_i and a fixed unit of labor to produce a simple homogeneous good y_i with a price normalized to one and production function:[1]

$$y_i = z_i^{1-\gamma} \ell_i^{\gamma}, \tag{B.1}$$

with $\gamma < 1$.

The efficient land allocation solves the social planner's problem of maximizing aggregate output $\sum_{j=1}^{N} y_j$ subject to the land endowment constraint $\sum_{j=1}^{N} \ell_j = L$. This maximization problem is given by

$$\textbf{max} \sum_{j=1}^{N} z_j^{1-\gamma} \ell_j^{\gamma} + \lambda \left(\sum_{j=1}^{N} L - \ell_j \right), \tag{B.2}$$

where λ is the Lagrangian multiplier associated to the land endowment constraint. The solution to the maximization problem in (B.2) is

$$\ell_i^* = \frac{z_i}{\sum_{j=1}^{N} z_j} L \quad \text{for} \quad i = \{1, ..., N\}. \tag{B.3}$$

Substituting into (B.1) gives

$$y_i^* = z_i \left(\frac{L}{\sum_{j=1}^{N} z_j} \right)^{\gamma} \quad \text{for} \quad i = \{1, ..., N\}. \tag{B.4}$$

Proof of Result 1. The distribution of farmers' productivity determines the farm size

[1] We thank Diego Restuccia for his generosity in sharing his valuable insights, which we closely follow to develop this section.

distribution and more productive farmers operate larger farms. From (B.3) it can be seen that the share of land allocated to each farmer in the efficient allocation is proportional to her individual productivity. Farmers with a larger z_i hold a larger share of land. Also, the only source of variation in ℓ_i^* across farmers is z_i, therefore, the distribution of z_i determines the distribution of ℓ_i^*.

Proof of Result 2. Land yields are constant across farms and output per worker increases with farm size. From (B.3) and (B.4) we can derive an expression for land yields in the efficient allocation:

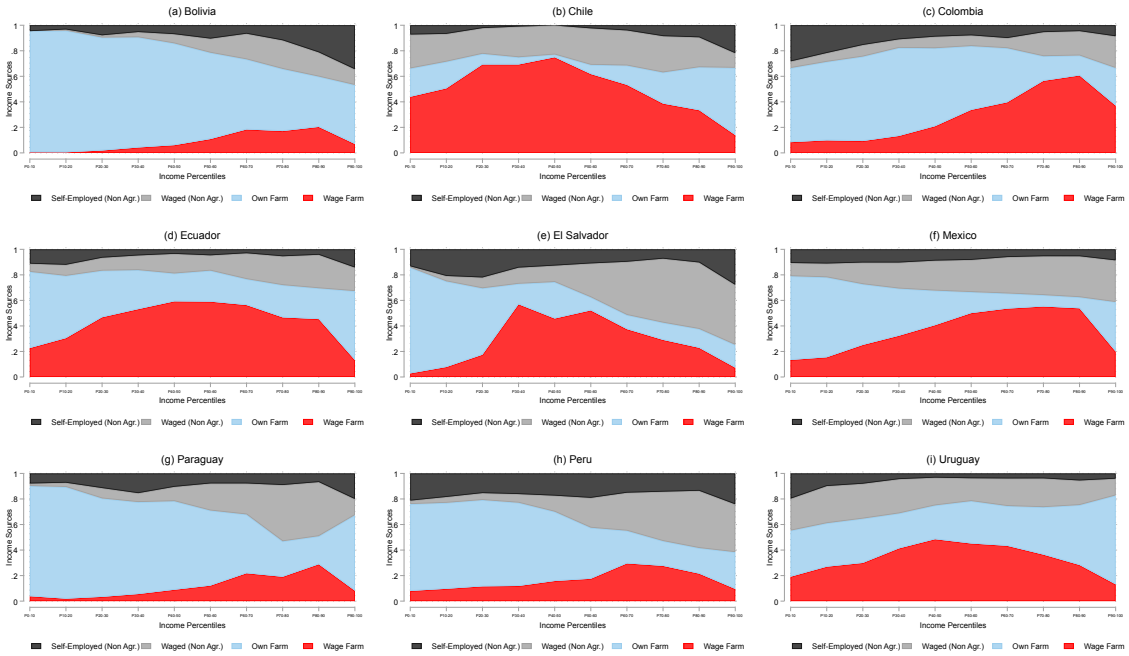
$$\frac{y_i^*}{\ell_i^*} = \left(\frac{\sum_{j=1}^{N} z_j}{L} \right)^{1-\gamma} \quad \text{for} \quad i = \{1, ..., N\}. \tag{B.5}$$

Land yields in expression (B.5) depend on aggregate productivity $\sum_{j=1}^{N} z_j$ and total land endowment L and do not vary across farmers. Output per worker is equal to farm output in (B.4), which increases with z_i.

Result 3. The distribution of farmers' productivity z_i determines the income distribution and equalizing farm size does not eliminate income inequality. In this simple framework, farm output determines farmers' income. From equation (B.4) it can be seen that farm output depends on individual productivity z_i, and therefore the distribution of income is a function of the productivity distribution. If land is evenly distributed across farmers farm size is $\hat{\ell}_i = \frac{L}{N}$ for $i = \{1, .., N\}$ and farm output $\hat{y}_i = z_i^{1-\gamma} \left(\frac{L}{N}\right)^{\gamma}$, which still depends on z_i.

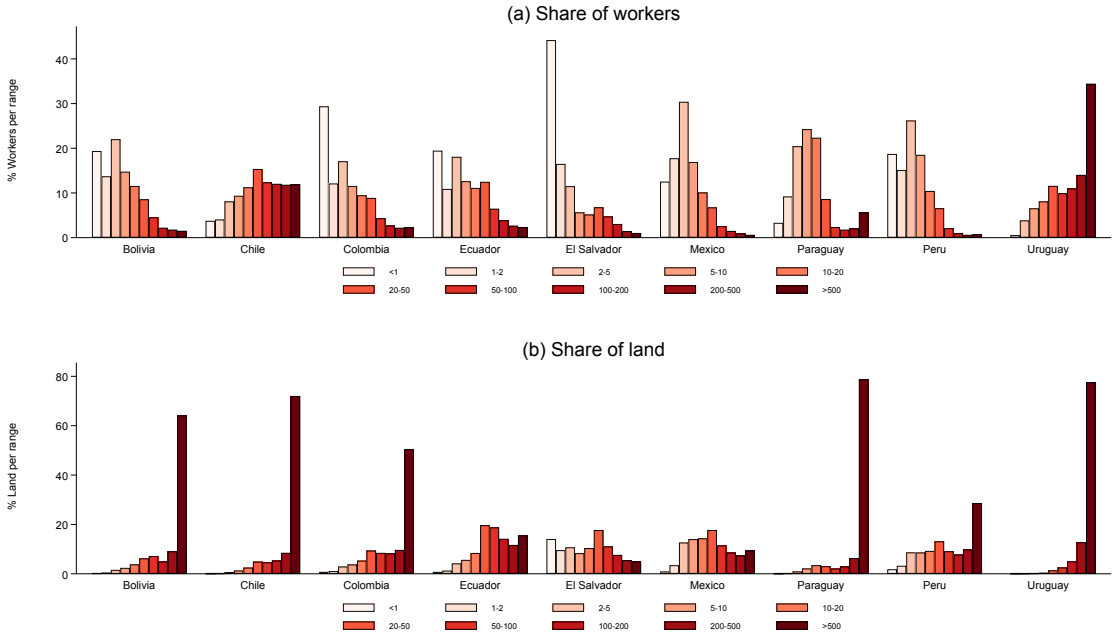
B.3 Additional Figures and Tables

Figure B.1. Individual Level Decomposition of Income Sources



Notes: We perform the same analysis at the individual level as a robustness check. Includes only individuals in agricultural households, whether they are rural or urban. Individuals are ranked in deciles across income distribution. Total income of all individuals in each decile are then plotted according to four different income sources.
Sources: Bolivia - EH (2014); Chile - CASEN (2006); Colombia - GEIH (2014); Ecuador - ENEMDU (2007); El Salvador - EHPM (2008); México - ENIGH (2016); Paraguay - EPH (2017); Perú - ENAHO (2012); Uruguay- ECH (2011).

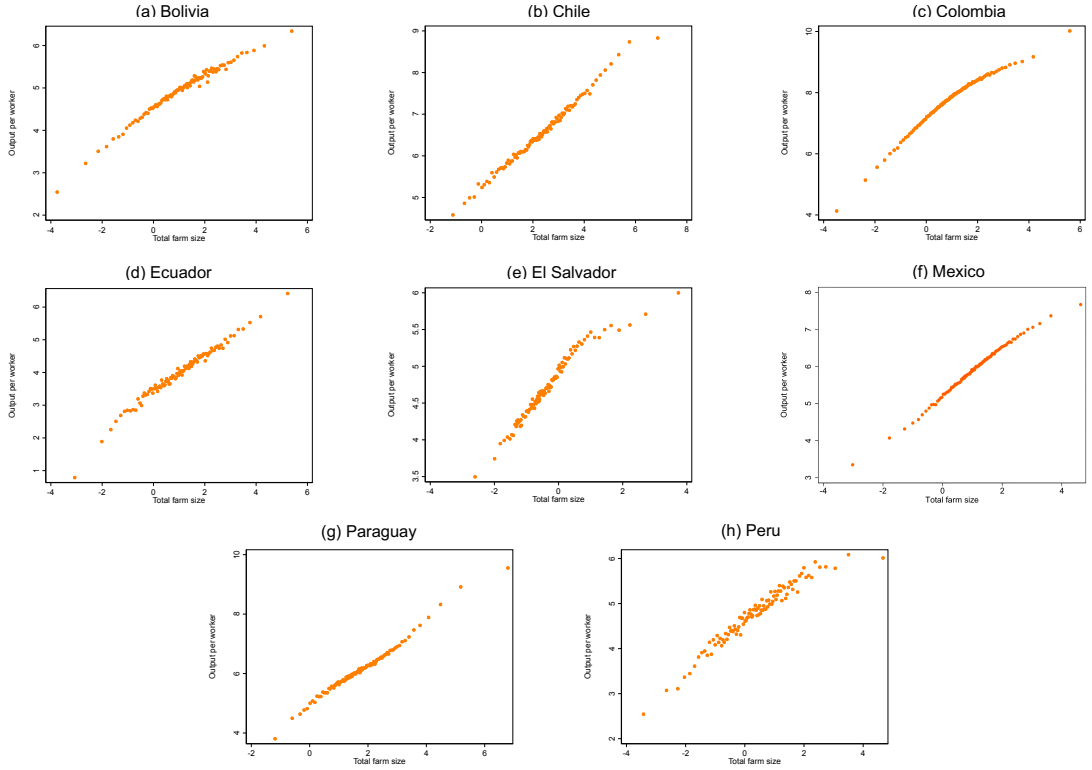
Figure B.2. Share of land and labor by total farm area



Notes: The sample includes all farms that report at least one worker. Farm size measured as total farm area (hectares). Farm workers include family and hired workers (permanent and seasonal).

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012); Uruguay - CGAXIX (2011).

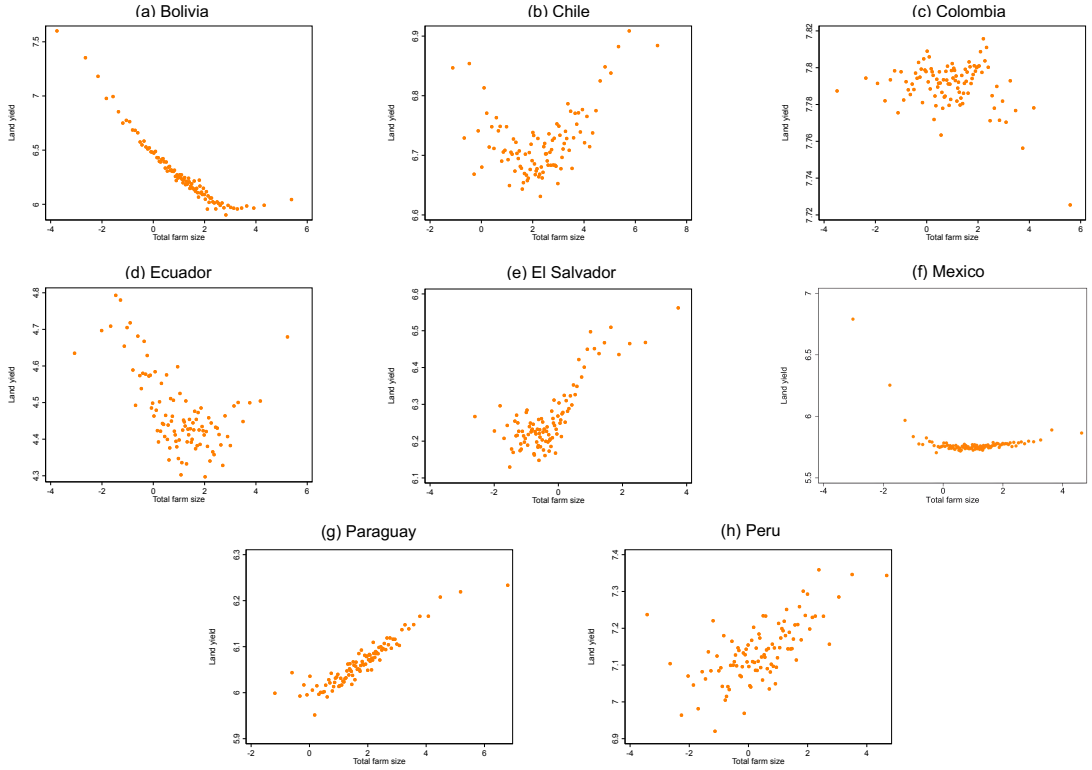
Figure B.3. Output per worker and total farm area



Notes: Binned scatterplots of the relationship between log of output per worker and log of total farm area controlling for fixed effects of regions within countries.

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - ENA (2019).

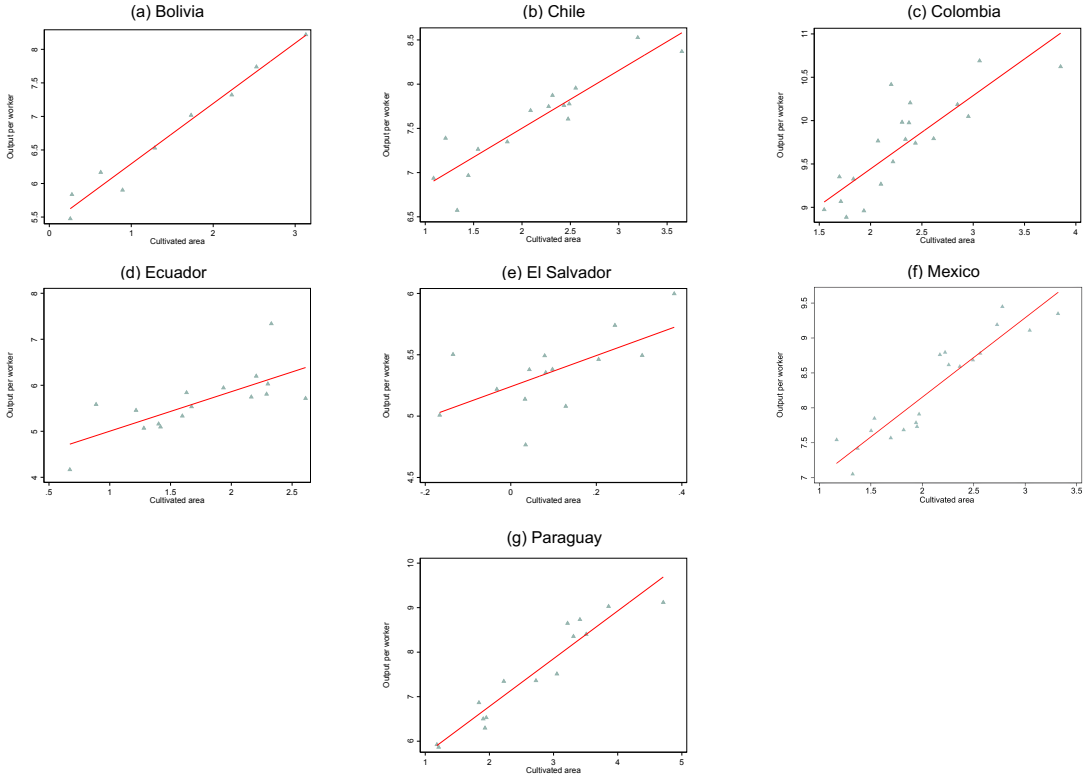
Figure B.4. Land yields and total farm area



Notes: Binned scatterplots of the relationship between log of land yields and log of total farm area controlling for fixed effects of regions within countries.

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - ENA (2019).

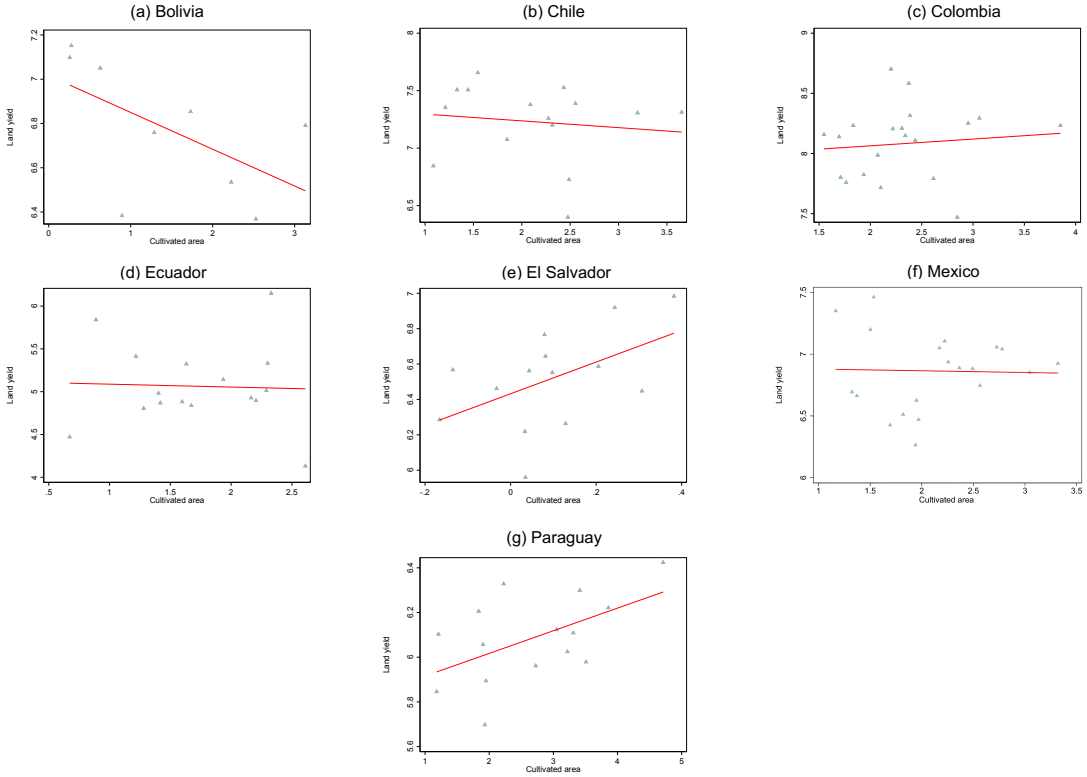
Figure B.5. Output per worker and farm size



Notes: Binned scatterplots of the relationship between log of average output per worker and log of average cultivated area controlling for agricultural suitability and total rainfall across regions. Farms without production are excluded.

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008).

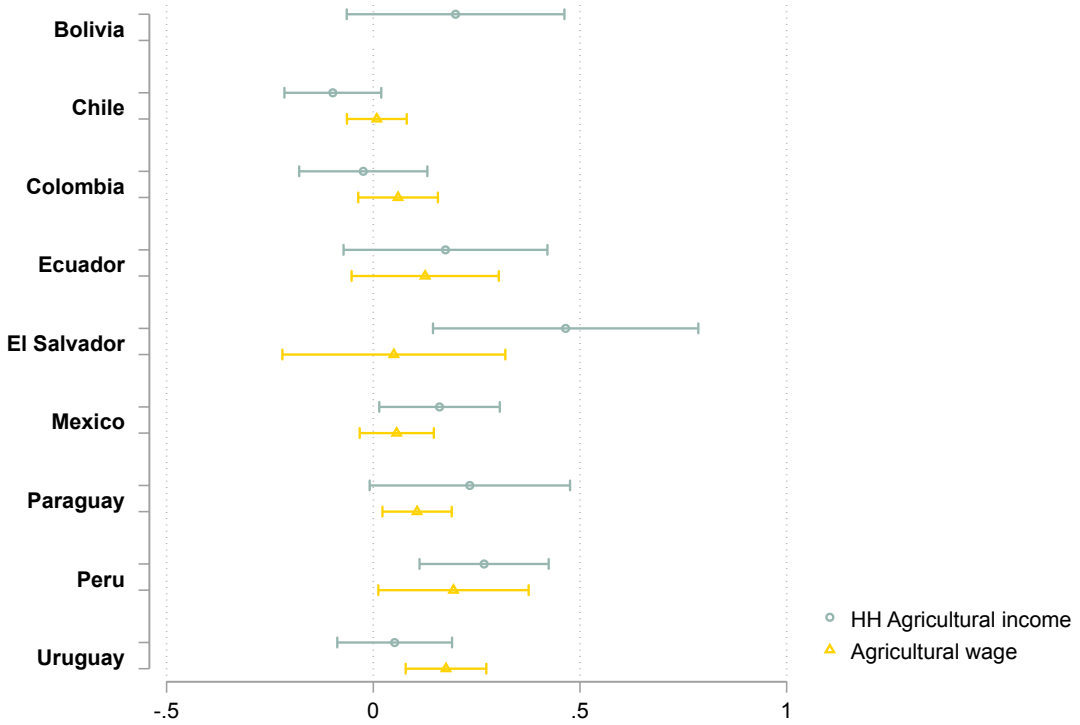
Figure B.6. Yields per hectare and farm size



Notes: Binned scatterplots of the relationship between log of average land yield and log of average cultivated area controlling for agricultural suitability and total rainfall across regions. Farms without production are excluded.

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008).

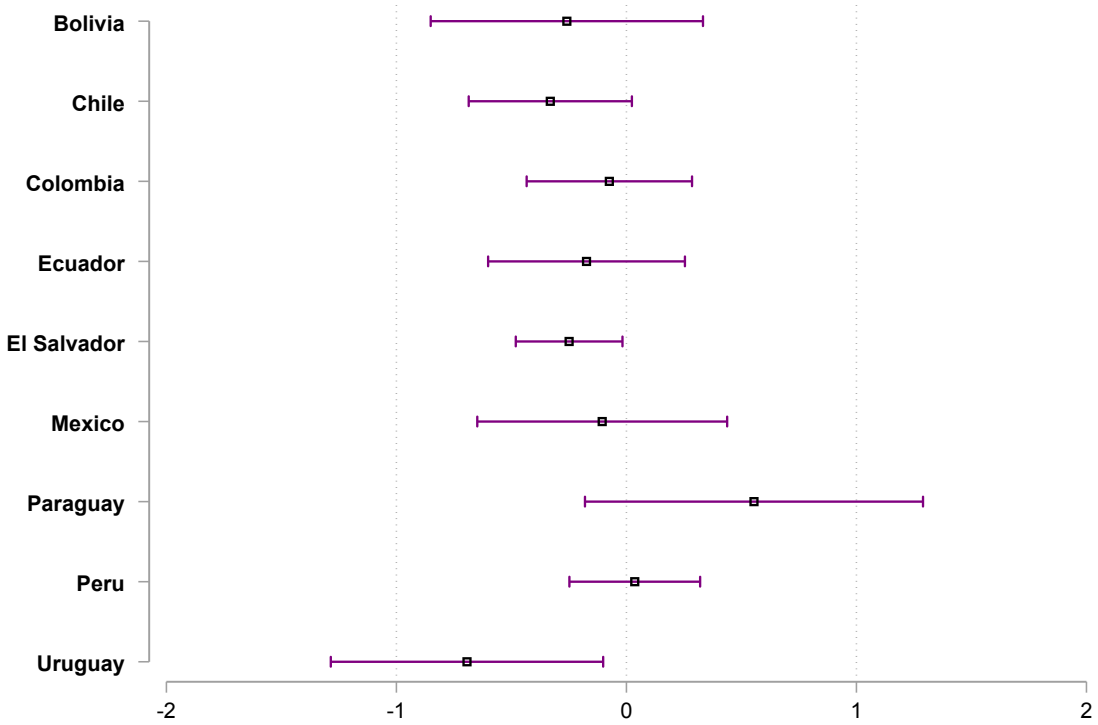
Figure B.7. Relationship between household agricultural income and total farm area



Notes: Coefficient estimates from independent linear regressions with log of average farm size (measured as total farm area) as independent variable and log of average agricultural income and log of average agricultural wages across regions as dependent variables. Regional averages of soil quality and rainfall are included as control. Bolivia is not included, because the sample that receives agricultural wages is not representative.

Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007);Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay- ECH (2011) and CGAXIX (2011).

Figure B.8. Relationship between agricultural income Gini and farm size Gini



Notes: Coefficient estimates from independent linear regressions with the Gini of agricultural income as dependent variable and the Gini cultivated area across regions as independent variable. Regional averages of soil quality and rainfall included as controls.
Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay - ECH (2011) and CGAXIX (2011).

Table B.1. Description of Agricultural Censuses by Country

Variable	Bolivia	Chile	Colombia	Ecuador	El Salvador	México	Paraguay	Perú	Uruguay
Census and year	1er Censo Agropecuario del Estado Plurinacional de Bolivia - CAI (2013).	VII Censo Nacional Agropecuario y Forestal - CNAFVII (2006-2007).	3er Censo Nacional Agropecuario - CNAIII (2014).	III Censo Nacional Agropecuario - CNAIII (2000).	IV Censo Agropecuario - CAIV (2007-2008).	VIII Censo Agrícola, Ganadera y Forestal - CAGFVIII (2007).	Censo Agropecuario Nacional - CAN (2008).	IV Censo Nacional Agropecuario - CNAIV (2012).	XIX Censo General Agropecuario - CGAXIX (2011)
Inclusion criteria	Agricultural farms settled in rural areas, peri-urban areas, and special cases in urban areas.	All rural plots with agricultural production, livestock and forestry and urban plots with agricultural production, livestock, and forestry.	Rural area registered on the Official Cadaster (IGAC) which covers 99% of the total rural area.	Rural agricultural farms.	Rural and urban agricultural producers. Cover 100% of rural census segments and 16% of urban census segments.	Rural geostatistical areas with and without agricultural and forestry activity, as well as urban geostatistical areas with livestock.	Agricultural farms active during the census. It does not cover all rural land plots.	Agricultural farms.	Agricultural farm greater than or equal to 1 ha.
Definition of farm (Agricultural Production Unit)	Plot, set of land or plots, used totally or partially for agricultural activities, conducted as an economic unit by a producer, regardless of size, tenure regime or legal status.	Plot used in agricultural production, livestock, and forestry by a one decision-making unit (household or firm). A farm may cover a part of a land plot, the total plot, or a sum of land plots (or parts). Plots can be contiguous or separated located in one comuna. Covers landless exploitations dedicated to animal or livestock production located in urban and rural areas.	A farm may cover a part of a land plot, the total plot, or a sum of land plots (or parts). Plots can be contiguous or separated in one or several municipalities. Criteria to consider a land plot as agricultural: (i) active in agricultural production, livestock, forestry, and/or fish farms; (ii) operated by one decision-making unit (household or firm): and (iii) uses at least one production input.	Plot of a size equal or larger than 0.05 hectares, dedicated fully or partially to agricultural production and under one decision -making unit. Plots below 0.05 hectares were included only if, during the year of the census, they produced and sold an agricultural product.	Plot used in agricultural production, livestock, and fish farms. A farm may cover a part of a land plot, the total plot, or a sum of land plots (or parts). Plots can be contiguous or separated in one municipality. Operated by one decision-making unit (household or firm) or a group of decision-making units. It includes landless exploitations.	Set of lands with or without agricultural or forestry activity in the rural area or with agricultural activity in the urban area, located in the same municipality; Animals that are owned regardless of the place where they are, as well as the equipment, machinery and vehicles destined for agricultural, livestock or forestry activities must be managed under the same administration.	Plot used in agricultural production, or livestock. A farm may cover a part of a land plot, the total plot, or a sum of land plots (or parts). Operated by one decision-making unit (household or firm) or with participation of additional decision-making units (household or firm). Inclusion criteria: (i) 0.1 hectares with seasonal crops, commercial vegetable garden, permanent crops, flowers, ornamental plants and/or greenhouses; (ii) three livestock units; (iii) five sheep, goat or porks; (iv) 100 fowls; and/or (v) 10 beehives.	Plot, or group of plots, used in agricultural production, or livestock under one decision-making unit. Exclusion criteria: (i) concession for the exploitation of natural products; (ii) fish farms; or (iii) production of fighting cocks, racing horses; and fighting bulls.	Includes all land dedicated totally or partially to agricultural, livestock and/or forestry purposes under one decision-making unit. Regardless of tenure, the condition legal and that productive activities are carried out or not for commercial purposes.
Crops	Annual and permanent summer crops, forage grass and forestry. No production for forage grass and forestry.	Yearly forage crops, permanent forage, vegetables, flowers, vineyards, fruits, and forestry. Production information for cereals, leguminous vegetables, root crops, and industrial crops.	Seasonal crops, permanent crops, forestry, and forage crops. Production data for all crops.	Seasonal crops, permanent crops, flowers, and forage grass. No production for forage grass.	Grains, vegetables, seasonal and permanent agro-industrial crops, fruits, and coffee. No production for coffee.	Annual spring-summer crops and perennials (Fruit trees, plantations or cultivated grass).	Seasonal crops, permanent crops, vegetables, forage crops, and forestry. No production information for forage crops and forestry.	Permanent crops (agro-industrial crops, fruit trees, forestry and cultivated grass) and seasonal crops (cereals, fruits, vegetables, grains, root crop and forage crops). No data for agricultural production.	Forestry, fruits, vineyards, vegetables, cereals, industrial crops and forage crops. No data for agricultural production.
Prices	Agricultural producer prices FAO (2013).	Agricultural producer prices FAO (2006).	Agricultural producer prices FAO (2013).	Agricultural producer prices FAO (2000).	Agricultural producer prices FAO (2006).	Agricultural producer prices FAO (2007).	Agricultural producer prices FAO (2008).		

Notes: Output data for Perú comes from the Agricultural Survey (2019), which includes agricultural units of small and medium producers and companies and large producers. The survey covers permanent and seasonal crops. Producer prices for each crop are taken from FAO (2019).

Sources: Own elaboration based on census methodological guides.

Table B.2. Number and area of farms by farm size range

Country	Variable	All sizes	<1 ha	1-2 ha	2-5 ha	5-10 ha	10-20 ha	20-50 ha	50-100 ha	100-200 ha	200-500 ha	>500 ha
Bolivia	Number of farm	844,104	223,044	108,692	163,538	114,454	93,902	68,305	39,032	12,655	9,593	10,889
	Average farm size (ha)	38.35	0.35	1.33	3.08	6.75	13.07	29.52	59.41	127	306.99	1,906.44
	Share of farm (%)	100	26.42	12.88	19.37	13.56	11.12	8.09	4.62	1.50	1.14	1.29
	Share of farm size (%)	100	0.24	0.45	1.56	2.39	3.79	6.23	7.16	4.96	9.10	64.12
Chile	Number of farm	218,487	26,196	23,691	39,259	35,759	33,936	30,049	12,809	7,397	5,319	4,072
	Average farm size (ha)	87.83	0.51	1.31	3.19	7.07	14	31.12	69.12	138.43	306.12	3,393.94
	Share of farm (%)	100	11.99	10.84	17.97	16.37	15.53	13.75	5.86	3.39	2.43	1.86
	Share of farm size (%)	100	0.07	0.16	0.65	1.32	2.48	4.87	4.61	5.34	8.48	72.02
Colombia	Number of farm	1,893,939	742,698	276,037	338,508	198,848	139,811	112,281	45,268	22,339	12,015	6,134
	Average farm size (ha)	19.71	0.38	1.45	3.23	7.10	14.11	31.19	69.73	137.99	299.76	3,070.11
	Share of farm (%)	100	39.21	14.57	17.87	10.50	7.38	5.93	2.39	1.18	0.63	0.32
	Share of farm size (%)	100	0.76	1.07	2.93	3.78	5.28	9.38	8.45	8.26	9.65	50.44
Ecuador	Number of farm	692,551	194,497	95,889	139,830	85,114	64,613	65,875	29,833	11,323	4,370	1,207
	Average farm size (ha)	14.94	0.4	1.33	3.1	6.83	13.46	30.91	65.19	129.48	276.47	1,328.18
	Share of farm (%)	100	28.08	13.85	20.19	12.29	9.33	9.51	4.31	1.64	0.63	0.17
	Share of farm size (%)	100	0.75	1.24	4.19	5.62	8.41	19.68	18.79	14.17	11.67	15.50
El Salvador	Number of farm	388,742	262,799	67,269	33,224	10,913	6,895	5,357	1,525	531	178	51
	Average farm size (ha)	2.37	0.49	1.31	2.97	7.04	13.91	30.55	67.04	132.88	286.84	908.23
	Share of farm (%)	100	67.69	17.30	8.55	2.81	1.77	1.38	0.39	0.14	0.05	0.01
	Share of farm size (%)	100	14.07	9.56	10.71	8.33	10.39	17.73	11.07	7.64	5.52	4.97
México	Number of farm	3,961,799	582,185	766,063	1,193,665	638,268	380,129	248,771	80,522	37,529	20,863	13,804
	Average farm size (ha)	16.77	0.4	1.15	2.88	6.66	13.21	28.55	67.25	132.27	293.79	2,099.52
	Share of farm (%)	100	24.69	19.34	30.13	16.11	9.59	6.28	2.03	0.95	0.53	0.35
	Share of farm size (%)	100	0.35	1.33	5.18	6.40	7.56	10.69	8.15	7.47	9.23	43.63
Paraguay	Number of farm	261,428	13,175	32,219	59,940	62,497	54,065	20,784	5,826	4,072	3,882	4,968
	Average farm size (ha)	70.71	0.45	1.18	2.90	6.30	11.86	26.99	66.66	133.60	303.63	2,931.17
	Share of farm (%)	100	5.04	12.32	22.93	23.91	20.68	7.95	2.23	1.56	1.48	1.90
	Share of farm size (%)	100	0.03	0.21	0.94	2.13	3.47	3.03	2.10	2.94	6.38	78.77
Perú	Number of farm	2,123,973	790,483	415,543	477,186	212,691	115,051	73,079	22,451	9,521	5,469	2,499
	Average farm size (ha)	7.54	0.37	1.25	2.91	6.49	12.87	28.80	65.01	129.99	286.79	1,836.36
	Share of farm (%)	100	37.22	19.56	22.47	10.01	5.42	3.44	1.06	0.45	0.26	0.12
	Share of farm size (%)	100	1.82	3.23	8.66	8.62	9.25	13.14	9.11	7.73	9.79	28.65
Uruguay	Number of farm	44,653		426	2,635	4,197	4,716	6,899	5,693	5,625	6,405	8,057
	Average farm size (ha)	364.17		1.1	3.27	7.07	14.2	32.71	72.23	144.43	324.5	1,567.31
	Share of farm (%)	100		0.95	5.90	9.40	10.56	15.45	12.75	12.60	14.34	18.04
	Share of farm size (%)	100		0.00	0.05	0.18	0.41	1.39	2.53	5	12.78	77.66

Notes: The sample includes all farms that report at least one worker. Farm size is measured as total farm area.

Sources: Own calculations based on: Bolivia - CAI (2013); Chile - CNVAFVII (2006-2007); Colombia - CNAIII (2014); Ecuador - CNAIII (2000); El Salvador - CAIV (2007-2008); México - CAGFVIII (2007); Paraguay - CAN (2008); Perú - CNAIV (2012); Uruguay - CGAXIX (2011).

Table B.3. Definition of Rural Population by Country

Country	**Definition**
Bolivia	The area of population dispersed or grouped in population entities that do not exceed 2,000 inhabitants.
Chile	Human settlement with: (i) a population lower or equal to 1,000 inhabitants; or (ii) between 1,001 and 2000 inhabitants in which more than 50% work on the primary sector.
Colombia	Area with a high dispersion of houses and plots with farming exploitation. The area lacks an official grid street plan, official street name, public services, or other amenities typical of urban areas.
Ecuador	Rural areas are human settlements with: (i) less than 100 adjacent housing unit and it is not a district capital; or (ii) more than 100 housing units that are dispersed.
El Salvador	The rural area is the residual of the urban area in a municipality. An urban area is: (i) the population center that contains the municipal mayorship and is known as urban zone of the municipality; or (ii) conglomerate of housing units located far from the urban zone, has a population density equal or above to 1,000 inhabitants per squared kilometer. If 5% or less of the population is rural, the municipality is defined as urban.
México	Locality that has less than 2500 inhabitants and is not the municipal seat.
Paraguay	The urban area are the district urban area and the rural area the residual of the municipality.
Perú	In the population census, rural areas are human settlements with: (i) less than 100 adjacent housing unit; or (ii) more than 100 housing units that are dispersed. In the ENAHO survey, the rural area of a district is any population center with less than 2,000 inhabitants.
Uruguay	For the National Institute of Statistics, the rural area is the one that is outside the urban area. The urban area is delimited according to Law No. 10,723 on Populated Centers.

Table B.4. Agricultural income and farm size

PANEL A	Agricultural income								
	Bolivia	Chile	Colombia	Ecuador	El Salvador	México	Paraguay	Perú	Uruguay
Total farm size	0.20	-0.10*	-0.02	0.17	0.47***	0.16**	0.23*	0.27***	0.05
	(0.10)	(0.05)	(0.07)	(0.11)	(0.14)	(0.07)	(0.11)	(0.07)	(0.07)
R-squared	0.56	0.41	0.01	0.52	0.53	0.52	0.41	0.73	0.16
PANEL B									
Cultivated area	0.25	-0.28**	0.09	0.22**	0.48***	0.24**	0.28***	0.30***	0.10
	(0.13)	(0.12)	(0.12)	(0.08)	(0.12)	(0.10)	(0.07)	(0.10)	(0.07)
R-squared	0.56	0.49	0.03	0.63	0.63	0.53	0.66	0.69	0.23
Observations	9	15	24	16	14	32	15	24	19

Notes: Dependent variable: log average agricultural income at the region level. Panel A uses log average total farm size and Panel B log average total cultivated area. Geographic Controls include: agricultural suitability and total rainfall. Standard errors in parentheses.
*** significance at the 1%, ** significant at the 5%, * significant at the 10%.
Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay- ECH (2011) and CGAXIX (2011).

Table B.5. Agricultural wage and farm size

PANEL A	Agricultural wage							
	Chile	Colombia	Ecuador	El Salvador	México	Paraguay	Perú	Uruguay
Total farm size	0.01	0.06	0.13	0.05	0.06	0.11**	0.19**	0.18***
	(0.03)	(0.05)	(0.08)	(0.12)	(0.04)	(0.04)	(0.09)	(0.05)
R-squared	0.30	0.11	0.29	0.31	0.48	0.66	0.50	0.58
PANEL B								
Cultivated area	-0.02	0.08	0.12	-0.03	0.04	0.10***	0.20*	0.21***
	(0.08)	(0.08)	(0.07)	(0.11)	(0.06)	(0.03)	(0.12)	(0.05)
R-squared	0.30	0.08	0.33	0.31	0.45	0.72	0.46	0.63
Observations	15	24	16	14	32	15	24	19

Notes: Dependent variable: log average agricultural wage at the region level. Panel A uses log average total farm size and Panel B log average total cultivated area. Geographic Controls include: agricultural suitability and total rainfall. Standard errors in parentheses.
*** significance at the 1%, ** significant at the 5%, * significant at the 10%.
Sources: Own calculations based on: Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay - ECH (2011) and CGAXIX (2011).

Table B.6. Interquartile range of agricultural income and farm size

PANEL A	Agricultural income								
	Bolivia	Chile	Colombia	Ecuador	El Salvador	México	Paraguay	Perú	Uruguay
Total farm size	0.15	-0.08	-0.01	0.08	0.02	-0.02	0.14*	0.25**	0.02
	(0.14)	(0.11)	(0.06)	(0.06)	(0.16)	(0.07)	(0.06)	(0.11)	(0.07)
R-squared	0.42	0.30	0.03	0.22	0.12	0.28	0.42	0.61	0.12
PANEL B									
Cultivated area	0.26	-0.10	0.06	0.11*	-0.05	0.00	0.17**	0.41***	0.12**
	(0.21)	(0.25)	(0.09)	(0.06)	(0.12)	(0.10)	(0.06)	(0.14)	(0.06)
R-squared	0.45	0.28	0.05	0.32	0.14	0.28	0.56	0.66	0.33
Observations	9	15	24	16	14	32	15	24	19

Notes: Dependent variable: log IQR of agricultural income at the region level. Panel A uses log IQR of total farm size and Panel B log IQR of total cultivated area. Geographic Controls include: agricultural suitability and total rainfall. Standard errors in parentheses.
*** significance at the 1%, ** significant at the 5%, * significant at the 10%.
Sources: Own calculations based on: Bolivia - EH (2014) and CAI (2013); Chile - CASEN (2006) and CNVAFVII (2006-2007); Colombia - GEIH (2014) and CNAIII (2014); Ecuador - ENEMDU (2007) and CNAIII (2000); El Salvador - EHPM (2008) and CAIV (2007-2008); México - ENIGH (2016) and CAGFVIII (2007); Paraguay - EPH (2017) and CAN (2008); Perú - ENAHO (2012) and CNAIV (2012); Uruguay - ECH (2011) and CGAXIX (2011).

Appendix C

Appendix to Chapter 3

C.1 Main Robustness Checks

C.1.1 Census–Survey and Alternative Survey Comparisons

This appendix compares inequality measures derived from agricultural censuses and household surveys. Agricultural censuses report the distribution of *operational holdings*, whereas household surveys allow us to compute inequality among *landowner households*. We focus on the Gini index of land area and land value, and assess the extent to which census-based measures approximate survey-based estimates.

Appendix Figure C.1a plots the Gini coefficient of land area from the census (operational holdings) against the corresponding land-area Gini among landowners from household surveys. The two measures are strongly correlated across countries, with an R^2 of 0.94. This suggests that census data provide a reliable proxy for land-area inequality among landowners.

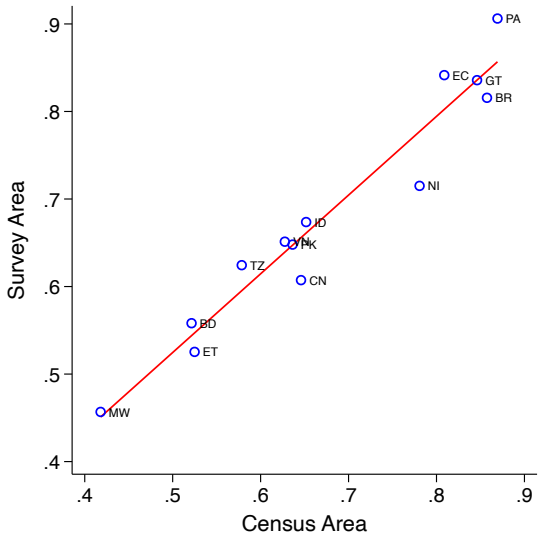
Appendix Figure C.1b compares census-based land-area inequality (operational holdings) with survey-based land-value inequality among landowners. The correlation is substantially weaker ($R^2 = 0.56$), indicating that land area is an imperfect proxy for the distribution of land wealth.

Finally, Appendix Figure C.1c plots survey-based land-area inequality against survey-based land-value inequality among landowners. The correlation remains moderate ($R^2 = 0.63$), confirming that differences between area and value inequality arise even within the same data source.

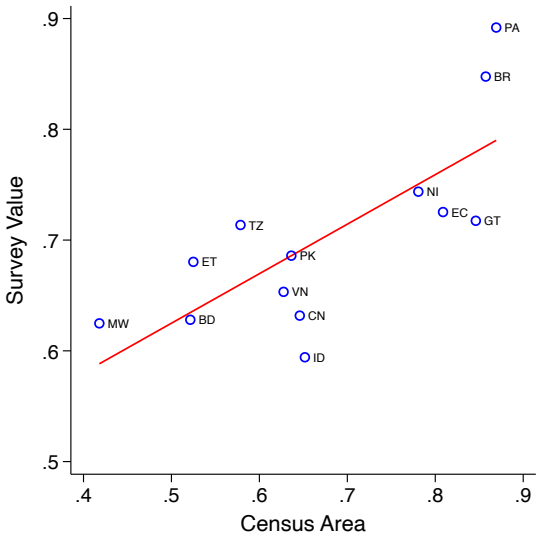
Taken together, these results show that while census data capture land-area inequality reasonably well, they are less informative about land-value inequality. Since land value is more closely related to wealth, relying exclusively on land-area measures may misrepresent cross-country differences in land inequality.

The following figure provides an additional validation by comparing land-area Gini coefficients from our benchmark household surveys with estimates obtained from the Demographic and Health Surveys (DHS) and from alternative waves of the same surveys. For each country, the figure plots Gini coefficients across available sources and years. The DHS-based estimates closely match the benchmark survey values, reinforcing the conclusion that household surveys provide reliable measures of land-area inequality. The stability across survey waves is also

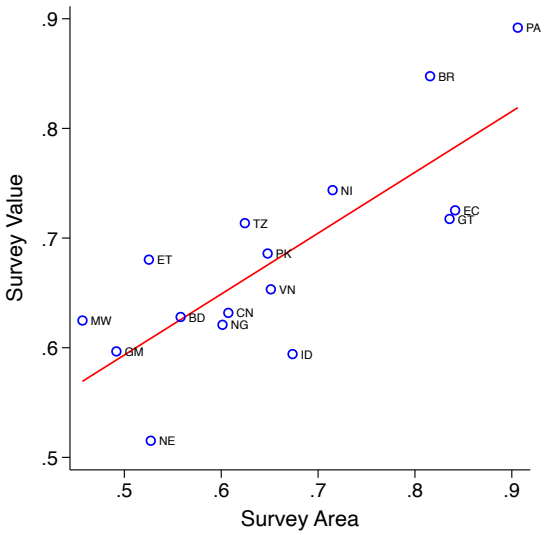
Figure C.1. Census and survey measures of land inequality



(a) Census vs. Survey: Land-area Gini



(b) Census (area) vs. Survey (value): Gini

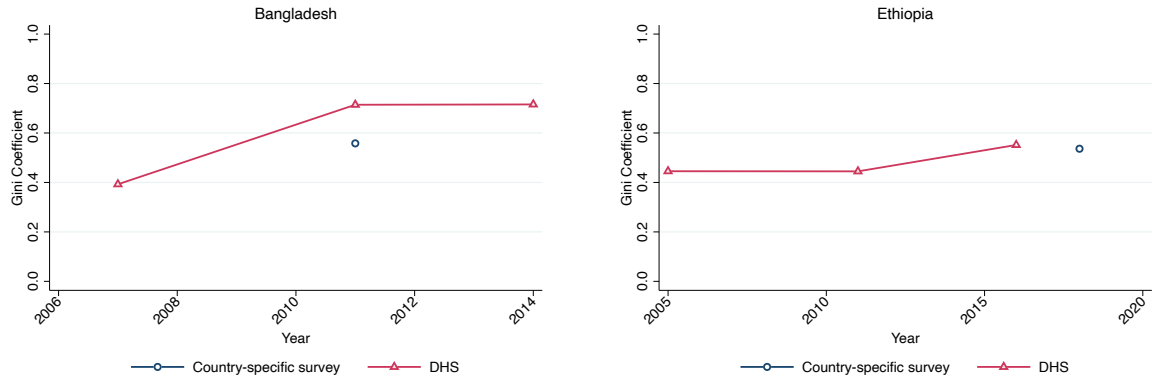


(c) Within-survey: Land-area vs. land-value Gini

Notes: Census-based Gini indices are computed over *operational holdings*, while survey-based Gini indices are computed over *landowner households*. Panel (a) plots census and survey land-area Gini coefficients. Panel (b) plots the census-based land-area Gini against the survey-based land-value Gini. Panel (c) compares, within surveys, land-area and land-value Gini indices among landowners. Each dot corresponds to a country; the fitted line is from an OLS regression. When reported in the text, the R^2 values are 0.94 in Panel (a), 0.56 in Panel (b), and 0.63 in Panel (c).

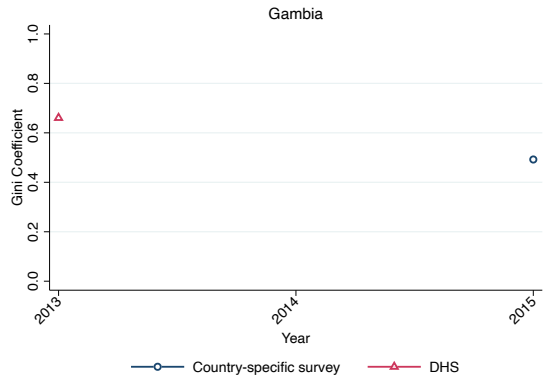
consistent with the high persistence of land inequality documented in the literature.

Figure C.2. Land inequality: alternative country-specific surveys and DHS

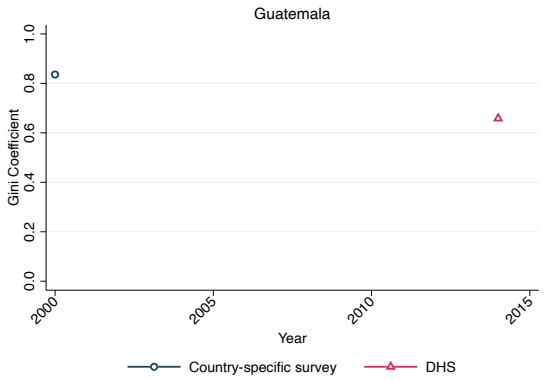


(a) Bangladesh

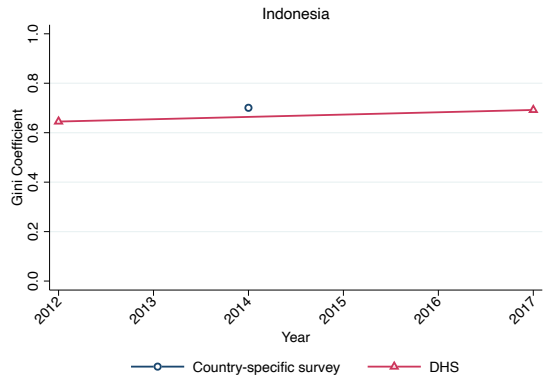
(b) Ethiopia



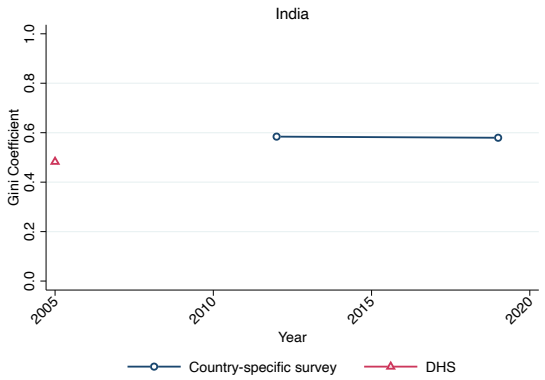
(c) Gambia



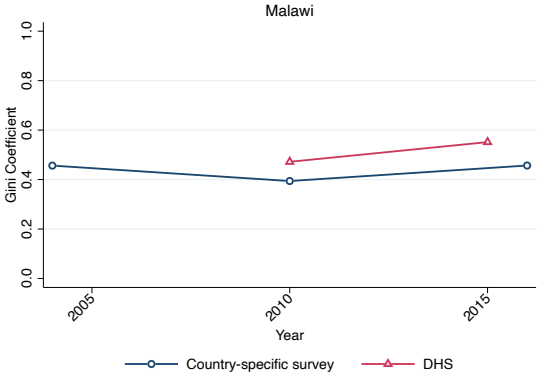
(d) Guatemala



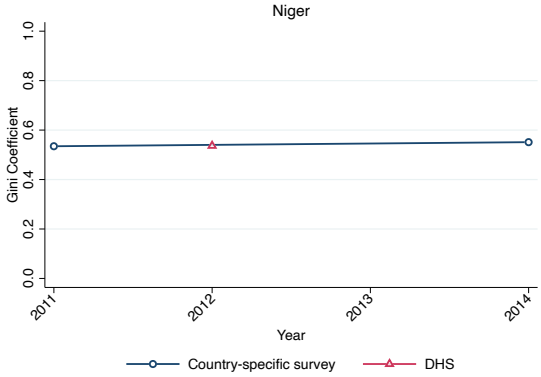
(e) Indonesia



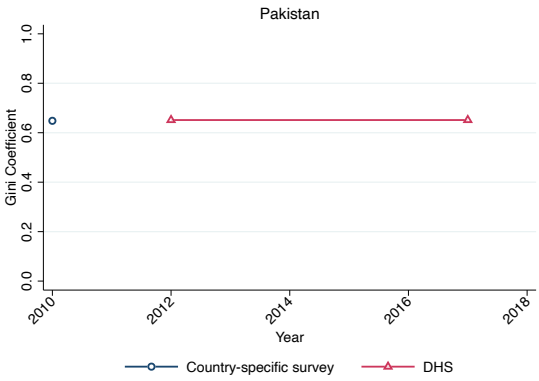
(f) India



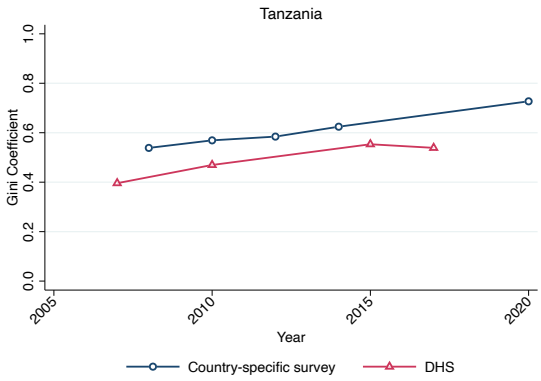
(g) Malawi



(h) Niger



(i) Pakistan



(j) Tanzania

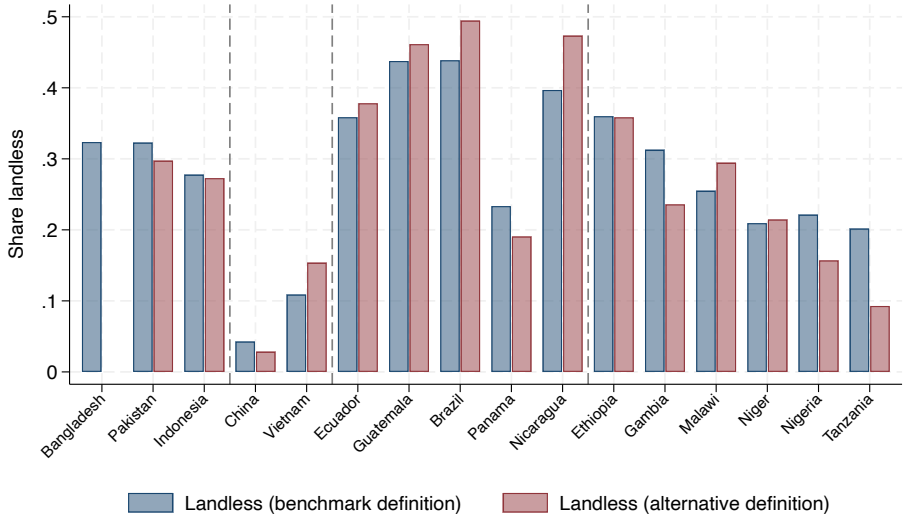
C.1.2 Alternative Definition of Landless Households

This subsection assesses the robustness of our results to an alternative definition of landless agricultural households. Our benchmark definition classifies as landless those households that report owning no agricultural land. The alternative definition relies on an agricultural classification based on labor participation, thereby modifying the reference population and, mechanically, the measured share of landless households. We show that this change has negligible implications for our main results. In Bangladesh, however, we are unable to construct the alternative definition due to data limitations, and the analysis therefore excludes this country.

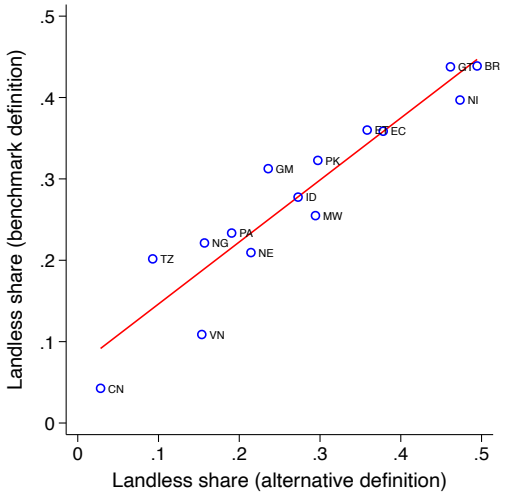
Figure C.3 presents a direct comparison of the two definitions. Panel (a) reports the landless share of agricultural households across countries using side-by-side bars, while Panel (b) displays a scatter plot contrasting the benchmark and alternative measures. The two definitions produce remarkably similar levels of landlessness. Differences are quantitatively small and do not alter the cross-country ranking. Countries with high landlessness under the benchmark definition (e.g., Guatemala, Brazil, and Nicaragua) remain so under the alternative definition, while countries with low landlessness (e.g., China and Vietnam) continue to exhibit very low shares. The scatter plot shows a tight clustering around the 45-degree line, indicating an almost one-to-one relationship between the two measures and a very high correlation between them.

Given this close correspondence, redefining the agricultural population has limited quantitative impact on our core results. While the alternative definition mechanically changes the reference group, the relative position of countries and the magnitude of landlessness remain broadly unchanged. This reinforces the conclusion that our findings are not driven by the particular way in which landless households are defined.

Figure C.3. Alternative Definitions of Landless Households



(a) Share of Landless Households: Benchmark vs. Alternative Definition



(b) Scatter Plot: Benchmark vs. Alternative Definition

Notes: Panel (a) reports the landless share of agricultural households under the benchmark definition and the alternative definitions of landless. Differences in levels are small and do not alter the cross-country ranking. Panel (b) plots the two measures against each other. Observations cluster tightly around the 45-degree line, indicating a very high correlation between definitions and confirming that the choice of definition has negligible impact on cross-country comparisons.

C.2 Robustness for Section 3.3

C.2.1 Quantile Effects of Land Ownership

The main specification (column 2 of Table 3.2) estimates the mean effect of land ownership on log income among agricultural households. Figure C.4 examines whether this effect varies across the conditional income distribution by estimating quantile regressions of log equivalized income on the landowner indicator at the 10th, 25th, 50th, 75th, and 90th percentiles. All regressions use the same sample restrictions as the main table (agricultural households, zero-income households dropped, income winsorized at the 99th percentile).

The ownership coefficient increases monotonically across the income distribution: it is negative and significant at the 10th percentile, negative at the 25th percentile, positive and significant at the median, and increases in magnitude while remaining significant at the 75th and 90th percentiles. The small and statistically insignificant mean effect reported in the main table therefore masks substantial heterogeneity: land ownership is associated with *lower* income at the bottom of the distribution and *higher* income at the top. This pattern is consistent with the interpretation that land ownership per se does not guarantee higher income; what matters is the *value* of the land owned, which is concentrated among higher-income agricultural households.

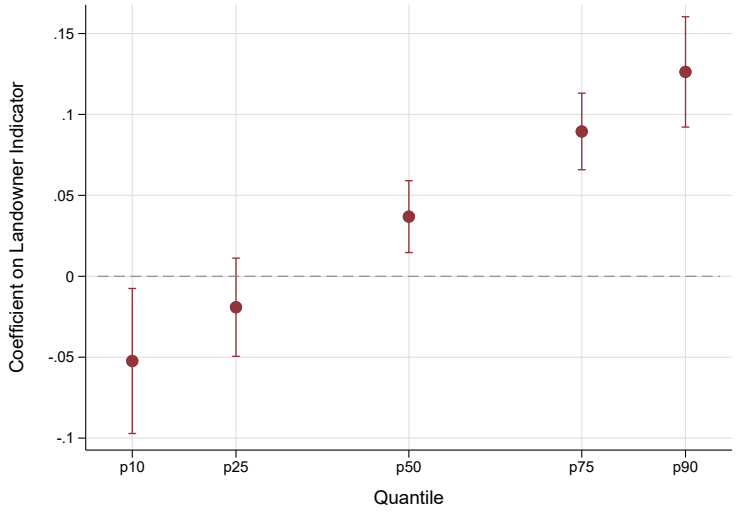


Figure C.4. Quantile regression coefficients: landowner indicator on log income

Notes: Each point shows the coefficient on the landowner dummy from a quantile regression of log equivalized income at the indicated percentile, with survey fixed effects (as indicator dummies). Vertical bars show 90% confidence intervals. The regressions are unweighted (quantile regression does not support probability weights). Sample: agricultural households; same cleaning as the main table.

C.2.2 Alternative Treatment of Zero Income

Table C.1 examines the robustness of the main results to the treatment of zero-income households. In the main table, households with zero or negative equivalized income are excluded before taking logarithms. Here we retain them using two alternative transformations: columns (1)–(4) replace $\ln(y)$ with $\ln(y+1)$, and columns (5)–(8) use the inverse hyperbolic sine $\text{asinh}(y)$. Both transformations are defined at zero. Within each panel, the four specifications parallel the main table: full population, agricultural households, landowners (log area only), and landowners (log area and value). All other sample restrictions, survey fixed effects, and probability weights are identical to the main table.

Table C.1. Robustness: Alternative Treatment of Zero Income

	ln(Y+1)				asinh(Y)			
	Retains zero income				Retains zero income			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Agric. HH	2.189***	–	–	–	2.529***	–	–	–
	(0.215)				(0.242)			
Landowner	–	0.130	–	–	–	0.125	–	–
		(0.118)				(0.129)		
Log Land Area	–	–	0.296***	0.138***	–	–	0.294***	0.160***
			(0.041)	(0.046)			(0.040)	(0.049)
Log Land Value	–	–	–	0.325***	–	–	–	0.275***
				(0.042)				(0.052)
Survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	471	449	444	444	471	449	444	444
Surveys	16	16	16	16	16	16	16	16
Sample	Full pop	Agric	Owners	Owners	Full pop	Agric	Owners	Owners
Observations	156,168	77,983	58,083	58,083	156,663	78,139	58,215	58,215
Adj. R^2	0.208	0.133	0.137	0.160	0.201	0.115	0.116	0.130
Within R^2	0.1050	0.0002	0.0168	0.0436	0.1092	0.0001	0.0131	0.0283

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions include survey fixed effects and probability weights. Standard errors (in parentheses) are clustered at the survey–strata level. Columns (1)–(4) use $\ln(y+1)$ as the income measure, retaining zero-income households. Columns (5)–(8) use the inverse hyperbolic sine $\text{asinh}(y)$. All 16 income surveys as in the main table.

A potential concern is that some household surveys in our sample were not designed primarily for constructing income aggregates, particularly in Sub-Saharan Africa, where income is notoriously difficult to measure and self-reported figures may capture only part of household earnings. If this is the case, the estimated returns to land area and value could partly reflect survey design rather than true economic relationships. To address this concern, Table C.2 replaces income with consumption aggregates built directly by the LSMS survey providers themselves, who designed these measures specifically for welfare comparisons. The sample is restricted to agricultural landowners in the eight surveys that collect household consumption data: China (2012), Ethiopia (2018), Guatemala (2000), Malawi (2016), Niger (2014), Nigeria (2015), Panama (2008), and Tanzania (2014). Within each panel, columns (1)–(2) regress log income on log land area and log area plus log value;

columns (3)–(4) repeat the same two specifications with log equivalized consumption as the outcome. The left panel (columns 1–4) applies the main-table data clean (dropping zero-income households), while the right panel (columns 5–8) retains them: income in columns (5)–(6) uses $\ln(y + 1)$ so that zero-income households contribute to the income regression, and columns (7)–(8) use $\ln(c)$ on the same retained sample.

Table C.2. Robustness: Income and Consumption Returns to Land

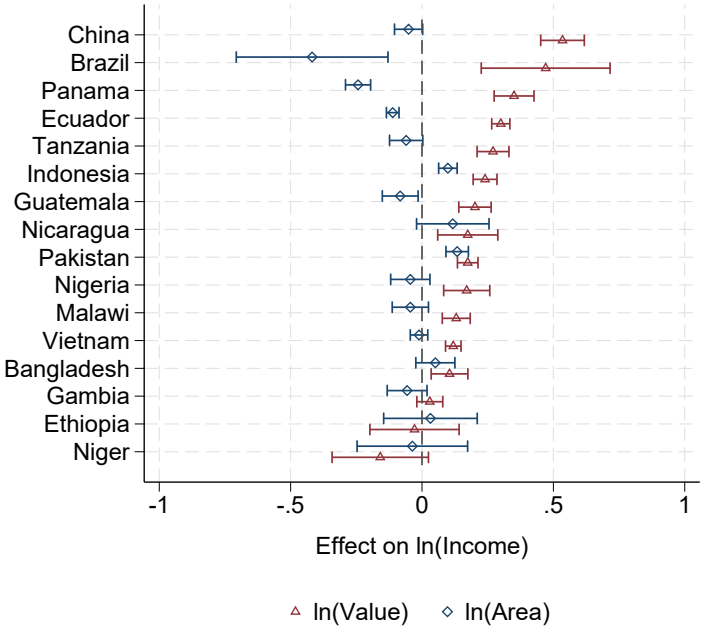
	Drop zeros				No drop zeros			
	ln(Income)		ln(Consumption)		ln(Y+1)		ln(Consumption)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Land Area	0.165***	-0.057*	-0.036	-0.058**	0.339***	0.171***	-0.035	-0.059***
	(0.036)	(0.031)	(0.025)	(0.024)	(0.051)	(0.056)	(0.023)	(0.023)
Log Land Value	–	0.506***	–	0.050***	–	0.371***	–	0.053***
		(0.035)		(0.008)		(0.045)		(0.008)
Survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	179	179	179	179	180	180	180	180
Surveys	8	8	8	8	8	8	8	8
Sample	Owners	Owners	Owners	Owners	Owners	Owners	Owners	Owners
Observations	19,738	19,738	19,738	19,738	23,722	23,722	23,722	23,722
Adj. R^2	0.032	0.217	0.412	0.416	0.036	0.068	0.446	0.450
Within R^2	0.0134	0.2019	0.0023	0.0089	0.0179	0.0503	0.0021	0.0095

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions include survey fixed effects and probability weights. Standard errors (in parentheses) are clustered at the survey–strata level. Sample: agricultural landowners in 8 surveys with consumption data: China2012, Ethiopia2018, Guatemala2000, Malawi2016, Niger2014, Nigeria2015, Panama2008, Tanzania2014. Consumption is equivalized identically to income (OECD scale, China: per capita). Columns (1)–(4) drop zero-income households before estimation. Columns (5)–(6) use $\ln(y + 1)$ retaining zero-income households; columns (7)–(8) retain them in the consumption regression sample.

C.2.3 Cross-Country Heterogeneity

Figure C.5 replicates the main specification for agricultural landowners, with zero-income households dropped, estimated separately for each of the 16 income surveys. Because wealth composition, agricultural structure, and land markets vary widely across countries, pooled fixed-effects estimates may mask substantial cross-country heterogeneity. The figure shows the OLS coefficient on log land value (open triangles, maroon) and log land area (open diamonds, navy) for each country, sorted by the land value coefficient. Despite heterogeneity in magnitudes, the land value coefficient is positive in the overwhelming majority of countries, confirming that the main result does not simply reflect the composition of the sample. The land area coefficient is small and often insignificant across most countries, mirroring the pooled finding.

Figure C.5. Land Returns by Country: Main Specification



Notes: Each point shows the OLS coefficient from regressing ln(income) on log land area and log land value for agricultural landowners within a single survey, using probability weights and the survey design. Horizontal bars: 95% confidence intervals. Surveys sorted by log land value coefficient (descending). Zero-income households dropped. Open triangle (▽) = ln(Value); open diamond (◇) = ln(Area).

C.2.4 Main Specification Variants

Table C.3 examines the sensitivity of the main land area and value estimates from Table 3.2 to five alternative data and sample choices. Column (1) includes non-titled land: households with non-titled plots are reclassified as landowners and their non-titled area and value are added to titled holdings. Column (2) retains observations with missing land area or land value flags rather than dropping them as in the main table. Column (3) uses income without any top-coding. Column (4) winsorizes income at the 95th percentile rather than the 99th. Column (5) uses total household income including transfers rather than income net of transfers.

Table C.3. Robustness: Main Specification Variants

	Non-titled	Miss. land/val.	No winsor.	Winsor. 95th	Incl. transfers
	(1)	(2)	(3)	(4)	(5)
Log Land Area	-0.030	-0.059**	-0.059**	-0.053**	-0.026
	(0.024)	(0.026)	(0.025)	(0.026)	(0.021)
Log Land Value	0.385***	0.445***	0.456***	0.439***	0.246***
	(0.037)	(0.034)	(0.032)	(0.034)	(0.018)
Survey FE	Yes	Yes	Yes	Yes	Yes
Clusters	442	441	442	439	442
Surveys	16	16	16	16	16
Sample	Owners	Owners	Owners	Owners	Owners
Observations	53,671	50,031	50,372	48,541	53,692
Adj. R^2	0.462	0.479	0.477	0.490	0.429
Within R^2	0.1440	0.1661	0.1716	0.1676	0.0597

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions include survey fixed effects and probability weights. Standard errors (in parentheses) are clustered at the survey–strata level. All columns estimate $\ln(\text{income}) \sim \ln(\text{land area}) + \ln(\text{land value})$ for agricultural landowners. Column (1): non-titled plots reclassified as owned, non-titled area and value added to totals. Column (2): retains observations with missing land area or value flags. Column (3): no winsorization of income. Column (4): income winsorized at the 95th percentile. Column (5): total household income including transfers.

C.2.5 Non-Response Analysis

Figure C.6 examines whether landowners with larger or more valuable holdings are more or less likely to have zero or missing income (non-response). The dependent variable is an indicator equal to one if a household reports zero or missing income. Panel A presents a pooled linear probability regression on within-country standardized log land area and log land value, with survey fixed effects, probability weights, and standard errors clustered at the survey–strata level. Panel B shows the same regression estimated country by country, using within-country standardized regressors and survey-weighted estimation, so each coefficient represents the effect of a one within-country standard deviation increase in log area or log value on the probability of non-response. The grey shaded band in Panel B marks the $[-0.01, +0.01]$ range: estimates falling within it imply that a one-standard-deviation increase in the land characteristic shifts the probability of non-response by less than one percentage point, which we treat as economically negligible. Non-response is especially pronounced in China and Panama, the two countries with the largest and most statistically significant biases. For these countries, the household surveys provide internally constructed income aggregates that impute income for non-reporting households. Panel C exploits these aggregates by regressing log survey-imputed income on log land area and log land value among landowners, without dropping zero-income observations. The resulting coefficients are consistent with the main findings, confirming that non-response does not drive the results.

Figure C.6. Non-Response and Standardized Land Returns

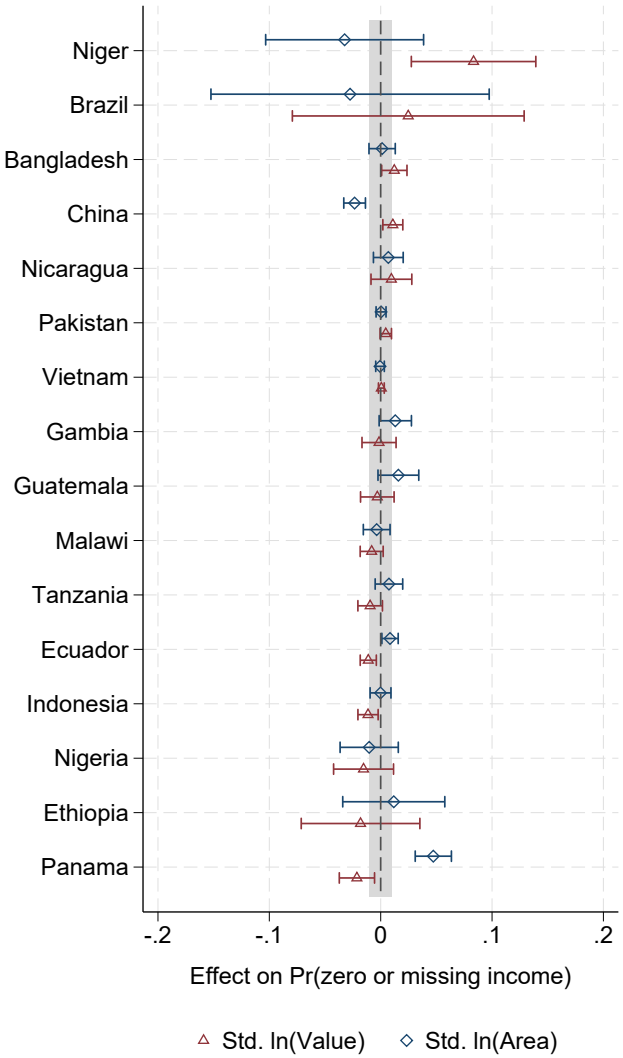
Panel A: Pooled Regression

	Pr(non-resp.)
Std. ln(Area)	-0.018***
	(0.003)
Std. ln(Value)	0.008***
	(0.003)
Survey FE	Yes
Clusters	444
Surveys	16
Sample	Owners
Observations	58,215
Adj. R^2	0.051
Mean dep. var.	0.078

Panel C: Survey-Imputed Income (China & Panama)

	China	Panama
Log Land Area	0.019	-0.142***
	(0.024)	(0.013)
Log Land Value	0.194***	0.256***
	(0.022)	(0.021)
Survey FE	No	No
Sample	Owners	Owners
Observations	7,453	2,062
Adj. R^2	0.038	0.100
Mean dep. var.	8.503	8.184

Panel B: Country-Level Estimates



Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Dependent variable: indicator for zero or missing income among agricultural landowners (Panels A–B); log survey-imputed income (Panel C). Panel A: survey fixed effects and probability-weighted OLS; standard errors clustered at survey–strata level; regressors standardized within each country. Panel B: survey-weighted OLS estimated separately for each country; regressors standardized within each country's owner sample; 95% confidence intervals shown. Open triangle (▽) = std. ln(value); open diamond (◇) = std. ln(area). Countries sorted by ln(value) coefficient (descending). Panel C: OLS with probability weights and robust standard errors; no survey fixed effects (single-country regressions). Outcome is log of the survey designers' own income aggregate, which imputes income for non-reporting households. Zero-income households are not dropped.

C.2.6 Poverty Lines

Table C.4 examines whether the estimated relationship between land holdings and poverty depends on the choice of poverty line or poverty definition. Columns (1)–(3) apply the main data clean (dropping zero-income households) and estimate the effect of log land area and log land value on the probability of per capita income falling below the USD 2.15, USD 3.65, and USD 6.85 per-capita per-day poverty lines (2017 PPP), respectively. Column (4) uses a binary poverty indicator constructed directly by the survey providers from household consumption aggregates rather than from our income measure; this alternative is available for seven surveys: Ethiopia (2018), Guatemala (2000), Malawi (2016), Nicaragua (2014), Niger (2014), Panama (2008), and Tanzania (2014). Columns (5)–(8) repeat the same four definitions retaining zero-income households (those households are classified as poor under all income-based thresholds, so the retained sample mechanically raises mean poverty rates but gives a complete picture of the population). Across all columns and poverty lines, the land value coefficient remains negative and highly significant, while the land area coefficient is small and imprecisely estimated once value is controlled for, consistent with the main results.

Table C.4. Robustness: Extensive Margin of Income Result: Poverty Lines

	Drop zeros				Retain zeros			
	USD 2.15 (1)	USD 3.65 (2)	USD 6.85 (3)	Other src. (4)	USD 2.15 (5)	USD 3.65 (6)	USD 6.85 (7)	Other src. (8)
Log Land Area	0.006	0.014**	0.015***	0.028***	-0.009	0.002	0.008*	0.026***
	(0.007)	(0.007)	(0.005)	(0.009)	(0.006)	(0.006)	(0.004)	(0.009)
Log Land Value	-0.105***	-0.080***	-0.044***	-0.059***	-0.081***	-0.062***	-0.034***	-0.051***
	(0.008)	(0.006)	(0.006)	(0.009)	(0.007)	(0.006)	(0.006)	(0.010)
Survey FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	441	441	441	146	444	444	444	147
Surveys	16	16	16	7	16	16	16	7
Sample	Owners	Owners	Owners	Owners	Owners	Owners	Owners	Owners
Observations	50,031	50,031	50,031	13,192	58,215	58,215	58,215	15,610
Adj. R^2	0.165	0.100	0.045	0.069	0.128	0.081	0.037	0.059
Mean dep. var.	0.608	0.718	0.844	0.496	0.663	0.758	0.866	0.481

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions: owners sample, 16 income surveys (cols 1–3, 5–7) or 7 surveys (cols 4, 8), survey FE, probability weights, cluster SEs at survey–strata level. Poverty lines (cols 1–3, 5–7): per-capita per-day income below USD 2.15, 3.65, or 6.85 (2017 PPP). Other sources (cols 4, 8): survey-provider consumption-based poverty indicator available for Ethiopia (2018), Guatemala (2000), Malawi (2016), Nicaragua (2014), Niger (2014), Panama (2008), Tanzania (2014).

C.2.7 Alternative Wealth Measures

The main table uses the inverse hyperbolic sine of wealth1, dwelling wealth plus plots and other assets, as the wealth outcome, which handles the large mass of zero-wealth households without dropping them. Table C.5 checks robustness across both wealth definitions and outcome transformations. Column (1) replicates the main table specification. Column (2) uses wealth2 (plot wealth plus other assets, excluding the dwelling), and column (3) uses other wealth only (financial and non-farm assets; composition varies by survey). Columns (4)–(6) use log levels, restricting each regression to the subsample with strictly positive wealth in that component, allowing direct comparison with the income log-level specification. The positive and significant land value coefficient is robust across all six columns; the land area coefficient is again small and often insignificant, mirroring the income results.

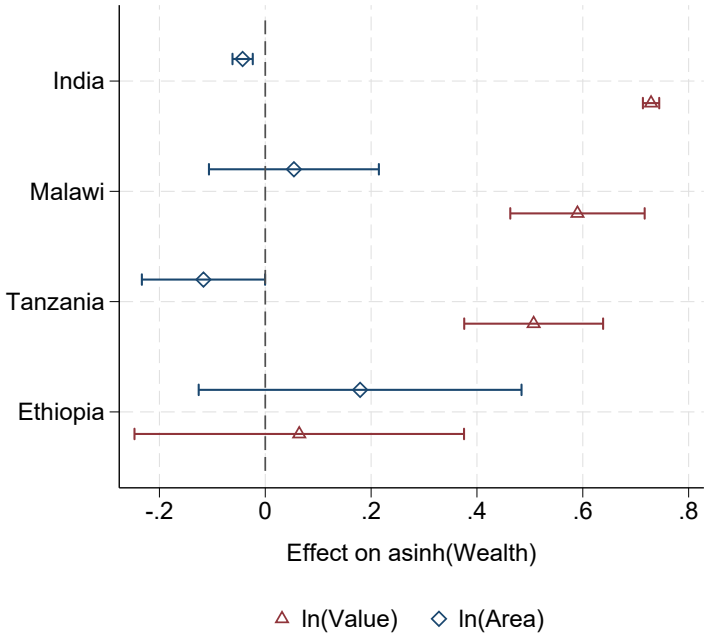
Table C.5. Robustness: Alternative Wealth Measures

	asinh(wealth)			ln(wealth), drop zero-wealth		
	Wealth 1	Wealth 2	Other	Wealth 1	Wealth 2	Other
	(1)	(2)	(3)	(4)	(5)	(6)
Log Land Area	-0.014	0.150***	0.126**	-0.010	0.103***	0.219***
	(0.030)	(0.044)	(0.061)	(0.023)	(0.031)	(0.049)
Log Land Value	0.622***	0.617***	0.339***	0.631***	0.637***	0.289***
	(0.050)	(0.062)	(0.051)	(0.046)	(0.054)	(0.042)
Survey FE	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	175	175	175	175	175	174
Surveys	4	4	4	4	4	4
Sample	Owners	Owners	Owners	Owners	Owners	Owners
Observations	34,404	34,404	34,404	34,102	33,752	32,405
Adj. R^2	0.595	0.626	0.559	0.740	0.771	0.305
Within R^2	0.304	0.288	0.082	0.541	0.573	0.170
Mean dep. var.	11.227	10.813	8.341	10.634	10.329	8.163

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions: agricultural landowners in 4 wealth surveys (Ethiopia 2018, India 2012, Malawi IHPS 2016, Tanzania 2019); survey FE, probability weights, cluster SEs at survey–strata level. Wealth 1 = dwelling + plots + other assets (used in main table). Wealth 2 = plot wealth + other assets (excludes dwelling). Other wealth = financial and non-farm assets only (composition varies by survey). Columns (4)–(6) restrict to households with strictly positive wealth in that component.

Because wealth components are measured differently across surveys, particularly the non-farm and financial asset categories, Figure C.7 shows the same main-table specification estimated separately for each of the four wealth surveys. The land value coefficient is consistent in sign and magnitude across all countries except Ethiopia, confirming that the pooled result is not driven by any single survey. Ethiopia appears as an outlier in the wealth results, but this is also the case for income, as shown in Figure C.5.

Figure C.7. Effects of Land and Value on Wealth by Country



Notes: Each point shows the OLS coefficient from regressing asinh(wealth) on log land area and log land value within a single wealth survey, using probability weights and the survey design. Wealth is total household wealth (dwelling + plots + other assets). Horizontal bars: 95% confidence intervals. Countries sorted by log land value coefficient (descending). Open triangle (▽) = ln(Value); open diamond (◇) = ln(Area).

C.2.8 Landless Analysis

Table 3.2, column (2) shows that the landowner dummy is not statistically significant for income among agricultural households. This section explores two dimensions of this result. First, we interact the owner dummy with an indicator for countries where agriculture accounts for a large share of value added at the time of the survey, labelled "agricultural countries." If landowners earn more than the landless primarily in agrarian economies, the pooled owner coefficient may mask this heterogeneity. Second, we replace log income with the share of agricultural income in total household income. Even if ownership does not predict total income, it may shift the composition of income toward agriculture.

Table C.6 defines agricultural countries as the three surveys with the highest agricultural share of value-added according to the World Bank at the time of the survey: Tanzania (2014), Ethiopia (2018), and Niger (2014). Column (1) replicates the main table owner regression with equivalized income. Column (2) adds the interaction of the owner dummy with the agricultural-country indicator. Columns (3)–(4) use the share of agricultural income in total household income as the outcome.

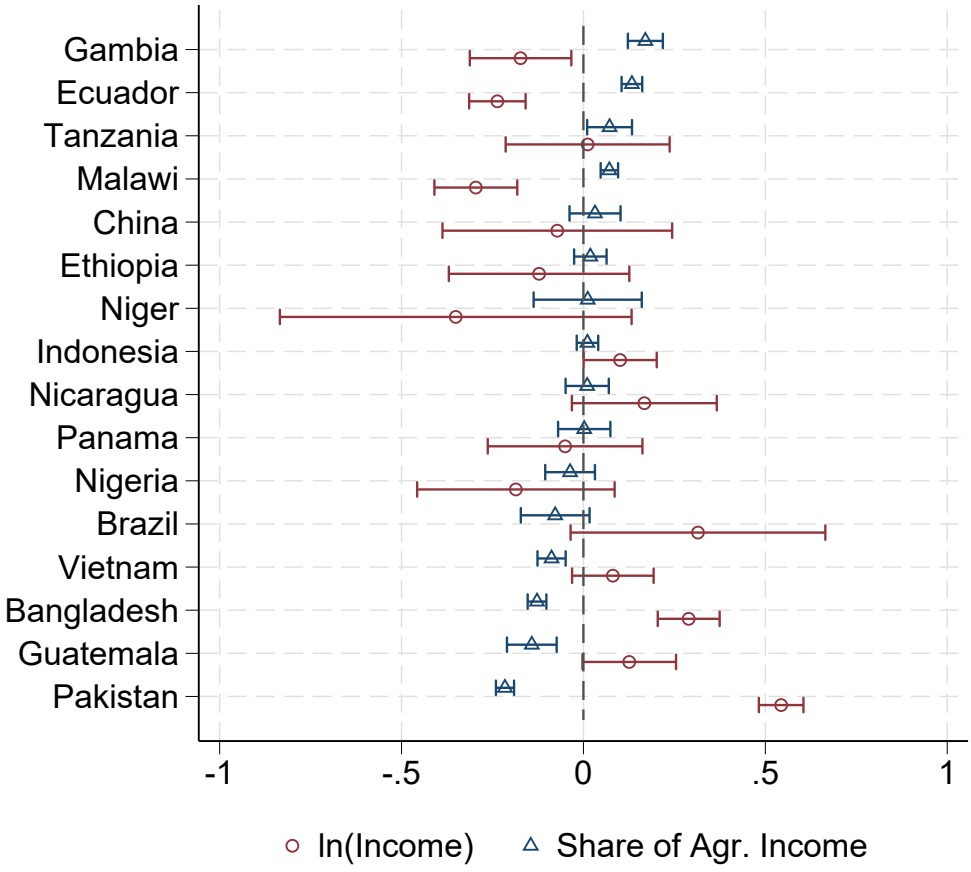
Table C.6. Landless Analysis: Ownership, Income, and Agricultural Share

	ln(Income)		Agr. Income Share	
	(1)	(2)	(3)	(4)
Owner	0.057	0.077	-0.024	-0.032*
	(0.067)	(0.075)	(0.016)	(0.017)
Owner × Agric. country	–	-0.163*	–	0.068***
		(0.098)		(0.023)
Survey FE	Yes	Yes	Yes	Yes
Clusters	447	447	447	447
Surveys	16	16	16	16
Sample	Agric	Agric	Agric	Agric
Observations	67,409	67,409	67,409	67,409
Adj. R^2	0.380	0.380	0.079	0.079
Within R^2	0.0001	0.0002	0.0003	0.0006

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. All regressions include survey fixed effects and probability weights. Standard errors (in parentheses) are clustered at the survey–strata level. Sample: agricultural households in all 16 income surveys. Columns (1)–(2): equivalized log income (same cleaning as main table). Columns (3)–(4): share of agricultural income in total household income (non-equivalized). Missing or negative agricultural income is imputed to zero before computing shares. *Agric. country* (devt = 1): Tanzania (2014), Ethiopia (2018), Niger (2014) — the three surveys with the highest agricultural share of value added.

Figure C.8 plots the country-by-country coefficient on the landowner dummy for both outcomes. The navy triangles show the effect on the share of agricultural income, while the maroon circles show the effect on log income. Horizontal bars indicate 95% confidence intervals.

Figure C.8. Landowner–Landless Gap by Country: Income and Agricultural Share



Notes: Each point shows the survey-weighted OLS coefficient on the landowner dummy from a country-level regression of the outcome on a landowner indicator. Maroon circles: ln(Income), equivalized using the OECD-modified scale (China: per capita). Navy triangles: share of agricultural income in total household income (non-equivalized). Horizontal bars show 95% confidence intervals. Sample: agricultural households; same cleaning as the main table.

C.3 Land Wealth Inequality

C.3.1 Elasticities, Area Inequality, and the Covariance Decomposition

This appendix provides additional evidence on the inverse value–size relationship and clarifies how it shapes the covariance term in Section 3.4. The analysis is motivated by the variance identity

$$\mathrm{Var}(\log V) - \mathrm{Var}(\log A) = \mathrm{Var}(\log P) + 2\,\mathrm{Cov}(\log A, \log P),$$

which shows that differences between land-value and land-area inequality arise from two components: price dispersion, $\mathrm{Var}(\log P)$, and the covariance between land size and land prices. The covariance term itself can be written as $\mathrm{Cov}(\log A, \log P) = \beta\,\mathrm{Var}(\log A)$, where β is the elasticity of land prices per hectare with respect to landholding size.

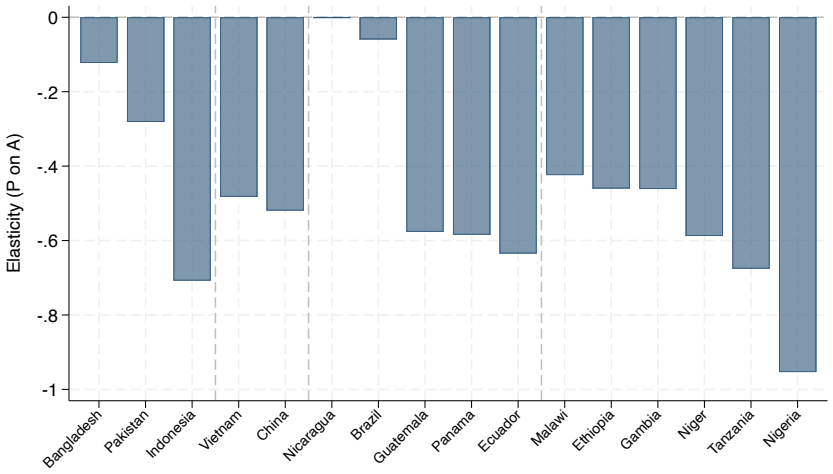
Elasticities across countries. Figure C.9a reports the elasticity of land prices per hectare with respect to landholding size (β) for each country. In all cases, the elasticity is negative, indicating that larger landholdings command lower prices per hectare. While the magnitude varies across countries, the inverse price gradient is a consistent feature of the data.

Elasticity and initial land-area inequality. Figure C.9b plots country-level elasticities against initial land-area inequality, measured by $\mathrm{Var}(\log A)$. Although elasticities are uniformly negative, there is no strong systematic relationship between the steepness of the inverse gradient and area inequality. Countries with highly unequal land distributions do not exhibit systematically more negative elasticities. This indicates that cross-country differences in the covariance term are unlikely to be driven primarily by variation in the slope of the inverse value–size relationship.

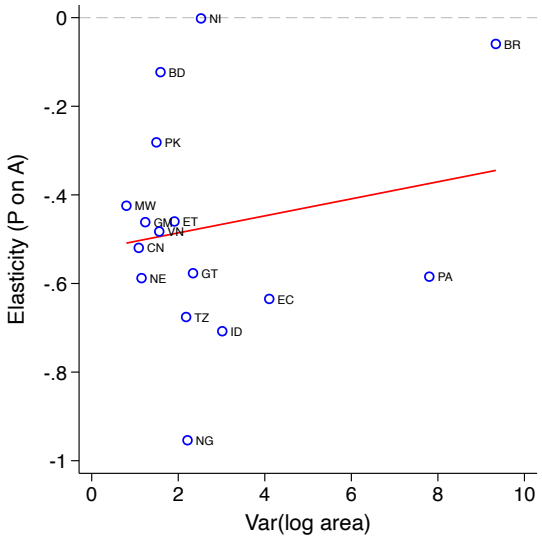
Price dispersion and land-area inequality. The variance identity also highlights a second potential channel: dispersion in land prices. Figure C.9c plots $\mathrm{Var}(\log P)$ against $\mathrm{Var}(\log A)$. While price dispersion is positively related to land-area inequality, the relationship is comparatively weak and considerably noisier than the tight association between the covariance term and land-area inequality documented in Figure 3.2b in the main text. This suggests that variation in price dispersion alone cannot account for the strong equalizing effect observed when moving from area- to value-based inequality measures.

Decomposing the covariance term. Given that $\mathrm{Cov}(\log A, \log P) = \beta\,\mathrm{Var}(\log A)$, cross-country differences in the covariance reflect differences in both elasticities and land-area dispersion. Table C.7 implements an exact Shapley decomposition of the between-group difference in the covariance effect shown in Figure 3.2c in the main paper. The results

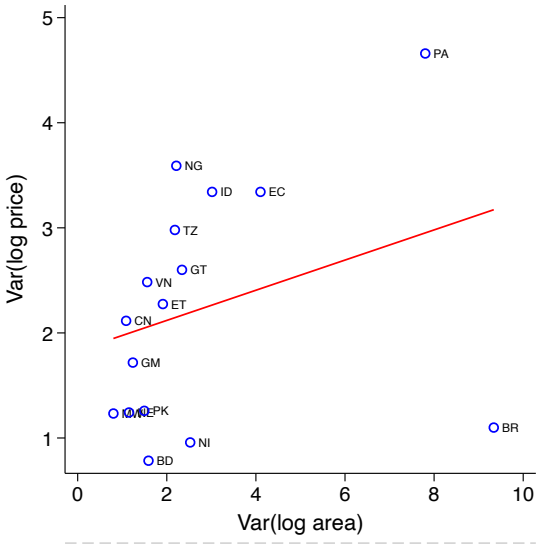
Figure C.9. Elasticities, land-area inequality, and price dispersion



(a) Elasticity of land prices with respect to land size



(b) Elasticity and land-area inequality



(c) Price dispersion and land-area inequality

Notes: Panel (a) reports country-level estimates of the elasticity of log land prices per hectare with respect to log land size. Panel (b) plots these elasticities against land-area inequality, measured by $\text{Var}(\log A)$. Panel (c) plots price per hectare dispersion, $\text{Var}(\log P)$, against $\text{Var}(\log A)$.

show that the more negative covariance in the descending group is driven primarily by higher dispersion in landholdings, with a secondary contribution from slightly more negative elasticities. Differences in the cross-country covariance between β and $\text{Var}(\log A)$ partially offset these mean effects. Overall, dispersion in land size accounts for the majority of the difference in the covariance effect across groups.

Is the inverse value–size relationship driven by productivity differences? An important remaining question is whether the negative elasticities documented above simply reflect the

Table C.7. Shapley decomposition of the covariance effect

	Descending	Ascending
Mean elasticity (β)	-0.5864	-0.3811
Mean Var(log A)	4.2810	1.5991
Mean 2Cov(log A, log P)	-3.9983	-1.1565
Between-group difference (Ascending − Descending)		
Total difference	2.8418	
Elasticity component	1.2072	
Area inequality component	2.5950	
Cov(β,VarA) component	-1.1231	

Notes: Countries are grouped into *descending* (where $\mathrm{Var}(\log V) - \mathrm{Var}(\log A) < 0$) and *ascending* (where the change is positive). β denotes the elasticity of log land prices with respect to log land size, and $\mathrm{Var}(\log A)$ is initial land-area inequality. The covariance term entering the variance identity can be written as $2\,\beta\,\mathrm{Var}(\log A)$. The lower panel reports an exact Shapley decomposition of the between-group difference in the mean covariance effect. The elasticity component captures differences in mean elasticities holding dispersion fixed at the midpoint; the area inequality component captures differences in mean land-area dispersion holding elasticities fixed; and the $\mathrm{Cov}(\beta, \mathrm{Var}(\log A))$ component reflects cross-country comovement between elasticities and dispersion within groups.

well-known inverse relationship between farm size and productivity. If smaller farms are more productive, they may command higher land values per hectare, mechanically generating a negative elasticity of land prices with respect to land area.

Table C.8 evaluates this channel in two steps. First, Column 2 restricts the sample to households with strictly positive agricultural income (Drop 0), ensuring comparability with specifications that incorporate productivity controls. The resulting elasticities are very similar to those obtained in Column 1 using the full sample, indicating that the sample restriction itself does not materially affect the inverse value–size relationship.

Second, Column 3 introduces a direct control for agricultural productivity, measured as farm income per hectare.[1] Controlling for productivity attenuates the inverse elasticity, but only partially. The unweighted mean elasticity declines from −0.4584 in Column 2 to −0.3708 in Column 3, a reduction of roughly 19 percent. In the vast majority of countries, the coefficient remains negative and economically significant.

These results indicate that while productivity differences account for part of the inverse value–size relationship, a substantial component remains unexplained. Consequently, the negative covariance term that compresses land-value inequality relative to land-area inequality

[1] Agricultural income is defined as farm income plus wage farm income. Farm income can in turn be decomposed into crop and animal income. Our results are nearly identical when we use agricultural income or crop income instead of farm income.

Table C.8. Country-level elasticity robustness checks

	Elasticity estimates				Sample size (N)	
Country	Baseline	Drop 0	Control prod.	SE	Baseline	Drop 0
Bangladesh	-0.1230	-0.1109	-0.0937	0.0173	3049	1623
Brazil	-0.0593	0.0746	0.1158	0.0410	452	283
China	-0.5196	-0.5987	-0.1581	0.0079	8218	5787
Ecuador	-0.6348	-0.5989	-0.4446	0.0113	8812	5614
Ethiopia	-0.4600	-0.5022	-0.5207	0.0411	1535	935
Gambia	-0.4647	-0.4740	-0.4412	0.0195	7306	4060
Guatemala	-0.5769	-0.5951	-0.4851	0.0296	1931	944
Indonesia	-0.7077	-0.6913	-0.5252	0.0195	3901	2723
Malawi	-0.4245	-0.3747	-0.3353	0.0216	7035	3231
Nicaragua	-0.0017	-0.0187	0.0128	0.0288	631	548
Niger	-0.5879	-0.6919	-0.7714	0.0776	1098	235
Nigeria	-0.9539	-0.9191	-0.8860	0.0322	2220	893
Pakistan	-0.2815	-0.2513	-0.0853	0.0247	3698	2078
Panama	-0.5846	-0.4913	-0.3583	0.0243	2080	705
Tanzania	-0.6757	-0.6211	-0.5321	0.0329	1616	1100
Vietnam	-0.4826	-0.4699	-0.4235	0.0204	5664	4045
Unweighted mean	**-0.4712**	**-0.4584**	**-0.3708**	**0.0281**		

Notes: Entries report country-level elasticities of log land prices per hectare with respect to log landholding size. Baseline estimates use all landowning households. Drop 0 restricts the sample to households with strictly positive agricultural income. Control prod. additionally controls for agricultural productivity, measured as agricultural income per hectare. All specifications are estimated using household sampling weights. SE reports the standard error of the elasticity in the productivity-controlled specification. The controlled regression is estimated on the Drop-0 sample (so its number of observations equals the Drop-0 column by construction). The final row reports the unweighted average across countries.

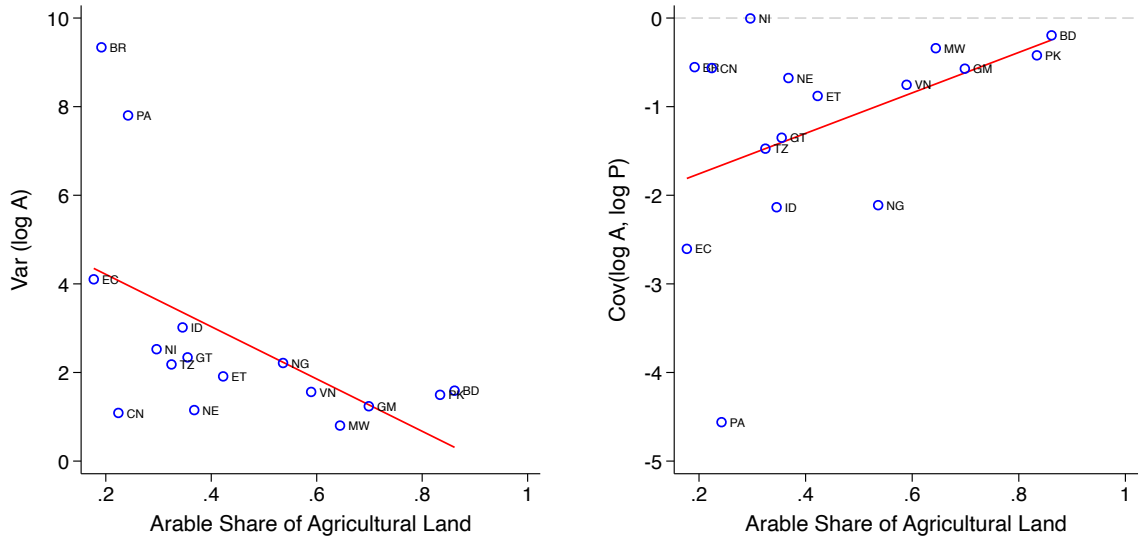
in Section 3.4 cannot be interpreted as merely a reflection of the inverse size–productivity relationship. Land prices are influenced by additional factors, such as land quality heterogeneity, location, infrastructure access, and market conditions, that could drive higher prices per hectare for smaller holdings.

C.3.2 Land use composition and land-area dispersion

This appendix provides suggestive evidence on whether cross-country differences in agricultural land use are correlated with the dispersion of landholdings. The motivation is the variance identity in Section 3.4.2, where the covariance term depends both on the elasticity of land prices with respect to land size and on the dispersion of land sizes. Using FAO country-level data aligned with the survey dates, we construct the share of arable land in total agricultural land.

Figure C.10 shows that countries with a higher arable share tend to have lower land-area inequality, measured by $\mathrm{Var}(\log A)$, and a less negative covariance between land size and land prices, $\mathrm{Cov}(\log A, \log P)$. By contrast, arable share exhibits no systematic relationship with the elasticity of land prices with respect to land size (not shown). Taken together, these patterns suggest that land-use structure is primarily associated with cross-country differences in land-area dispersion, and thus with the magnitude of the covariance effect, rather than with differences in the slope of the inverse value–size gradient itself.

Figure C.10. Land use composition and the covariance mechanism



(a) Arable share and land-area inequality

(b) Arable share and the covariance term

Notes: The horizontal axis reports the FAO share of arable land in total agricultural land. Panel (a) plots land-area inequality $\mathrm{Var}(\log A)$. Panel (b) plots the covariance between land size and land prices, $\mathrm{Cov}(\log A, \log P)$. Each dot is a country.

C.3.3 Crop versus Animal Income Mix and Land Area Dispersion

This appendix provides household-level evidence on the mechanism linking agricultural land use to the dispersion of landholdings documented in Appendix C.3.2. We exploit within-country variation in production patterns by classifying landowning households according to the share of net crop income in total farm income, which serves as a micro-level analog of the FAO arable share.

We group households into four categories: pure animal (crop share = 0; 24% of the sample), animal-dominant (crop share in $(0, 0.5)$; 9%), crop-dominant (crop share in $[0.5, 1)$; 14%), and pure crop (crop share = 1; 53%). The analysis includes 33,457 landowning households with strictly positive farm income across 14 countries.

Panel (a) of Figure C.11 plots the variance of log land area, $\mathrm{Var}(\log A)$, by crop-income group. Two patterns stand out. First, dispersion declines sharply with crop intensity: $\mathrm{Var}(\log A)$ falls from 3.30 among pure-animal households to 1.49 among pure-crop households, a decline of more than 50%. Second, by contrast, average log land area varies little and shows no monotonic relationship with crop intensity.[2] The contrast between a strong gradient in dispersion and a flat relationship in the mean indicates that the crop–animal distinction operates through *who occupies the upper tail* of the farm-size distribution, rather than through shifts in typical farm size.

To confirm that this dispersion gradient is not driven by cross-country composition, we regress the squared deviation from the country mean, $(\log A - \overline{\log A}_c)^2$, on crop share and country fixed effects. The coefficient is -0.49 ($p< 0.01$). Restricting the sample to pure-animal and pure-crop households, the coefficient on the pure-crop indicator is -0.46 ($p< 0.01$).

The dispersion effect is concentrated in the upper tail. A one-unit increase in crop share reduces the probability of being in the top decile of land area (within country) by 4.35 percentage points ($p< 0.01$), while it has no statistically meaningful effect on the probability of being above the median (marginal effect = 0.017, p=0.41). Estimating the two equations jointly, we reject equality of the effects (chi-squared(1) = 59.4, $p< 0.01$). This asymmetry indicates that animal-oriented production is associated with the presence of very large landholdings at the top of the distribution, rather than with a general rightward shift of the land-size distribution.

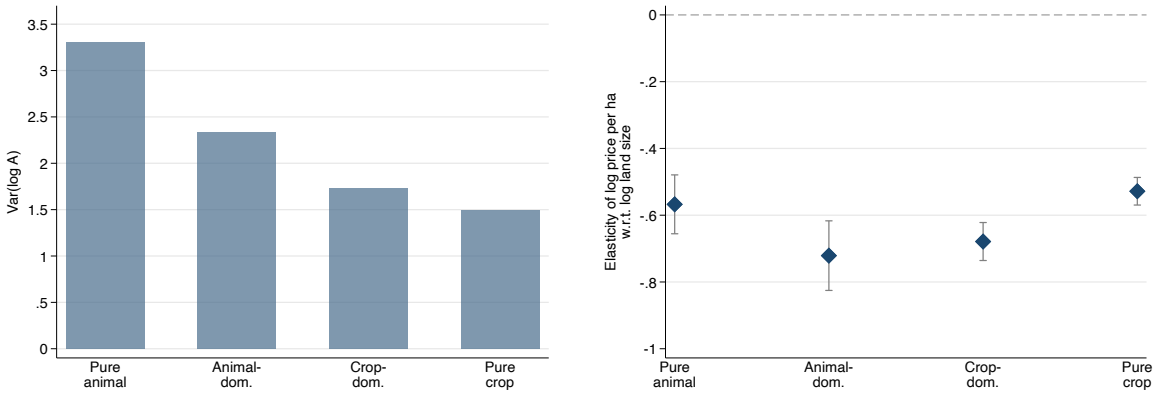
Panel (b) of Figure C.11 reports the elasticity of land price per hectare with respect to land size, estimated separately within each group. The elasticities range from -0.53 (pure crop) to -0.72 (animal-dominant), with overlapping confidence intervals and no monotonic pattern. The inverse price gradient is present across all groups and does not vary systematically with the crop–animal mix, consistent with the interpretation that land-use composition shapes land

[2] In a weighted regression of $\log A$ on crop share and country fixed effects, the coefficient on crop share is 0.017 (p=0.72).

inequality primarily through dispersion in farm sizes rather than through changes in the slope of the price gradient. This is also in line with the Shapley decomposition in Table C.8, which attributes cross-country differences in the covariance term mainly to differences in $\mathrm{Var}(\log A)$ rather than to differences in the elasticity.

Overall, the micro evidence points to a specific mechanism: pastoral and mixed animal–crop production is associated with polarization in farm sizes, with very large ranching or pastoral operations coexisting with small-scale herders, thereby inflating $\mathrm{Var}(\log A)$ through the upper tail. This pattern helps interpret the cross-country relationship between the FAO arable share and land-area inequality in Appendix C.3.2. At the same time, average crop income shares within countries are essentially uncorrelated with country-level $\mathrm{Var}(\log A)$ ($r = -0.02$), suggesting that what varies across countries is not the income mix of the typical household but the scale of large pastoral and ranching operations, a feature the FAO land-use measure captures more directly than household income shares.

Figure C.11. Crop–animal income mix and the land-inequality mechanism



(a) $\mathrm{Var}(\log A)$ by crop-income group

(b) Elasticity of log price per ha w.r.t. log land size

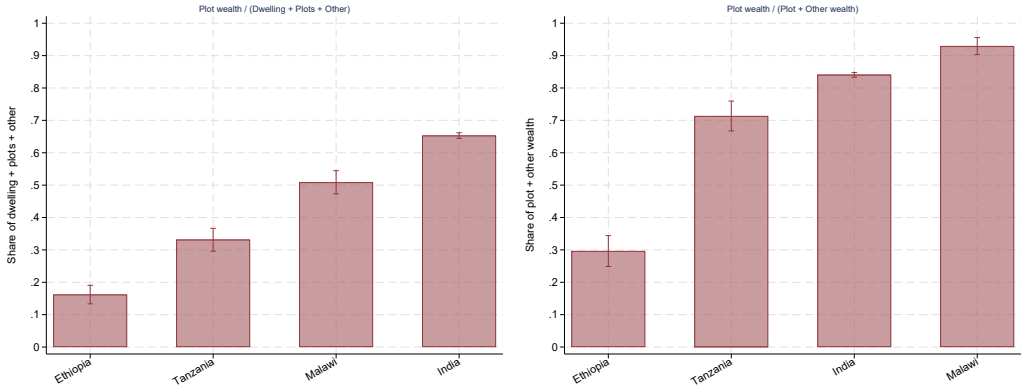
Notes: The sample includes landowning households in 14 countries with strictly positive farm income (net crop income and/or net animal income). Households are grouped by the share of net crop income in total farm income: pure animal (crop share = 0), animal-dominant (crop share in $(0, 0.5)$), crop-dominant (crop share in $[0.5, 1)$), and pure crop (crop share = 1). Panel (a) plots the weighted variance of $\log A$ within each group. Panel (b) plots the coefficient on $\log A$ from separate weighted OLS regressions of $\log P$ on $\log A$ within each group, controlling for country fixed effects; whiskers denote 95% confidence intervals based on robust standard errors. All estimates use survey sampling weights.

C.4 Land Value, Asset Poverty, and Minimum Endowments

C.4.1 Land as a Share of Household Wealth

Figure C.12 decomposes wealth by showing the survey-weighted share of total household wealth attributable to plot wealth (the wealth module's own plot valuation). The left panel expresses plot wealth as a share of Wealth 1 (total household wealth); the right panel expresses it as a share of Wealth 2 (plot wealth plus other assets, excluding the dwelling).

Figure C.12. Plot Wealth as a Share of Total Wealth, by Country



Notes: Each bar shows the survey-weighted mean share of household wealth attributable to plot wealth (from the wealth survey module), with 95% confidence intervals. Countries sorted ascending by share within each panel. *Left panel*: plot wealth as a fraction of total wealth (dwelling + plots + other assets). *Right panel*: plot wealth as a fraction of plot wealth + other assets (excluding the dwelling). Sample: agricultural landowners in the four wealth surveys.

C.4.2 Income Poverty and Land Poverty: Alternative Specifications

This appendix describes the set of linear probability models (LPM) used to link land poverty to income poverty among agricultural households. In the main text, we report results from a pooled specification without survey fixed effects, where income poverty is defined using the international US$2.15 (PPP) per day line and households are grouped into three mutually exclusive categories: landless, land-poor landowners, and non-poor landowners. Here we document additional specifications that clarify the role of cross-survey heterogeneity and assess robustness to alternative land-poverty thresholds.

Three-group specification (no survey fixed effects). Our baseline pooled model estimates income poverty as a function of indicators for land-poor and non-poor landowners, with landless households as the omitted category:

$$y_{is} = \alpha + \beta \mathbf{1}\{\text{land-poor}_{is}\} + \gamma \mathbf{1}\{\text{non-poor}_{is}\} + u_{is},$$

where y_{is} is an indicator for income poverty of household i in survey s. Observations are weighted using household survey weights and standard errors are clustered at the survey level. The coefficient α captures the poverty rate among landless households, while $\alpha + \beta$ and $\alpha + \gamma$ correspond to land-poor and non-poor landowners, respectively. The associated three-bar figure (Figure 3.3a) displays these predicted poverty levels.

The results show a clear monotonic gradient: land-poor landowners exhibit the highest poverty rates, followed by landless households, with non-poor landowners displaying the lowest incidence. The differences are economically meaningful in magnitude.

Three-group specification with survey fixed effects. To account for cross-survey differences in overall poverty levels and institutional environments, we augment the baseline model with survey fixed effects:

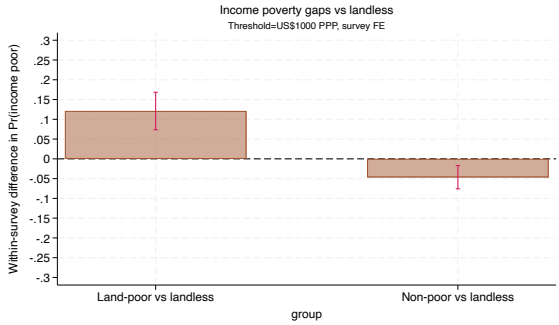
$$y_{is} = \alpha_s + \beta \mathbf{1}\{\text{land-poor}_{is}\} + \gamma \mathbf{1}\{\text{non-poor}_{is}\} + u_{is}.$$

The survey fixed effects α_s absorb survey-level heterogeneity. The coefficients β and γ therefore measure within-survey differences relative to landless households.

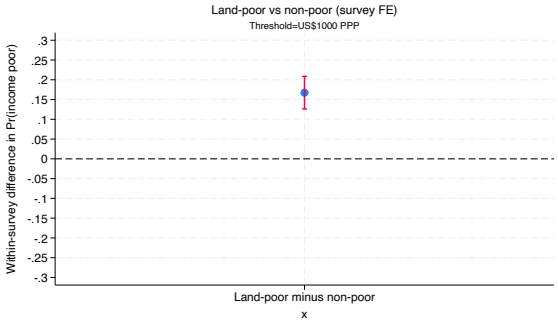
Panel (a) of Figure C.13 reports the predicted income-poverty rates for landless households, land-poor landowners, and non-poor landowners implied by this fixed-effects specification at the benchmark US$1,000 threshold. The ordering across groups is unchanged relative to the baseline pooled specification in the main text: land-poor landowners exhibit the highest poverty rates, followed by landless households, with non-poor landowners displaying the lowest incidence. Moreover, the magnitudes of the differences are quantitatively very similar

to those in the baseline no–fixed-effects specification, indicating that the main pattern is not driven by cross-country differences in average poverty levels.

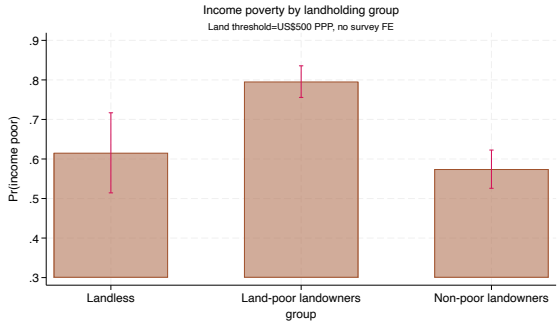
Figure C.13. Income poverty by landholding group: within-survey gaps and robustness to alternative thresholds



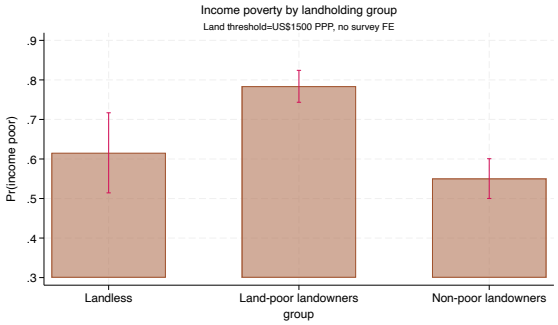
(a) Three-group specification (US$1,000, survey FE)



(b) Within-survey gap: land-poor minus non-poor (US$1,000)



(c) Alternative threshold (US$500, no survey FE)



(d) Alternative threshold (US$1,500, no survey FE)

Notes: Panel (a) reports predicted income-poverty rates for landless households, land-poor landowners, and non-poor landowners from a three-group specification with survey fixed effects at the benchmark land-value threshold (US$1,000 PPP) with the landless as the omitted group. Panel (b) displays the within-survey contrast between land-poor and non-poor landowners ($\beta - \gamma$) from the same fixed-effects specification, shown as a point estimate with a 95% confidence interval. Panels (c) and (d) replicate the pooled three-group specification without survey fixed effects using alternative land-value thresholds (US$500 and US$1,500 PPP). All regressions use household survey weights and cluster standard errors at the survey level. Income poverty is defined using the international US$2.15 (PPP) per day line.

Within-survey contrast between land-poor and non-poor landowners. We focus in particular on the linear combination $\beta - \gamma$, which directly tests whether land-poor and non-poor landowners differ within surveys. This contrast isolates inequality within the landowning population and removes cross-country heterogeneity. Panel (b) of Figure C.13 reports this estimate as a point estimate with a 95% confidence interval. At the benchmark US$1,000 threshold, the gap is positive, economically large, and statistically significant. Its magnitude is close to that observed in the pooled three-group specification, reinforcing the conclusion that the relationship between land poverty and income poverty reflects within-country differences rather than compositional effects.

Variation across land-value thresholds. All specifications are estimated for alternative land-value thresholds (US$500, US$1,000, and US$1,500 PPP) that define the land-poor category.

To avoid excessive duplication of figures, we report in the appendix only the pooled three-group specification for the alternative thresholds (Figures C.13c and C.13d). The qualitative pattern remains unchanged across thresholds: land-poor landowners consistently display the highest income poverty rates, followed by landless households, with non-poor landowners exhibiting the lowest poverty incidence.

The corresponding fixed-effects results for the US$500 and US$1,500 thresholds yield coefficients that are quantitatively very similar to those reported for the benchmark US$1,000 threshold. In particular, the within-survey gaps closely mirror the pooled differences, reinforcing the conclusion that the relationship between land poverty and income poverty is robust to alternative threshold definitions and to the inclusion of survey fixed effects. Additional results are available upon request.

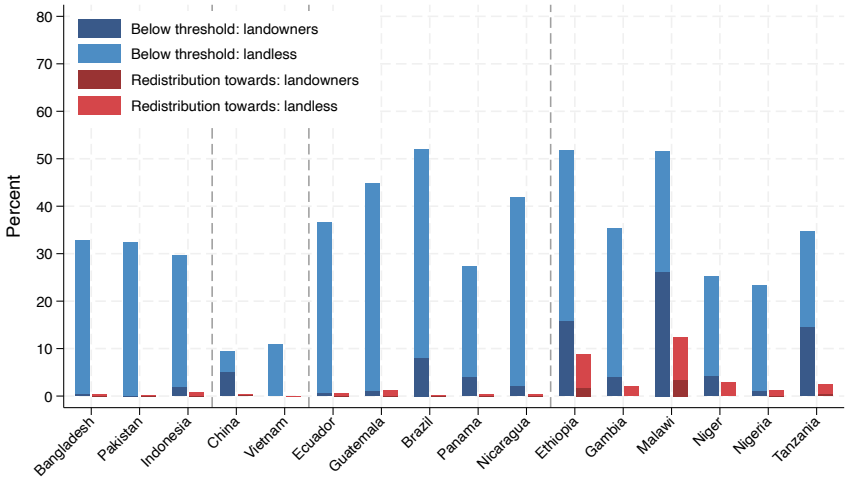
C.4.3 Robustness: Land Poverty and Redistribution Shares

This subsection assesses the robustness of the land-poverty and redistribution measures reported in Section 3.5 along two dimensions: the land-value threshold and the definition of agricultural households.

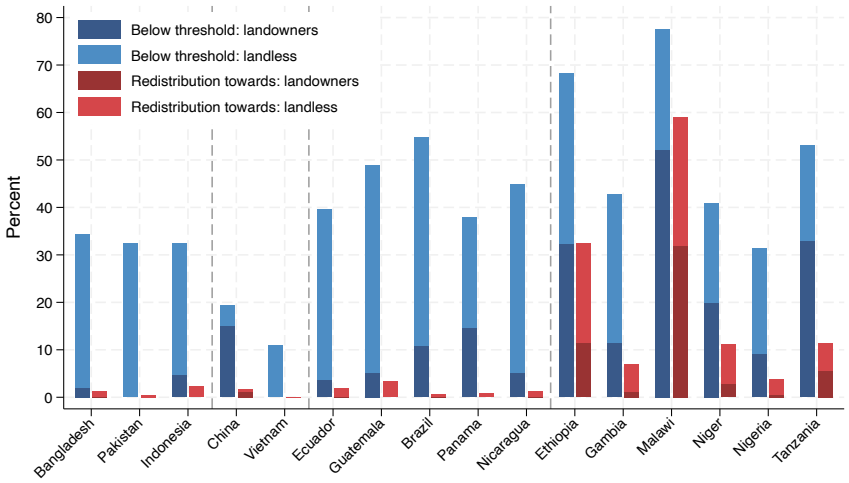
Figure C.14 presents two panels. Each panel plots, by country, (i) the share of agricultural households with land value below a given threshold ("Below threshold") and (ii) the land value required to lift all households to the threshold, expressed as a percentage of aggregate land value ("Redistribution"). Bars are stacked to distinguish the contributions of landowners and landless households.

Panels (a) and (b) report results for alternative thresholds of US$500 and US$1,500 (PPP). As expected, higher thresholds mechanically increase both the incidence of land poverty and the implied redistribution share. Cross-country rankings, on the other hand, remain largely unchanged.

Figure C.14. Land poverty and potential redistribution: alternative thresholds



(a) Broad definition, Threshold = US$500 PPP



(b) Broad definition, Threshold = US$1500 PPP

Notes: Each panel plots, by country, (i) the share of agricultural households below the land-value threshold ("Below threshold") and (ii) the total land value required to bring all households up to the threshold, expressed as a share of aggregate land value ("Redistribution"). Bars are stacked by household type (landowners vs. landless). Panels A and B report results for alternative land value-poverty thresholds: US$500 and US$1,500 (PPP). Landless households are assigned zero land value in all cases.

C.5 Definitions

C.5.1 Ownership

Table C.9. Ownership questions and definitions by country

Country	Ownership question	Ownership definition
Bangladesh	How was the plot acquired?	Purchased/bought; Inherited (wife's or husband's family)
Brazil	If reported land-owned is non-zero	N/A
China	The amount collectively owned land allocated to your household	N/A
Ecuador	How was the plot acquired?	Purchased; Inherited; Awarded; Usufruct
Ethiopia	How did your household acquire this parcel?	Gift/Inherited; Purchased
Gambia	How did the household acquire this parcel of land?	Inherited; Purchased
Guatemala	What ownership documents do you have?	Registered title deed; Title deed pending; Inheritance (pending); Public deed
India	If reported land-owned is non-zero	N/A
Indonesia	Do you have land for farming?	Yes
Malawi	How did your household acquire this plot?	Inherited by the death of a family member; Bride price; Purchased; Allocated by family member
Nicaragua	If answered to module on information on farms or plots owned by a member of the household	N/A
Niger	Mode d'acquisition de cette parcelle / Way of acquiring this plot	Purchased; Inherited
Nigeria	How was the plot acquired?; Have you or any other household member acquired a legal title	Outright purchase; Family inheritance; Yes
Pakistan	If reported land-owned is non-zero	N/A
Panama	If reported land-owned is non-zero	N/A
Tanzania	How was this plot acquired?	Inheritance; Gift; Purchased
Vietnam	If reported land-owned is non-zero	N/A

C.5.2 Income Aggregate

This section describes the harmonized protocol used to construct total household income from survey microdata across all countries in our sample. Total household income is defined as:

$$\textit{Total Income} = \textit{Agricultural Income} + \textit{Non-Agricultural Income} + \textit{Non-Labour Income}$$

where each component is constructed as follows:

- *Agricultural Income* = Own Farm Income + Wage Farm Income + Land Rent Income. Own farm income is net income from crop, livestock, forestry, and fishing production on privately held land. Wage farm income is earnings from agricultural wage work. Land rent income is payments received for renting out owned agricultural land.
- *Non-Agricultural Income* = Non-Agricultural Business Income + Non-Agricultural Wage Income. Business income is net self-employment or employer income in non-agricultural sectors. Wage income is earnings from non-agricultural wage employment.
- *Non-Labour Income* includes pensions, dividends, remittances, cash transfers, rental income from non-farm assets, and scholarships.

Several harmonization rules apply across all countries. Survey responses coded as "don't know" or similar are recoded to missing. All incomes are annualized. Business incomes are net of costs where revenue data are available; where missing, independent income from the labour module is substituted. Negative own-farm or business incomes are set to zero. One-off receipts (asset sales, lotteries, inheritances) are excluded. If all components are missing for a household, total income is assigned missing; if at least one component is observed, missing components are set to zero. When an activity filter confirms the household did not engage in a given activity, the corresponding missing income is recoded to zero. Only cash income is included where the distinction between cash and in-kind is available.

Worked Example: Guatemala (2000)

Table C.10 illustrates the income aggregate construction for Guatemala, based on the *Encuesta Nacional sobre Condiciones de Vida* (2000).

Table C.10. Income aggregate components: Guatemala (2000)

Component	Description	Survey File
Own Farm Income	Net crop income + net forestry income + net livestock income − rent of land	ECV2#A16; ECV3#A16
Wage Farm Income	Gross cash salary (incl. bonuses/tips) from main job + second job + other jobs (last 12 months)	ECV11P10
Land Rent Income	Total money received for renting out land in the last 12 months	ECV21A16; ECV22A16
Agricultural Income	= Own Farm + Wage Farm + Land Rent	
Own Business Income	Net independent income from main + second + other non-agricultural jobs	ECV11P10
Wage Income	Gross cash salary from main + second + other non-agricultural jobs	ECV11P10
Non-Agric. Income	= Own Business + Wage	
Non-Labour Income	Pensions + dividends + remittances + cash assistance + non-farm rent + scholarships	ECV16I13; ECV11P10
Total HH Income	= Agricultural + Non-Agricultural + Non-Labour	